

UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA
KATEDRA MATEMATICKÉ ANALÝZY A APLIKACÍ MATEMATIKY

BAKALÁŘSKÁ PRÁCE

Lorenzova křivka



Vedoucí bakalářské práce:
Mgr. Ondřej Vencálek
Rok odevzdání: 2010

Vypracovala:
Gabriela Večerková
ME, III. ročník

Prohlášení

Prohlašuji, že jsem bakalářskou práci zpracovala samostatně pod vedením pana Mgr. Ondřeje Vencálka s použitím uvedené literatury.

V Olomouci dne 20. dubna 2010

Poděkování

Na tomto místě bych chtěla poděkovat především svému vedoucímu bakalářské práce panu Mgr. Ondřeji Vencálkovi, že měl se mnou dostatek trpělivosti a pomohl mi dovést tuto práci ke zdárnému konci. Ráda bych také poděkovala své rodině a přátelům, že mě po celou dobu studia podporovali.

Obsah

Úvod	4
1 Podstata Lorenzovy křivky	5
1.1 Užití Lorenzovy křivky	5
1.2 Konstrukce Lorenzovy křivky	6
2 Matematické vyjádření Lorenzovy křivky	7
2.1 Pojem kvantilu náhodné veličiny X	7
2.2 Matematická formule Lorenzovy křivky	8
3 Tvar Lorenzovy křivky pro některá spojitá rozdělení	13
3.1 Rovnoměrné rozdělení	13
3.2 Exponenciální rozdělení	17
3.3 Normální rozdělení	20
4 Simulace dat	24
4.1 Princip simulace z exponenciálního rozdělení	25
4.2 Lorenzova křivka pro nasimulovaná data	26
4.3 Porovnání teoretické a nasimulované Lorenzovy křivky	27
5 Praktická aplikace	29
Závěr	32
Literatura	33

Úvod

Cílem mé bakalářské práce je seznámit čtenáře s tím, jak je možné zjistit nerovnoměrnost rozdělení dat. V této práci budeme pracovat s metodou zvanou Lorenzova křivka. Předpokládá se, že čtenář má alespoň základní znalosti z teorie pravděpodobnosti.

Práce je rozdělena do pěti kapitol. V první kapitole je stručně vysvětleno, za jakým účelem byla Lorenzova křivka sestrojena, k čemu dnes může sloužit a jakým způsobem můžeme danou křivku zkonstruovat.

V kapitole druhé je definován kvantil náhodné veličiny X , což je důležitý pojem pro tuto práci. Dále v této kapitole pomocí teoretického příkladu vysvětlujeme, jak postupně dochází k vyjádření matematického vzorce Lorenzovy křivky.

V kapitole třetí odvozujeme tvar Lorenzovy křivky pro některá spojitá rozdělení pravděpodobnosti a zkoumáme jak parametry daného rozdělení ovlivňují (případně neovlivňují) tvar křivky a o čem vypovídá tvar Lorenzovy křivky.

V kapitole čtvrté je provedena simulační studie, kde naším cílem je pro náhodně generovanou sadu dat exponenciálního rozdělení vykreslit Lorenzovu křivku. Výsledky simulační studie porovnáme s teoretickými výsledky získanými v předešlé kapitole.

V kapitole páté je zkonstruována Lorenzova křivka pro reálná data.

1. Podstata Lorenzovy křivky

1.1. Užití Lorenzovy křivky

V roce 1905 byl publikován článek *Metody měření koncentrace bohatství* od amerického statistika M.O.Lorenze, viz [4]. Jak už říká sám název, tato práce pojednává o tom, jak je možné popsat koncentraci bohatství ve společnosti. Autor se zaměřil na grafickou stránku věci. Byla zkonstruována křivka popisující rozložení bohatství v určité společnosti a dnes tato křivka nese jeho jméno.

Lorenzova křivka nám mimo jiné může posloužit ke srovnávání nerovnoměrnosti rozdělení důchodu¹ v porovnání v různých ekonomikách nebo stejné ekonomiky v různých časech, může znázornit změnu koncentrace obyvatelstva na daném území v určitém časovém období. Při srovnávání Lorenzových křivek v různých časech lze zjistit, zda se v čase bohatství koncentruje (nerovnoměrnost ve společnosti roste) nebo zda má tendenci se rozptýlit rovnoměrně.

Uplatnění Lorenzovy křivky je dosti široké. Může se použít např. v následujících oborech:

- ekonomie

Př. nerovnoměrnost rozdělení důchodu v ekonomice - vztah mezi celkovými důchody a obyvatelstvem (domácnostmi) v České republice

- demografie

Př. nerovnoměrnost rozdělení obyvatelstva na určitém území - např. vztah mezi celkovým počtem obyvatel jednotlivých krajů a rozlohou těchto krajů

- bankovní sektor

Př. diverzifikace klientů - vztah mezi „dobrými“ a „špatnými“ klienty banky

Poznámka 1.1. *Lorenzova křivka se bude dále označovat LC*

¹Důchodem se rozumí celkové množství peněz, které osoba nebo domácnost obdrží během daného časového období (obvykle roku). Důchod sestává z mezd nebo pracovních výdělků, vlastnického důchodu, jako např. rent, úroku a dividend, transférových plateb neboli příjmů od vlády, jako sociální zabezpečení nebo pojištění v nezaměstnanosti, viz [5]

1.2. Konstrukce Lorenzovy křivky

Konstrukci LC vysvětlíme na příkladu měření koncentrace příjmů v populaci:

Je zavedena kartézská soustava souřadnic. Na horizontální osu zakreslujeme kumulovaná (postupně načítaná) procenta počtu obyvatel populace, kde obyvatelé jsou seřazeni podle příjmů od těch s nejmenšími příjmy po největší. Na vertikální osu zakreslujeme kumulovaná procenta celkových příjmů, které jsou drženy daným procentem obyvatelstva.

Mohou nastat dva extrémní případy, které se v reálném životě nevyskytují, ale slouží k porovnání s případy skutečnými.

1. Tzv. absolutně rovnoměrné rozdělení příjmů.

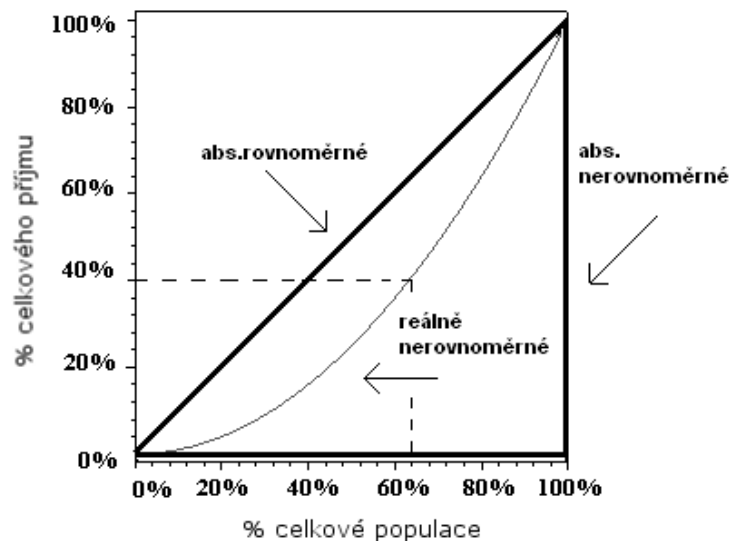
Nejchudší 1% populace získává 1% celkových příjmů.
Nejchudší 2% populace získává 2% celkových příjmů.
Nejchudší 3% populace získává 3% celkových příjmů.
Atd.

2. Tzv. absolutně nerovnoměrné rozdělení příjmů.

Veškerý příjem náleží jediné osobě.

Absolutně rovnoměrné rozdělení příjmů je v grafu zakresleno přímkou ve tvaru $y = x$. Křivka absolutně nerovnoměrného rozdělení prochází po x -ové ose a poté se přimyká k rovnoběžce k ose y ($x = 1$). Křivka reálně nerovnoměrného rozdělení příjmů se pohybuje mezi absolutně rovnoměrným a nerovnoměrným rozdělením a její tvar můžeme přirovnat ke tvaru luku.

První a poslední bod křivky je u všech případů stejný, tj. body $[0;0]$ a $[100\%;100\%]$ resp. $[1;1]$.



Obr.1 Lorenzova křivka pro abs. rovnoměrné, abs. nerovnoměrné a reálně nerovnoměrného rozdělení

Z obrázku vyčteme např., že 64% „nejchudších“ obyvatel získává jen 40% celkových příjmů.

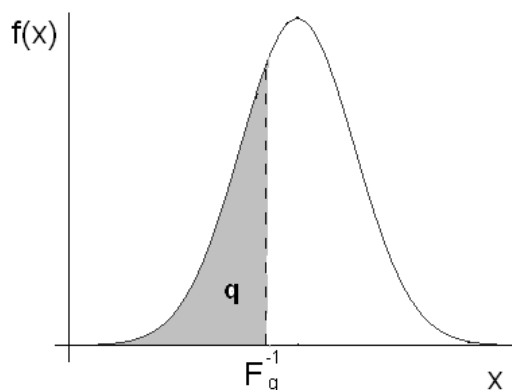
2. Matematické vyjádření Lorenzovy křivky

2.1. Pojem kvantilu náhodné veličiny X

Kvantil náhodné veličiny X je potřebným pojmem pro matematické vyjádření LC a proto zde bude tento pojem definován. Předvedeme jaký má význam a jakým způsobem ho lze získat. Následující definici najdeme v [3].

Definice 2.1. *Nechť $q \in (0, 1)$. q -kvantil náhodné veličiny X je takové reálné číslo F_q^{-1} , pro které platí*

$$P(X \leq F_q^{-1}) \geq q \quad \text{a současně} \quad P(X \geq F_q^{-1}) \geq 1 - q.$$



Obr.2 Vztah hustoty a kvantilu náhodné veličiny X

Nás zajímá náhodná veličina X , pomocí které modelujeme příjmy jednotlivých lidí v populaci. q -kvantilem rozumíme takovou výši příjmu, že právě $100 \cdot q\%$ populace má příjem nepřevyšující tuto částku.

q -kvantil lze zapsat i pomocí distribuční funkce $F_X(x)$ náhodné veličiny X (viz [1]).

Nechť F je nějaká distribuční funkce. Zavedme funkci F^{-1} předpisem

$$F_q^{-1} = \inf\{x : F(x) \geq q\}, \quad 0 < q < 1$$

Pak se F^{-1} nazývá kvantilová funkce odpovídající distribuční funkci F . Hodnotám F_q^{-1} se říká kvantily. Je-li F rostoucí funkce, pak F^{-1} je obyčejná inverzní funkce k F .

2.2. Matematická formule Lorenzovy křivky

Soudobou definici Lorenzovy křivky zavedl na počátku 70.let Gastwirth (viz [2]). Matematický vztah LC bude vysvětlen na jednoduchém příkladě: Máme čtyřicet náhodně vybraných osob určité populace. V tabulce jsou uvedené příjmy, které představují realizace náhodné veličiny X popisující příjem jednotlivých osob této populace.

i	j_i	x_i	p_i	q_i
1	2	8 000	$\frac{2}{40}$	$\frac{2}{40}$
2	3	9 000	$\frac{3}{40}$	$\frac{5}{40}$
3	6	11 000	$\frac{6}{40}$	$\frac{11}{40}$
4	4	12 000	$\frac{4}{40}$	$\frac{15}{40}$
5	4	14 000	$\frac{4}{40}$	$\frac{19}{40}$
6	9	16 000	$\frac{9}{40}$	$\frac{28}{40}$
7	7	19 000	$\frac{7}{40}$	$\frac{35}{40}$
8	3	25 000	$\frac{3}{40}$	$\frac{38}{40}$
9	1	30 000	$\frac{1}{40}$	$\frac{39}{40}$
10	1	35 000	$\frac{1}{40}$	1

Tabulka 1:

Označení:

$i \dots$ index skupiny (seřazeno podle velikosti příjmu)

$j_i \dots$ počet osob mající příjem x_i

$x_i \dots$ hodnota příjmu i -té skupiny lidí

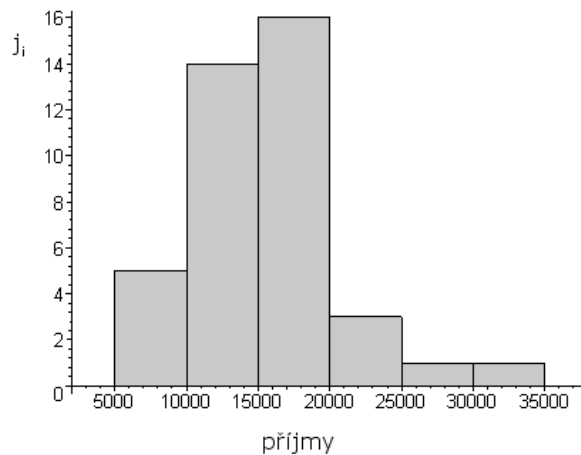
$p_i \dots$ relativní četnost i -té skupiny lidí

$q_i \dots$ kumulovaná relativní četnost, kde $q_i = \sum_{k=1}^i p_k$

V praxi se mohou rozdělit příjmy do příjmových tříd, jelikož zjištění kategorizovaných dat je snadnější. Na našem příkladu by to vypadalo následovně :

příjem v tis.Kč	5-10	10-15	15-20	20-25	25-30	30-35
počet osob	5	14	16	3	1	1

Tabulka 2:



Obr.3 Histogram příjmů náhodné veličiny X

Máme graficky znázorněnou četnost osob vyskytujících se v jednotlivých příjmových třídách.

Poznámka 2.1. *Počet tříd příjmů můžeme zvolit pomocí tzv. Sturgesova pravidla.*

$k = 1 + 3.3 \log n = 1 + 1.43 \ln n$, kde k je počet tříd a n počet dat (hodnoty příjmů), viz [3].

Body Lorenzovy křivky:

LC vznikne spojením bodů

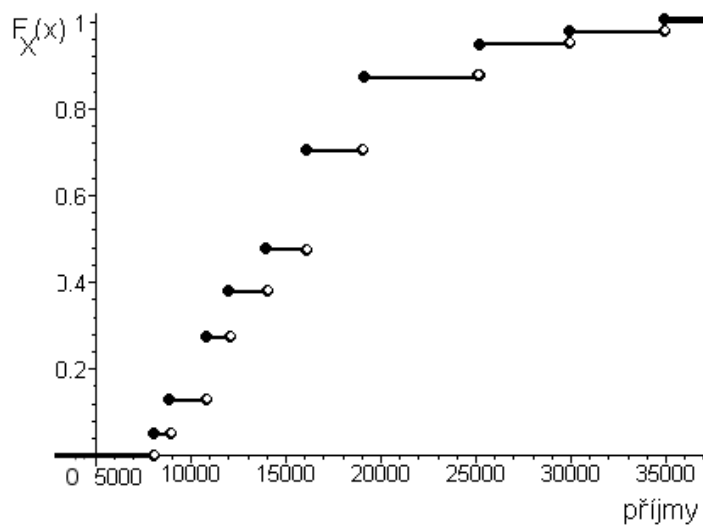
$[0,0]$, $[q_i, L(q_i)]$, kde $q_i = \sum_{k=1}^i p_k$, $i = 1, \dots, n$, a $[1,1]$.

Označení:

$q_i \dots$ postupně načítané relativní četnosti p_i

$n \dots$ obecné značení pro počet příjmových skupin (v našem případě 10)

Máme tedy diskrétní náhodnou veličinu X a známe i její distribuční funkci $F_X(x)$.



Obr.4 Distribuční funkce náhodné veličiny X

Nyní se věnujme hodnotám $L(q_i)$:

$L(q_1)$: podíl příjmu osob skupiny 1 (tj. s nejmenšími příjmy) na celkových příjmech.

$$L(q_1) = \frac{2 \cdot 8\,000}{630\,000} = \frac{x_1 j_1}{\sum_{k=1}^n x_k j_k}$$

$L(q_2)$: podíl příjmu osob skupiny 1 a 2 (tj. dvou tříd s nejnižšími příjmy) na celkových příjmech.

$$L(q_2) = \frac{2 \cdot 8\,000 + 3 \cdot 9\,000}{630\,000} = \frac{x_1 j_1 + x_2 j_2}{\sum_{k=1}^n x_k j_k}$$

obecně:

$$L(q_i) = \frac{\sum_{k=1}^i x_k j_k}{\sum_{k=1}^n x_k j_k} \quad i = 1, \dots, n \quad (1)$$

Tento vztah lze upravit tak, že počet osob v jednotlivých skupinách $j_1, \dots, j_n$ vyjádříme pomocí relativních četností tříd $p_1, \dots, p_n$.

Platí:

$$j_i = mp_i, \quad i = 1, \dots, n$$

kde m vyjadřuje počet osob sledované třídy.

$$L(p_1) = \frac{x_1 mp_1}{m \sum_{k=1}^n x_k p_k} = \frac{x_1 p_1}{\sum_{k=1}^n x_k p_k}$$

$$L(p_1 + p_2) = \frac{x_1 p_1 + x_2 p_2}{\sum_{k=1}^n x_k p_k}$$

$\vdots$

$$L\left(\sum_{k=1}^i p_k\right) = \frac{\sum_{k=1}^i x_k p_k}{\sum_{k=1}^n x_k p_k}$$

Funkce $L(q)$ je pro diskrétní náhodnou veličinu X vyjádřena vztahem :

$$L(q_i) = \frac{\sum_{k=1}^i x_k p_k}{\sum_{k=1}^n x_k p_k}, \quad i = 1, \dots, n \quad (2)$$

Analogicky pro spojitou náhodnou veličinu X je funkce $L(q)$ dána vztahem:

$$L(q) = \frac{\int_{-\infty}^{F_q^{-1}} x f(x) dx}{\int_{-\infty}^{\infty} x f(x) dx}, \quad q \in (0, 1) \quad (3)$$

kde F_q^{-1} je q -kvantilem náhodné veličiny X a $f(x)$ je hustotou této náhodné veličiny

Poznámka 2.2. *Jmenovatel funkce $L(q)$, tzn. $\sum_{k=1}^n x_k p_k$ resp. $\int_{-\infty}^{\infty} x f(x) dx$ vyjadřuje střední hodnotu $E(X)$ náhodné veličiny X .*

Poznámka 2.3. *V dalším textu budeme uvažovat, že X je spojitá náhodná veličina.*

Následující kapitola se bude zabývat tvarem závislosti $L(q)$ pro některá známá pravděpodobnostní rozdělení.

3. Tvar Lorenzovy křivky pro některá spojitá rozdělení

Předpokládáme, že naše data jsou realizacemi spojitě náhodné veličiny X . V této kapitole si předvedeme jak se LC bude „chovat“ v případě, že rozdělení pravděpodobnosti náhodné veličiny X bude rovnoměrné, exponenciální a normální. Budeme pozorovat, jaký vliv na tvar LC budou mít parametry jednotlivých rozdělení pravděpodobnosti.

Dosazením příslušné hustoty, kvantilu a mezí integrálu do (3), získáme vztah pro konstrukci LC .

3.1. Rovnoměrné rozdělení

Náhodná veličina X má rovnoměrné rozdělení pravděpodobnosti na intervalu (a, b) , kde $a, b \in \mathcal{R}$,

$$X \sim Ro(a, b)$$

má-li hustotu

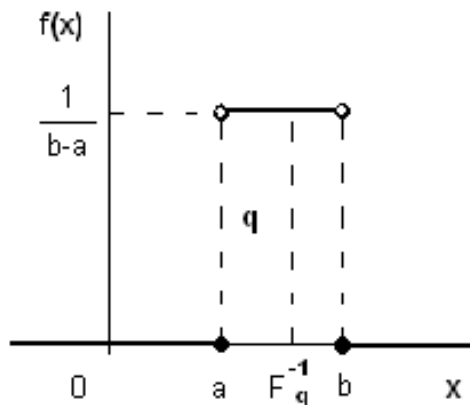
$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in (a, b) \\ 0, & x \notin (a, b), \end{cases}$$

viz [3].

LC po dosazení hustoty do (3):

$$L(q) = \frac{\int_a^{F_q^{-1}} x \frac{1}{b-a} dx}{\int_a^b x \frac{1}{b-a} dx}, \quad q \in (0, 1) \quad (4)$$

Prvním krokem je zjištění kvantilu F_q^{-1} :



Obr.5 Hustota rovnoměrného rozdělení

$$q = \frac{F_q^{-1} - a}{b - a}$$

Po upravě získáme kvantil

$$F_q^{-1} = q(b - a) + a$$

F_q^{-1} dosadíme do vztahu (4) a po úpravě získáme vztah pro znázornění LC :

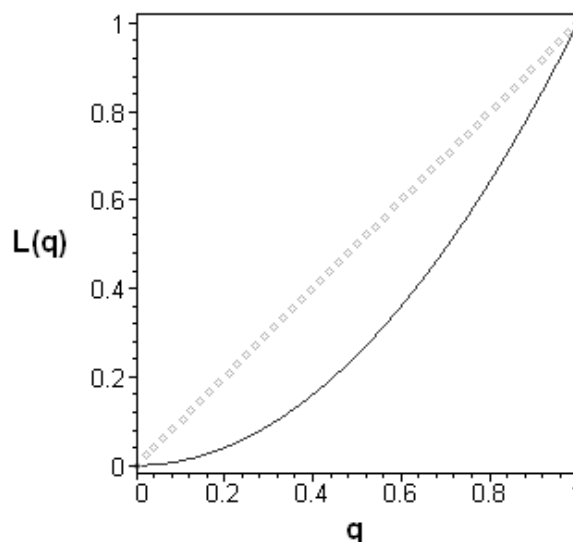
$$L(q) = \frac{\int_a^{q(b-a)+a} x \frac{1}{b-a} dx}{\int_a^b x \frac{1}{b-a} dx} = \frac{q^2(b-a) + 2aq}{a+b}$$

Lorenzova křivka je dána vztahem:

$$L(q) = \frac{q^2(b-a) + 2aq}{a+b}, \quad q \in (0, 1) \quad (5)$$

Graficky znázorněna LC

pro volbu parametrů: $a = 0, \quad b = 1$



Obr.6 LC pro rovnoměrné rozdělení náhodné veličiny X

Závislost křivky na parametrech

Hodnota $L(q)$ závisí na parametrech a a b . Jak bude tvar křivky ovlivněn parametry, bude předvedeno na dvou případech. V prvním případě bude rozptyl $var(X)$ konstantní a bude se měnit pouze střední hodnota $E(X) = \frac{a+b}{2}$. A v případě druhém střední hodnota $E(X)$ je konstantní a mění se rozptyl $var(X) = \frac{(b-a)^2}{12}$.

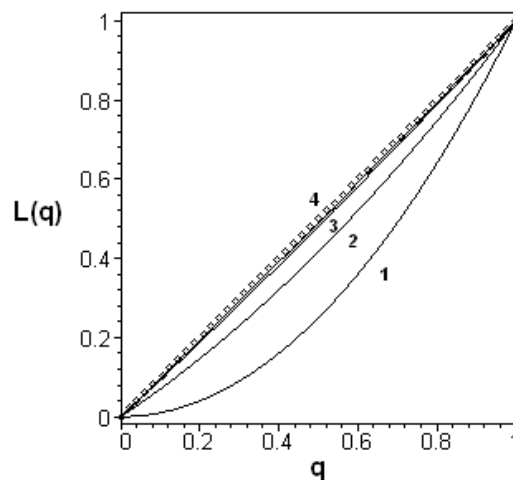
- 1.případ

$$LC_1 \dots a = 0, \quad b = 1 \quad \dots \quad E(X_1) = 0.5, \quad var(X_1) = \frac{1}{12}$$

$$LC_2 \dots a = 1, \quad b = 2 \quad \dots \quad E(X_2) = 1.5, \quad var(X_2) = \frac{1}{12}$$

$$LC_3 \dots a = 6, \quad b = 7 \quad \dots \quad E(X_3) = 6.5, \quad var(X_3) = \frac{1}{12}$$

$$LC_4 \dots a = 12, \quad b = 13 \quad \dots \quad E(X_4) = 12.5, \quad var(X_4) = \frac{1}{12}$$



Obr.7 LC pro rovnoměrné rozdělení při konstantním $var(X)$

Čím více se střední hodnota $E(X)$ vzdaluje od hodnoty 0, tím více se funkce $L(q)$ přibližuje ke křivce absolutně rovnoměrného rozdělení, tj. $L(q) = q$.

Srovnajme například LC_2 a LC_4 . Zatímco LC_2 je poměr největší a nejmenší hodnoty náhodné veličiny X roven

$$\frac{b}{a} = \frac{2}{1} = 2,$$

v případě LC_4 je tento poměr roven

$$\frac{b}{a} = \frac{13}{12} = 1,08.$$

Tzn., že v prvním případě je větší nerovnoměrnost než v druhém případě a tuto skutečnost vidíme i na obrázku 7.

- 2.případ

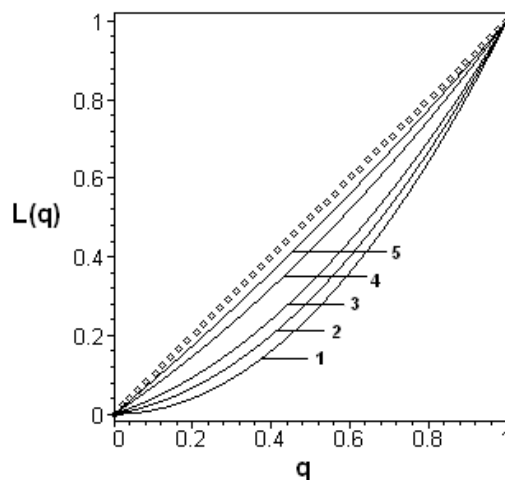
$$LC_1 \dots a = 0, \quad b = 12 \quad \dots \quad E(X_1) = 6, \quad var(X_1) = 12$$

$$LC_2 \dots a = 1, \quad b = 11 \quad \dots \quad E(X_2) = 6, \quad var(X_2) = 8.33$$

$$LC_3 \dots a = 2, \quad b = 10 \quad \dots \quad E(X_3) = 6, \quad var(X_3) = 5.33$$

$$LC_4 \dots a = 4, \quad b = 8 \quad \dots \quad E(X_4) = 6, \quad var(X_4) = 1.33$$

$$LC_5 \dots a = 5, \quad b = 7 \quad \dots \quad E(X_5) = 6, \quad var(X_5) = 0.33$$



Obr.8 LC pro rovnoměrné rozdělení při konstantní $E(X)$

Hodnota rozptylu $\text{var}(X)$ náhodné veličiny X postupně klesá a tedy i nerovnoměrnost rozdělení (v našem případě) příjmů se snižuje.

Poznámka 3.1. Pokud zvolíme parametr $a = 0$, tvar LC není ovlivněn parametrem b .

Dosadíme do (5)

$$L(q) = q^2.$$

Tzn. rozptyl $\text{var}(X)$ náhodné veličiny X , ať je jakkoli velký, nerovnoměrnost rozdělení příjmů je stále stejné.

3.2. Exponenciální rozdělení

Náhodné veličina X má exponenciální rozdělení pravděpodobnosti s parametrem λ , kde $\lambda > 0$

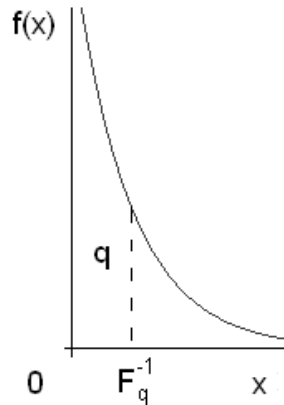
$$X \sim \text{Exp}(\lambda)$$

má-li hustotou

$$f(x) = \begin{cases} 0, & x \leq 0 \\ \frac{1}{\lambda} e^{-\frac{x}{\lambda}}, & x > 0, \end{cases}$$

viz [3].

Výpočet kvantilu F_q^{-1}



Obr.9 Hustota exponenciálního rozdělení

Distribuční funkci F_X můžeme vyjádřit jako integrál z hustoty a chceme, aby platilo $q = F_X(F_q^{-1})$.

$$q = F_X(F_q^{-1}) = \int_0^{F_q^{-1}} \frac{1}{\lambda} e^{-\frac{x}{\lambda}} dx = 1 - e^{-\frac{F_q^{-1}}{\lambda}}$$

Po úpravě

$$F_q^{-1} = -\lambda \ln(1 - q)$$

Dosadíme do vztahu (3) pro LC:

$$L(q) = \frac{\int_0^{-\lambda \ln(1-q)} x \frac{1}{\lambda} e^{-\frac{x}{\lambda}} dx}{\int_0^{\infty} x \frac{1}{\lambda} e^{-\frac{x}{\lambda}} dx}$$

Čitatel:

Integrál

$$\int_0^{-\lambda \ln(1-q)} x \frac{1}{\lambda} e^{-\frac{x}{\lambda}} dx$$

můžeme vyjádřit pomocí substituce ($t = -\frac{x}{\lambda}$) jako

$$-\lambda \int_{\ln(1-q)}^0 t e^t dt$$

Dále použijeme metodu per partes

$$-\lambda \int_{\ln(1-q)}^0 t e^t dt = -\lambda ([te^t]_{\ln(1-q)}^0 - \int_{\ln(1-q)}^0 dt) = -\lambda(-\ln(1-q) + q\ln(1-q) - q)$$

Jmenovatel:

Při výpočtu jmenovatele použijeme stejný postup jako u výpočtu čitatele.

Máme tedy integrál

$$\int_0^\infty x \frac{1}{\lambda} e^{\frac{-x}{\lambda}} dx,$$

který vyjádříme pomocí substituce a metody per partes

$$\int_0^\infty \frac{1}{\lambda} x e^{\frac{-x}{\lambda}} dx = \lim_{y \rightarrow \infty} \left(\lambda \frac{-y}{\lambda} e^{\frac{-y}{\lambda}} - e^{\frac{-y}{\lambda}} \right) - \lambda \left(\frac{0}{\lambda} e^{\frac{0}{\lambda}} - e^{\frac{0}{\lambda}} \right) = \lambda.$$

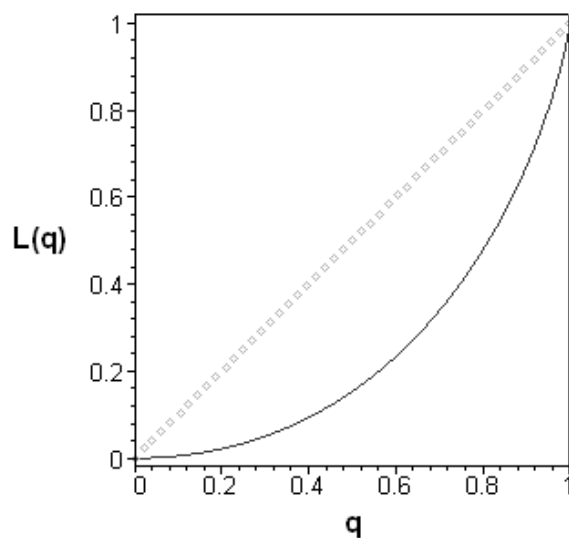
A dostáváme konečný tvar LC :

$$L(q) = \frac{-\lambda(-\ln(1-q) + q\ln(1-q) - q)}{\lambda} = \ln(1-q) - q\ln(1-q) + q$$

Lorenzova křivka je dána vztahem:

$$L(q) = \ln(1-q) - q\ln(1-q) + q, \quad q \in (0, 1) \quad (6)$$

Graficky znázorněna LC



Obr.10 LC pro exponenciální rozdělení

Závislost křivky na parametrech

Parametr λ (tzn.střední hodnota $E(X)$) neovlivňuje tvar LC , což je patrné ze vztahu (6).

3.3. Normální rozdělení

Náhodná veličina X má normální rozdělení pravděpodobnosti

$$X \sim N(\mu, \sigma^2)$$

má-li hustotu

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

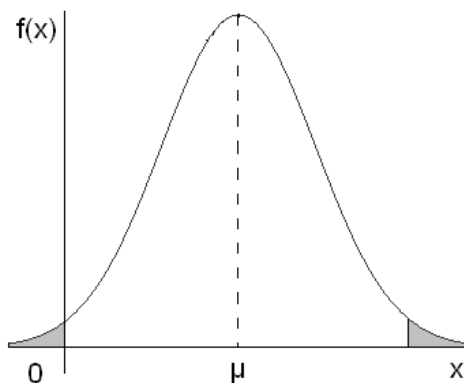
viz [3].

Směrodatná odchylka σ

Hodnoty, které používáme k sestrojení LC jsou nezáporná reálná čísla. Směrodatnou odchylku σ vyjádříme pomocí střední hodnoty μ tak, aby data nabývala záporných hodnot jen se zanedbatelnou pravděpodobností, tj.

$$P(X < 0) = q,$$

kde $q \dots$ zanedbatelná pravděpodobnost, např. $q = 0.0001$.



Obr.11 Hustota normálního rozdělení

Provedeme normalizaci náhodné veličiny X , tj. $\frac{X-\mu}{\sigma} \sim N(0,1)$

$$P\left(\frac{X-\mu}{\sigma} < \frac{0-\mu}{\sigma}\right) = q$$

Chceme, aby platilo

$$u_q = \frac{0-\mu}{\sigma} \quad \text{a tedy} \quad \sigma = -\frac{\mu}{u_q}$$

$u_q \dots q$ -kvantil normálního normovaného rozdělení $N(0,1)$ (tabelovaná hodnota).

Obecně budeme uvažovat vztah

$$\sigma = \frac{\mu}{k},$$

kde k je dostatečně velká konstanta ($k \geq 4$).

LC po dosazení hustoty do (3):

$$L(q) = \frac{\int_{-\infty}^{F_q^{-1}} x \frac{1}{\frac{\mu}{k}\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\frac{\mu^2}{k^2}}} dx}{\int_{-\infty}^{\infty} x \frac{1}{\frac{\mu}{k}\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\frac{\mu^2}{k^2}}} dx}, \quad q \in (0,1) \quad (7)$$

kde F_q^{-1} je q -kvantilem normálního rozdělení $N(\mu, (\frac{\mu}{k})^2)$.

Výpočet kvantilu F_q^{-1}

Kvantil F_q^{-1} vyjádříme pomocí kvantilu normálního normovaného rozdělení u_q .

Máme náhodnou veličinu $Y = \frac{X-\mu}{\mu}k$ s normálním normovaným rozdělením $Y \sim N(0, 1)$, kde platí

$$P(Y < u_q) = q$$

$$P\left(\frac{X-\mu}{\mu}k < u_q\right) = q$$

$$P(X < u_q \frac{\mu}{k} + \mu) = q$$

odtud

$$F_q^{-1} = u_q \frac{\mu}{k} + \mu$$

Čitatel:

Integrál v čitateli ze vztahu (7) po dosazení F_q^{-1}

$$\int_{-\infty}^{u_q \frac{\mu}{k} + \mu} \frac{1}{\frac{\mu}{k} \sqrt{2\pi}} x e^{-\frac{(x-\mu)^2}{2(\frac{\mu}{k})^2}} dx$$

vyjádříme jako

$$\int_{-\infty}^{u_q} \frac{1}{\frac{\mu}{k} \sqrt{2\pi}} \frac{\mu}{k} \left(\frac{\mu}{k}t + \mu\right) e^{-\frac{t^2}{2}} dt = \int_{-\infty}^{u_q} \frac{1}{\sqrt{2\pi}} \frac{\mu}{k} t e^{-\frac{t^2}{2}} dt + \mu \int_{-\infty}^{u_q} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

(byla použita substituce $(t = \frac{x-\mu}{\frac{\mu}{k}})$).

Dále využijeme substituci $(y = -\frac{t^2}{2})$ a vztah $\Phi(u_q) = q$, kde $\Phi(u_q)$ je distribuční funkce normálního normovaného rozdělení $N(0,1)$

$$-\frac{1}{\sqrt{2\pi}} \frac{\mu}{k} \lim_{s \rightarrow -\infty} (e^{-\frac{u_q^2}{2}} - e^s) + \mu q = -\frac{1}{\sqrt{2\pi}} \frac{\mu}{k} e^{-\frac{u_q^2}{2}} + \mu q = \mu \left(-\frac{1}{k\sqrt{2\pi}} e^{-\frac{u_q^2}{2}} + q\right)$$

Jmenovatel:

Při výpočtu integrálu

$$\int_{-\infty}^{\infty} \frac{1}{\frac{\mu}{k} \sqrt{2\pi}} x e^{-\frac{(x-\mu)^2}{2\frac{\mu^2}{k^2}}} dx$$

můžeme opět využít substituce použité při výpočtu čitatele

$$\frac{1}{\sqrt{2\pi}} \left(\int_{-\infty}^{\infty} \frac{\mu t}{4} e^{-\frac{t^2}{2}} dt + \mu \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt \right) = \frac{1}{\sqrt{2\pi}} (0 + \mu \sqrt{2\pi}) = \mu$$

(pro integrál $\int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt$ jsme použili transformaci pomocí polárních souřadnic).
Jednotlivé výsledky dosadíme do (7) a upravíme:

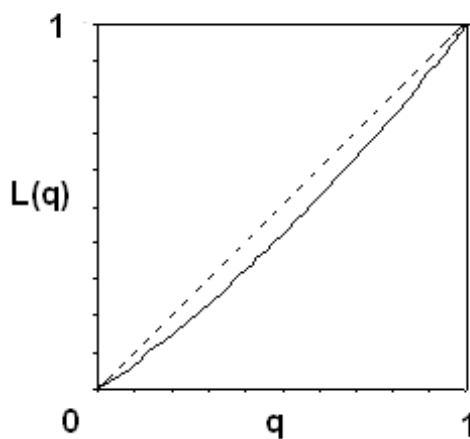
$$L(q) = \frac{\mu \left(-\frac{1}{k\sqrt{2\pi}} e^{-\frac{u_q^2}{2}} + q \right)}{\mu} = -\frac{1}{k\sqrt{2\pi}} e^{-\frac{u_q^2}{2}} + q$$

Lorenzova křivka je dána vztahem:

$$L(q) = q - \frac{1}{k\sqrt{2\pi}} e^{-\frac{u_q^2}{2}} \quad q \in (0, 1) \quad (8)$$

Graficky znázorněna LC

pro volbu $k=5$



Obr.12 LC pro normální rozdělení

Závislost křivky na parametrech

Tvar LC je ovlivněn střední hodnotou a rozptylem náhodné veličiny X (stejně jako u rovnoměrného rozdělení). Proto si opět uvedeme dva příklady, kde jedna z uvedených číselných charakteristik bude konstantní a druhá ne, a naopak.

1. Střední hodnota μ je konstantní, budeme měnit hodnotu konstanty k a tím i rozptyl $(\frac{\mu}{k})^2$.

Při snižování hodnoty rozptylu $(\frac{\mu}{k})^2$ dochází ke snížení nerovnoměrnosti rozdělení např. příjmů.

2. Rozptyl σ^2 je konstantní, střední hodnotu vyjádříme pomocí rozptylu, tj. $\mu = k\sigma$. Budeme opět měnit hodnotu konstanty k .

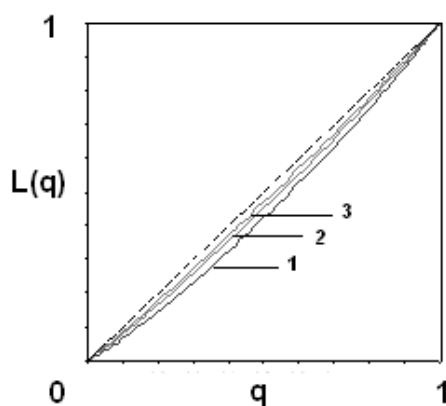
Při zvyšování střední hodnoty $\mu = k\sigma$ dochází ke snížení nerovnoměrnosti rozdělení.

Změna konstanty k :

pro $LC_1 \dots k=5$

pro $LC_2 \dots k=8$

pro $LC_3 \dots k=12$



Obr.13 LC pro normální rozdělení při změně hodnoty konstanty k

4. Simulace dat

V této kapitole se budeme zabývat simulací dat a to pro případ exponenciálního rozdělení náhodné veličiny X . Získané výsledky budeme konfrontovat s výsledky teoretickými, které jsme zjistili v kapitole 3. Simulaci dat budeme provádět na různém rozsahu výběru náhodných veličin.

4.1. Princip simulace z exponenciálního rozdělení

Většina generátorů pseudonáhodných čísel je schopna generovat posloupnost pseudonáhodných čísel z rovnoměrného rozdělení pravděpodobnosti na intervalu $(0,1)$. Budeme tedy generovat data z rovnoměrného rozdělení a poté je transformovat na požadované rozdělení pravděpodobnosti.

Následující větu použijeme na transformaci dat z rovnoměrného rozdělení pravděpodobnosti na data s exponenciálním rozdělením.

Věta 4.1. *Nechť náhodná veličina U má spojitě rovnoměrné rozdělení na intervalu $(0,1)$. Nechtě F^{-1} je kvantilová funkce odpovídající nějaké rostoucí spojitě distribuční funkci F . Pak náhodná veličina $X = F^{-1}(U)$ má distribuční funkci F .*

Máme náhodnou veličinu U , která má rovnoměrné rozdělení pravděpodobnosti,

$$U \sim Ro(0, 1).$$

Chceme, aby náhodná veličina X měla exponenciální rozdělení pravděpodobnosti

$$X \sim exp(\lambda)$$

s distribuční funkcí F

$$F = 1 - e^{-\frac{x}{\lambda}}.$$

Transformace dat:

Použijeme větu 4.1., odkud víme

$$X = F^{-1}(U).$$

$F^{-1}(U)$ je inverzní funkce k distribuční funkci F , tj.

$$F^{-1}(U) = -\lambda \ln(1 - u)$$

a tedy

$$X = -\lambda \ln(1 - u), \quad u \in (0, 1)$$

Na následující rovnici si ukážeme, že takto transformovaná náhodná veličina X má skutečně námi potřebné exponenciální rozdělení pravděpodobnosti.

$$\begin{aligned} F_X(x) &= P(X < x) = P(-\lambda \ln(1 - u) < x) = P(u < 1 - e^{-\frac{x}{\lambda}}) = F_U(1 - e^{-\frac{x}{\lambda}}) = \\ &= 1 - e^{-\frac{x}{\lambda}} \end{aligned}$$

Poznámka 4.1. *Obdobně bychom mohli transformovat data např. i pro normální rozdělení.*

4.2. Lorenzova křivka pro nasimulovaná data

Algoritmus konstrukce LC má několik kroků. Následující popis algoritmu je doplněn zdrojovým kódem jazyku Maple.

1. Generování 1 000 náhodných hodnot rovnoměrného rozdělení a transformace těchto dat na data z exponenciálního rozdělení

```
u:=stats[ random, uniform ](1000);
x:=-8*ln( 1-u[n] ) $ n=1...1000;
kde λ=8
```

2. Seřazení transformovaných dat od nejmenší po největší hodnotu

```
X:=sort([x]);
```

3. Výpočet x -ových souřadnic bodů LC (q_i)

```
for i from 1 to 1000 do
q[i]:=i/1000;
end do;
```

4. Výpočet kumulovaných hodnot realizací náhodné veličiny X , tj. $a_i = \sum_{k=1}^i x_k$

```
i = 1, ..., n
soucet:=0;
for i from 1 to 1000 do
soucet:=soucet+x[i];
```

```

a[i]:=soucet;
end do:

```

5. Výpočet bodů LC na y -ových souřadnic získané znormováním jednotlivých kumulovaných součtů

```

for i from 1 to 1000 do
q[i]:=a[i]/a[1000];
end do:

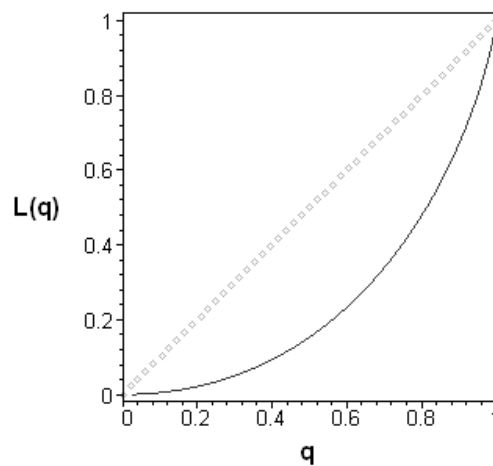
```

6. Vykreslení LC

```

LC:=( [q[n],L[n]] $ n=1..1000);
plot([LC]);

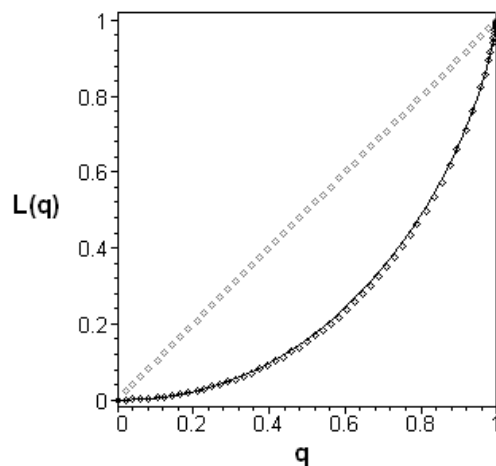
```



Obr.14 LC pro 1 000 náhodně generovaných dat

4.3. Porovnání teoretické a nasimulované Lorenzovy křivky

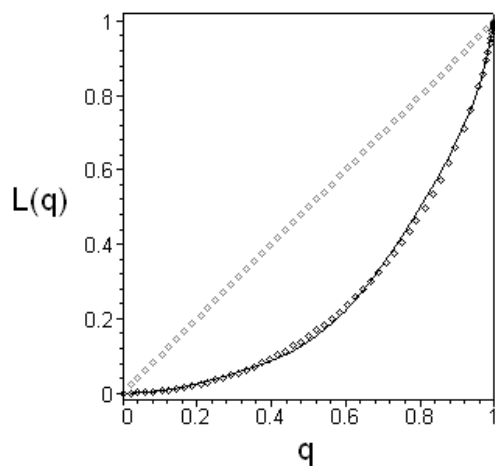
Ukážeme si, zda LC zkonstruována z 1 000 náhodně generovaných hodnot exponenciálního rozdělení bude tvarově odlišná od LC ze vztahu (6).



Obr.15 LC pro 1 000 náhodně generovaných dat a teoretické výsledky

Na obrázku 15 jsou plnou čarou vykresleny teoretické hodnoty LC pro exponenciální rozdělení, dle námi spočteného vztahu (6). Dále jsou zakresleny jednotlivé body odhadu LC , který je založen na simulačních datech. Tento odhad je dostatečně blízký teoretickým hodnotám.

Nasimulujeme 50 dat z exponenciálního rozdělení stejným způsobem jako v předešlém textu. A bude nás zajímat jak se tvar LC bude lišit od předešlého nasimulovaného příkladu.



Obr.16 LC pro 50 náhodně generovaných dat a teoretické výsledky

Ani pro mnohem menší rozsah náhodné veličiny X se křivka příliš neodchýlila od teoretického modelu.

5. Praktická aplikace

V této kapitole se budeme zabývat reálnými daty. Budeme konstruovat LC znázorňující nerovnoměrnost rozmístění obyvatelstva na území České republiky za rok 2006. Data, která jsou pro konstrukci nutná, jsme získali z Českého statistického úřadu, viz [6]. Jedná se o počet obyvatel v jednotlivých okresech a rozlohu území podle těchto okresů, kterých je 77.

Postup pro konstrukci LC bude obdobný jako v předcházejícím textu.

1. Zjistíme hustotu zalidnění v jednotlivých okresech.

$$hustota = \frac{O}{R},$$

kde O je počet obyvatel v okrese a R je rozloha okresu

2. Seřadíme okresy podle jejich hustoty zalidnění od nejmenší po největší (od Prachatic po Hlavní město Prahu).

3. Zjistíme x-ové souřadnice bodů LC

Na x-ovou osu zakreslujeme kumulovaná procenta z celkové rozlohy

$$x_i = \frac{\sum_{k=1}^i r_k}{\sum_{k=1}^{77} r_k}, \quad i = 1, \dots, 77$$

r_k je hodnota rozlohy v k -tém okrese

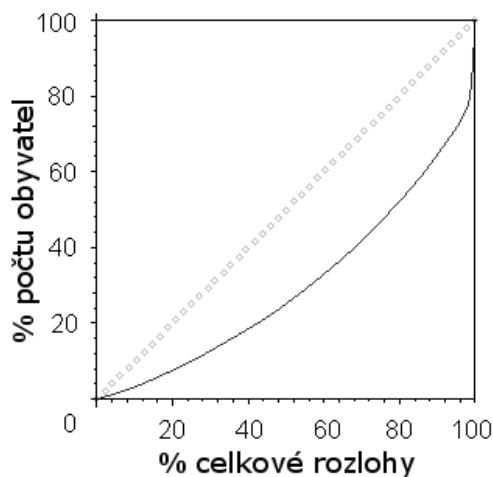
4. Zjistíme y-ové souřadnice bodů LC

Na y-ovou osu zakreslujeme kumulovaná procenta z celkového počtu obyvatel

$$y_i = \frac{\sum_{k=1}^i s_k}{\sum_{k=1}^{77} s_k}, \quad i = 1, \dots, 77$$

s_k je počet obyvatel v k -tém okrese

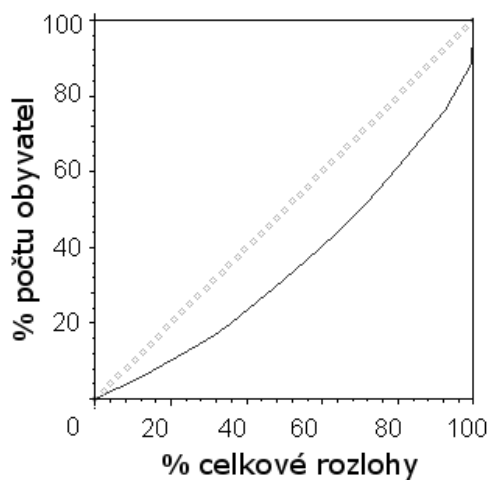
5. Zakreslíme body $LC [x_i, y_i]$, $i = 1, \dots, 77$



Obr.17 LC pro okresy

Z obrázku 17 můžeme např. říci, že 80% obyvatel osidlovalo okolo 98% území České republiky v roce 2006, tedy že na 2% nejhustěji osídleného území žilo 20% obyvatelstva.

Stejnou analýzu provedeme na datech pro kraje České republiky z roku 2006, LC znázorňuje následující obrázek.



Obr.18 LC pro kraje

Provedeme-li stejnou úvahu, jako v případě okresů, řekneme, že 80% obyvatel osidlovalo okolo 92% území České republiky v roce 2006. V porovnání z obrázkem 18 je v tomto případě nerovnoměrnost rozmístění menší. Na obrázku 17 a 18 jsme vypočítali body LC a to pro stejný počet celkových obyvatel a stejnou celkovou rozlohu, avšak příklady se různily v územních jednotkách. Můžeme říci, že tvar LC v aplikaci na demografická data silně závisí na velikosti územních jednotek.

Závěr

Tato bakalářská práce se zabývala způsobem, jak popsat stupeň nerovnoměrnosti rozdělení dat. Lorenzova křivka tuto nerovnoměrnost graficky znázorňuje. V porovnání s jinými metodami je grafická metoda jednodušší např. pro srovnání nerovnoměrnosti důchodů po dvou po sobě následujících letech. Snadno z grafu zjistíme, jak je nerovnoměrnost velká.

Při konstrukci bodů Lorenzovy křivky pracujeme s jednoduchými výpočty. Pokud však nemáme k dispozici nějaký počítačový program, výpočet těchto bodů je dosti zdoluhavý.

V průběhu této práce jsme zjistili, že tvar LC závisí na velikosti střední hodnoty a rozptylu náhodné veličiny X , jejíž nerovnoměrnost LC popisuje. Právě tvar LC ukazuje, jak velká je nerovnoměrnost rozdělení. Změnu tvaru LC jsme předváděli pro některá spojitá rozdělení pravděpodobnosti. Zjistili jsme, že při konstantní střední hodnotě a rostoucím rozptylu roste i nerovnoměrnost. A naopak při konstantním rozptylu a rostoucí střední hodnotě dochází ke snížení nerovnoměrnosti. Toto pozorování jsme učinili pro rovnoměrné a normální rozdělení náhodné veličiny X , avšak tuto úvahu nelze provést pro exponenciální rozdělení náhodné veličiny X , jelikož střední hodnota je rovna směrodatné odchylce.

Protože se jedná o jednoduchý způsob vyjádření nerovnoměrnosti a to jak z hlediska interpretace, tak i z hlediska konstrukce, můžeme tuto metodu nalézt např. ve většině knih o ekonomii.

Literatura

- [1] Anděl, J.: Základy matematické statistiky. MATFYZPRESS, Praha, 2005.
- [2] Gastwirth, J.L.: A general definition of The Lorenz curve, *Econometrica*, 39 (1971), 1037-1038.
- [3] Kunderová, P.: Základy pravděpodobnosti a matematické statistiky. Vydavatelství UP, Olomouc, 2004.
- [4] Lorenz, M.O.: Methods of Measuring the Concentration of Wealth, *Publications of the American Statistical Association Vol.9* , No 70 (1905), 209–219.
- [5] Samuelson, P.A., Nordhaus W.D.: EKONOMIE. Nakladatelství Svoboda, Praha, 1991.
- [6] Rozloha území ČR podle okresů, počet obyvatel, hustota obyvatelstva na 1 km² a počet obcí, [online], dostupné z $\langle$ <http://vdb.czso.cz/vdbvo/tabparam.jsp?voa=tabulka&cislatab=2005-02-03&stranka=1&kapitola.id=5> $\rangle$ [citováno 1.4.2010]