УНИВЕРЗИТЕТ У БЕОГРАДУ
МАТЕМАТИЧКИ ФАКУЛТЕТ

Јована Дубљанин

Лоренцова крива и Ђинијев коефицијент – оцене показатеља и примене

— мастер рад —

Београд, 2017.

Садржај

Поглавље 1

Лоренцова крива и Ђинијев коефицијент – настанак, особине и значај

Значај Лоренцове криве и Ђинијевог коефицијента, као показатеља неједнакости у расподели, присутан је у најразличитијим друштвеним и научним сферама и њихово поље деловања се шири са потребама друштва. Идеја о овим статистикама се најпре родила у економији, па не чуди што су појмови Лоренцове криве и Ђинијевог коефицијента најприсутнији у савременим економским наукама, како у изучавању националне економије и расподеле дохотка, тако и у теорији цена и макроекономији. У циљу мерења економских неједнакости, намеће се претпоставка постојања одређених показатеља који омогућавају рангирање и упоређивање расподела дохотка, односно богатства, различитих посматраних земаља, региона, па чак и тренутака у времену. Лоренцова крива и Ђинијев коефицијент заиста и јесу најпознатије мере испитивања доходне неједнакости, како због своје поузданости, тако и због лаког тумачења и визуелне интерпретације података. *Графичко представљање података* доприноси интерпретацији резултата, јер се информације које добијамо визуално лакше памте и пореде. Подаци који су правилно презентовани и интерпретирани помажу у доношењу разних одлука.

1.1 Лоренцова крива

Одговор на питање, да ли је укупна количина новчаних доходака распоређена равномерно или неравномерно, не можемо добити из дескриптивних статистика. Међутим, на ово питање, као и многа слична њему, одговор даје Лоренцова крива.

Лоренцова крива пружа веома поуздану визуелну представу о расподели дохотка у једној земљи. Име је добила по свом творцу Максу Лоренцу[1]. Лоренц је био незадовољан и критички настројен према методама које су се користиле за утврђивање кретања расподеле дохотка, односно да ли расподела дохотка постаје више или мање неједнака (неравномерна).

Лоренц је сматрао:,,*Желимо да будемо способни да кажемо у ком тренутку ће заједница бити смештена између два екстремна случаја, једнакости са једне стране и поседовања индивидуалног богатства са друге стране!”*.[6]

Да би остварио овај циљ Лоренц уводи нови приступ, касније назван Лоренцова крива, који у рачун симултано уводи промене у дохотку и популацији, што било које две заједнице са потпуно супротним условима може довести у поредбени положај. Ова крива данас има веома широку примену у анализи економских и социо-економских појава, о чему ће у наставку бити више речи, мада је првобитно примењивана искључиво у проучавању расподеле новчаног дохотка између појединачних чланова једног друштва, поређењу расподеле дохотка одређене земље у различитим временским периодима или поређење расподеле дохотка различитих земаља у истом временском периоду.

1.2 Формирање и смисао Лоренцове криве

Лоренцова крива се посматра у првом квадранту координатног система, где x и y узимају вредности ≤ 1. Смештена је у квадрату странице 1 приказаном на *Слици 1*, који је касније назван *Ђинијев квадрат*, према нумеричком показатељу неједнакости изведеном из Лоренцове криве.

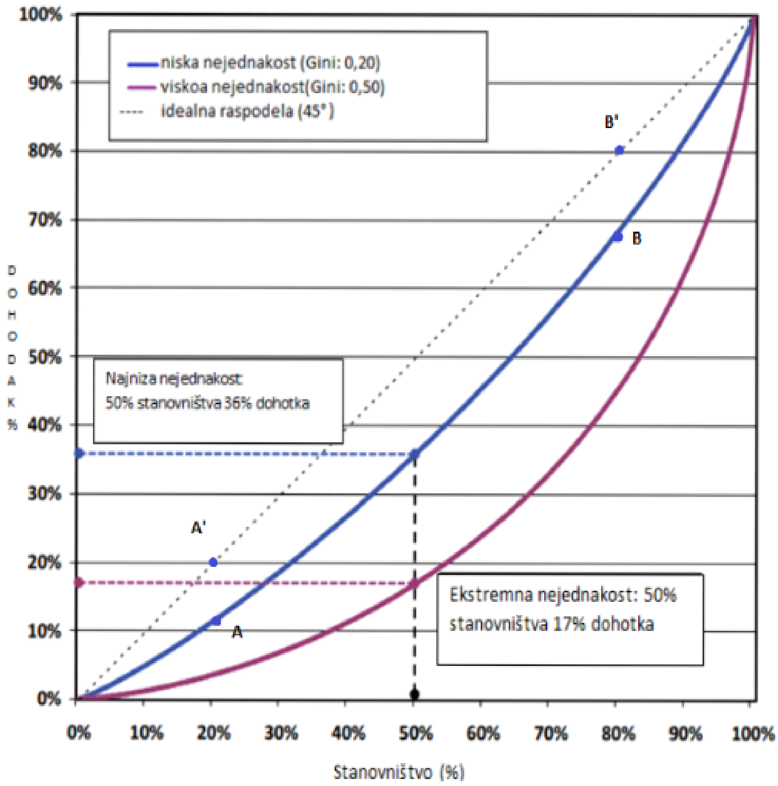
Поступак формирања Лоренцове криве је следећи:

1. На апсцису се наносе кумулативни проценти становништва (од најсиромашнијег ка најбогатијем), а на ординату процентуално изражен доходак.

2. Подаци се преносе на координатну раван, тако да се добијају различите тачке. На пример, на *Слици 1*, тачки A одговарају вредности од 20% на скали становништва и 11% на скали дохотка. То значи да најсиромашнијих 20% становника присваја само 11% укупног дохотка. Док са

[1] Max Otto Lorenz (1876-1959), амерички економиста

друге стране, тачка B има координате (80%, 68%). Аналогно, закључујемо да 80% најсиромашнијих становника остварује 68% националног дохотка. Информација садржана у тачки B може се и другачије интерпретирати, посматрањем Лоренцове криве "одозго" – најбогатијих 20% становника присваја 32% укупног дохотка. Према томе, тачка A нам даје податак да најбогатијих 80% становника присваја чак 89% укупног дохотка. Линија која спаја описане тачке назива се **Лоренцова крива**.



Слика 1

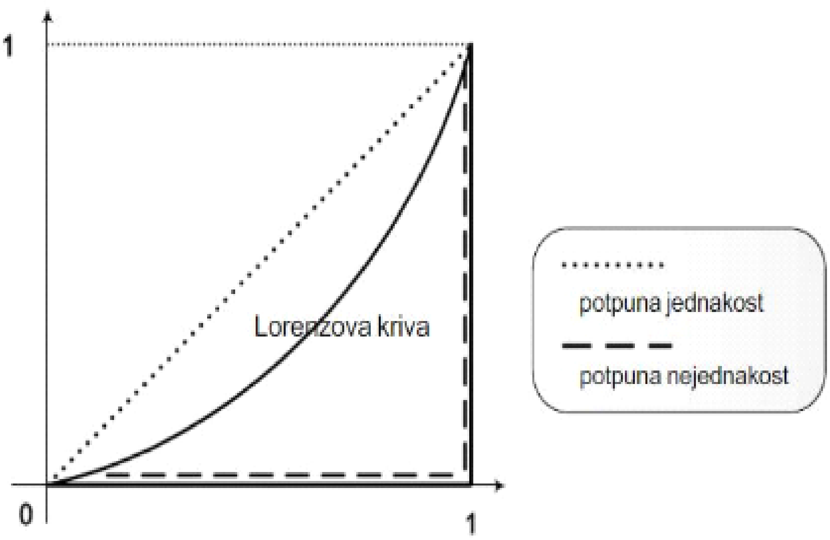
Лоренцова крива представља изломљену линију која полази од координатног почетка (0%,0%), спајајући одговарајуће тачке, и завршава се у тачки (100%,100%). То је са економског аспекта и логично, јер 0% становника присваја 0% националног дохотка, док цела популација, 100% становника, присваја цео доходак, односно свих 100% дохотка. Сада се додатно може прецизирати да је Лоренцова крива у потпуности садржана у троуглу чија су темена тачке (0,0), (1,1) и (0,1).

Екстремне вредности:

Сада ће бити представљене две крајње екстремне ситуације:

Потпуна једнакост — Пожељно је размотрити случај када би сваки појединац присвојио једнак доходак и како би крива тада изгледала. У том циљу, на *Слици 1* могу се уочити и тачке, $A'(20\%,20\%)$ и $B'(80\%,80\%)$. Повезивањем ових и сличних тачака, које такође имају исте вредности апсцисе и ординате, формира се линија **потпуне једнакости**, односно тражена Лоренцова крива. Јасно је да се у случају равномерне расподеле дохотка, ради о правој линији, под углом од 45^o, што представља дијагоналу квадрата у ком су тачке распоређене. Потпуну једнакост, дакле, означава дијагонална линија која представља хипотетичку расподелу дохотка у којој сви појединци имају једнаке дохотке.

Можемо рећи да се у реалним економским приликама никада нећемо срести са овим екстремним случајем. Светска економија је таква да се ствара све већа разлика између најбогатијих и најсиромашнијих слојева становништва. Са повећањем разлика у расподели дохотка, Лоренцова крива се савија испод линије једнакости. Што су диспаритети у дохоцима већи, то је крива више удаљена од линије равномерне расподеле.



Слика 2

Потпуна неједнакост — Супротан екстремни случај у расподели дохотка био би када би се појединцу придружио читав национални доходак, док други становници не присвајају ништа.

Потпуна неједнакост се такође не може очекивати у реалним околностима. Она заузима најнижу позицију на дијаграму, тј. ломи се под углом од 90^o и одређена је катетама раније поменутог троугла.

Да резимирамо претходна разматрања: Лоренцова крива представља графичку репрезентацију расподеле дохотка, приказујући за $p_1\%$ најсиромашнијих становника популације колики проценат $p_2\%$ од укупног дохотка присвајају. Ако је $p_1 = p_2$, Лоренцова крива је права линија која указује на потпуну једнакост у расподели дохотка. Свако закривљење испод линије 45^o представља различите облике **неједнакости у расподели**.

1.2.1 Лоренцова крива - математичка дефиниција и интерпретације

Јасно је да Лоренцова крива једна функција која описује начин расподеле укупног дохотка у одређеној популацији. Узимајући у обзир разматрања из претходног поглавља, може се формирати математичка дефиниција **емпиријске Лоренцове криве:**

Нека постоји узорак од n особа из дате популације и нека је обележје X_i приход особе i, $i = 1,2,3,\ldots,n$, тако да важи $X_1 \leq X_2 \leq \cdots \leq X_n$. Лоренцова крива узорка представља изломљену линију која редом спаја тачке $(h/n,\ L_h/L_n)$, $h = 0,1,2,\ldots,n$, где је $L_0 = 0$ и $L_h = \sum_{i=1}^{h} X_i$ укупни приход h најсиромашнијих особа. Отуда, Лоренцова крива има за своју апсцису кумулативни проценат становника који примају доходак, поређаних растуће, док за своју ординату има одговарајућу пропорцију примљених доходака.

1.2.2 Теоријска Лоренцова крива

Настанак Лоренцове криве проистекао је из великог интересовања економиста и статистичара за расподелу дохотка у популацији, па даље и неједнакостима у расподели дохотка. Касније је Лоренцова крива нашла примену у свим областима које проучавају расподеле, али се идеја примарно родила у расподели дохотка. Са тим у вези, нека се посматра одређена популација са обележјем X које представља, нпр. управо присвојени удео у националном дохотку. Претпоставља се да је X ненегативна променљива, сматрајући да је најмањи могући износ прихода једнак нули. Нека је $F(x) = p$ одговарајућа функција расподеле, чија је густина $f(x)$ и коначно математичко очекивање $\mu = E(X)$.

Теоријска дефиниција Лоренцове криве, за $p = F(x)$, се тада формира на следећи начин:

$$L(p) = \frac{\int_0^x tf(t)dt}{\int_0^{+\infty} tf(t)dt} = \frac{\int_0^x tf(t)dt}{\mu} = \frac{1}{\mu}\int_0^x tf(t)dt.$$

Популарнију нотацију за записивање Лоренцове криве увео је Гаствирт[2] 1971.године. Наиме, он је у обзир узео чињеницу да важи $x = F^{-1}(p)$. Како су учесници у расподели поређани растуће по приходима, а проценат присвојене добити кумулативно, функције расподеле прихода посматрамо као строго растуће и непрекидно-диференцијабилне функције. Стога је $x(F) = F^{-1}(p)$ добро дефинисано. Овим је омогућено да се Лоренцова крива запише на следећи начин, уводећи инверзну функцију $G = F^{-1}$:

$$L(p) = \frac{1}{\mu}\int_0^p G(t)dt = \frac{1}{\mu}\int_0^p F^{-1}(t)dt.$$

Алтернативна дефиниција Лоренцове криве постиже се применом релације $\mu = \int_0^1 G(t)dt$, дајући директну формулацију преко инверзне функције:

$$L(p) = \frac{\int_0^p G(t)dt}{\int_0^1 G(t)dt} = \frac{\int_0^p x(F)dF}{\int_0^1 x(F)dF}.$$

Може се десити да F^{-1} не постоји, јер функција расподеле има интервале константних вредности. У том случају, функција се може додефинисати на следећи начин:

$$F^{-1}(t) = \inf\{y : F(y) \geq F\}.$$

У интерпретацији преко инверзне функције, бројилац сумира дохотке најсиромашнијих $p\%$ популације, а именилац сумира доходак читаве популације. Када су појединци уређени растуће по вредностима присвојеног дохотка, Лоренцова крива $L(p)$ тако указује на пропорцију укупног дохотка присвојену од најсиромашнијих $p\%$ учесника популације.

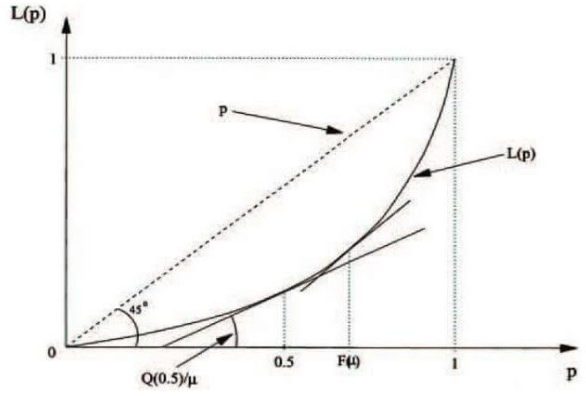
[2] Joseph L. Gastwirth

1.2.3 Особине Лоренцове криве

Неке особине Лоренцове криве за класу непрекидно-диференцијабилних функција расподеле са коначним математичким очекивањем:

1. У екстремном случају када сви чланови популације присвајају једнак приход, подела популације и подела укупног дохотка је идентична. Тада је Лоренцова крива еквирасподела, односно идентичка функција $F = L$, $L(p) = p$.
2. Лоренцова крива се дефинише за позитивне промењиве, са позитивним математичким очекивањем.
3. Лоренцова крива је потпуно садржана у јединичном квадрату, јер проценат p и *L(p)* узимају вредности из сегмента $[0, 1]$.
4. Ако је основна промењива позитивна и има густину, Лоренцова крива је непрекидна функција. $L(p) < p$, па је крива увек испод линије 45^o или се у случају потпуне једнакости са њом поклапа.
5. Лоренцова крива је функција са $L(0) = 0$ и $L(1) = \lim_{p \to 1-} L(p) = \lim_{x \to \infty} F(x) = 1$.
6. Лоренцова крива је инваријантна у односу на множење случајне промењиве позитивном константом, тј. X и cX имају исту Лоренцову криву, за $c > 0$.
7. Средња вредност дохотка у популацији се добија у проценту p за који важи да је нагиб функције *L(p)* једнак 1, односно где је $q(p) = \mu$, тј. у тачки *F(μ)*. Ово је лако доказати јер први извод Лоренцове криве износи управо $\frac{x}{\mu}$.
8. Медијана, као проценат средње вредности, дата је као нагиб Лоренцове криве у тачки где је $p = 0.5$. Како су многе расподеле дохотка накривљене удесно, средња вредност је често већа од медијане.
9. *L(p)* је растућа, конвексна функција од p, тј. коефицијент правца сваког одсечка Лоренцове криве већи је од коефицијента правца претходног одсечка. Наиме, први извод ове функције је $\frac{dL(p)}{dp} = \frac{q(p)}{\mu} = \frac{x}{\mu}$ где је $x = F^{-1}(p)$. Први извод је увек позитиван, јер је и доходак увек позитиван. Тако даље важи и за други извод, што указује на конвексност. Лоренцова крива је конвексна у p, па како p расте, нови дохоци који се кумулативно додају постају већи од оних који су до тада обрачунати. Конвексност Лоренцове криве доста открива о густини расподеле дохотка у различитим процентима. Што је већа густина расподеле *f(q(p))* у квантилу *q(p)*, то је мања конвексност Лоренцове криве у *L(p)*. На *Слици 3*, густина расподеле је видно већа за мање вредности процента p, како се управо у тим вредностима нагиб од *L(p)*

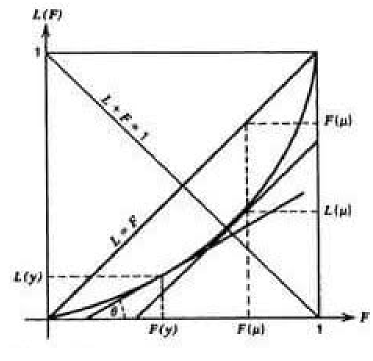
мења спорије уз пораст процента p.



Слика 3[5]

10. $\frac{E|X-\mu|}{2\mu} = F(\mu) - L(\mu) = \max_x[F(x) - L(x)]$.

Ова особина је јаснија помоћу *Слике 4*, узевши у обзир следеће:
- $L(y) = \int_0^y x dF(x)/E(Y)$,
- $F(\mu) - L(\mu) = \max_y[F(y) - L(y)]$,
- $F(\mu) + L(\mu) > 1, dL/dF|_{y=\mu} = 1$,
- $dL/d\mu = \tan\Theta = y/E(Y)$.



Слика 4[6]

Додатне особине Лоренцове криве могу се формулисати следећим двема теоремама.[5]

Теорема 1: *Претпоставимо да је L(p) дефинисана и непрекидна на* $[0,1]$ *са другим изводом* $L''(p)$. *Функција L(p) је Лоренцова крива ако и само ако*

$L(0) = 0, L(1) = 1, L'(0+) \geq 0, L''(p) \geq 0$ *на (0, 1).*

Ако је крива Лоренцова крива, она одређује расподелу случајне промењљиве X до фактора скалирања, што јесте средња вредност μ. Средња вредност се лако може израчунати, обрнутим поступком из формуле Лоренцове криве.

$$L(p) = \frac{1}{\mu} \int_0^p F^{-1}(t)dt,$$

што се може записати и као:

$$\mu L(F(x)) = \int_0^x ydF(y).$$

Овај израз се може диференцирати.

$$\mu L'(F(x))f(x) = xf(x).$$

Дељењем са *f(x)* и поновним диференцирањем, коначно се добија

$$\mu L''(F(x))f(x) = 1.$$

Из претходно наведеног следи и друга теорема.

Теорема 2: *Ако* $L''(p)$ *постоји и позитиван је свуда на интервалу* (x_1, x_2), *тада* F_X *има коначну позитивну густину расподеле на интервалу* $(\mu L'(x_1^+), \mu L'(x_2^-))$ *која је задата са*

$$f_X(x) = \frac{1}{\mu L''(F_X(x))}.$$

1.2.4 Лоренцове криве и графичко поређење неједнакости

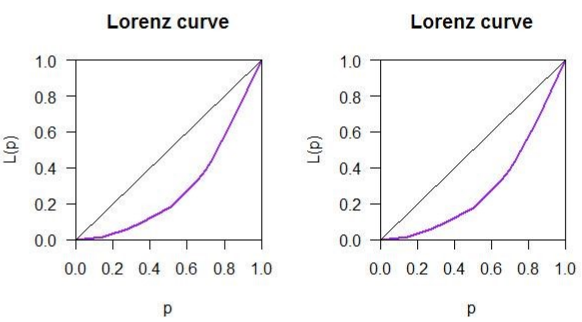
Могућност формирања и исцртавања Лоренцове криве даје простор за процену релативног степена неједнакости у одређеној расподели. Ипак, доста важније, даје простор за поређење две или више кривих. Познато својство Лоренцове криве јесте да свака расподела која има коначно очекивање, јединствено одређује своју Лоренцову криву до на трансформацију скалирања, што у овом случају и омогућава доношење закључака о расподелама, на основу упоређивања њихових Лоренцових кривих.

У том смислу, важе следећа правила:
– Ако једна Лоренцова крива лежи у потпуности изнад друге, расподела дохотка је равномернија у првом, него у другом случају.
– Ако се две Лоренцове криве поклапају, степен неједнакости је у оба случаја исти.
– Ако се две Лоренцове криве секу, потребне су додатне информације да би се извршило поређење неједнакости.

Пример 1:

Имајући у виду наведена правила и употребом статистичког софтвера R, може се графички приказати особина Лоренцове криве о непромењљивости при скалирању.

Случајни узорак v_1 из униформне расподеле и исти узорак помножен позитивним скаларом c, $v_2 = v_1 * c$, дају идентичну Лоренцову криву, што се може видети на графицима са *Слике 5.* Графици су резултат R-кода датог у *Додатку 1.* Функција која омогућава генерисање Лоренцове криве, зове се **Lc** функција из пакета **ineq**, узимајући за своје параметре векторе узорка и одговарајућих фреквенција.

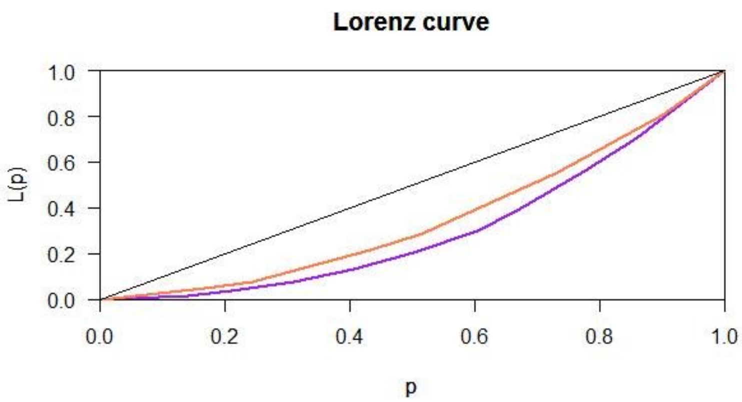


Слика 5

Како друго правило наводи, за случај када су две Лоренцове криве идентичне, тада је и степен неједнакости за оба случаја идентичан.

Пример 2:

Графичким поређењем Лоренцових кривих, може се илустровати додатна особина овог показатеља, а то је да је на подскупу већа неједнакост у расподели, гледано у односу на шири узорак из популације, или пак целу популацију. Пример илуструје Лоренцове криве униформне расподеле на датом сегменту $[0, 100]$ и његовом подскупу $[10, 30]$. Лоренцова крива која одговара ширем сегменту је свуда изнад Лоренцове криве подскупа, са највећом удаљеношћу од потпуне једнакости у тачки која одговара 60% становништва. R-код који генерише криве на сличан начин описан у *Примеру 1*, дат је у *Додатку 2*:



Слика 6

Приликом процене неједнакости, базиране на поређењу Лоренцових кривих, простор за дискусију и додатно истраживање се, јасно, јавља у трећем случају, када се Лоренцове криве секу. Чињеница је да се две Лоренцове криве могу сећи, као било које две криве у равни. То значи да свака крива у једном интервалу може лежати испод, а у следећем интервалу изнад неке друге криве. У овом случају се Лоренцов критеријум не може применити као критеријум за поређење. Стога су изграђени прецизнији, нумерички показатељи неједнакости, на основу којих се могу поредити све расподеле позитивног обележја, додељујући им одговарајуће нумеричке вредности. Један такав показатељ, директно изведен из Лоренцове криве, јесте *Ђинијев коефицијент*.

1.3 Ђинијев коефицијент

За Ђинијев коефицијент се може рећи да је најпознатији нумерички показатељ неједнакости. Њега је 1912. године увео Корадо Ђини[3], те је по њему и добио име. Сликовита графичка презентација чини ову меру лако разумљивом и допадљивом за статистичаре.

Веома корисна информација добијена из Лоренцове криве је њено одстојање, $p - L(p)$, од линије потпуне једнакости у расподели. Неједнакост у расподели одваја проценат $p - L(p)$ укупног дохотка од најсиромашнијих $p\%$ популације. Што је већа та разлика, то је већа неједнакост у расподели. Са повећањем степена неједнакости у расподели дохотка Лоренцова крива се све више удаљава од линије 45^o, а самим тим се повећава и површина која лежи између њих.

Отуда долази интерес за израчунавањем просечне дистанце Лоренцове криве од линије једнакости, па даље и израчунавањем површине између њих, као потенцијалног нумеричког показатеља неједнакости. Ђинијев коефицијент управо изражава однос између површине коју Лоренцова крива затвара са линијом једнакости и укупне површине троугла испод линије 45^o.

1.4 Формирање и смисао Ђинијевог коефицијента

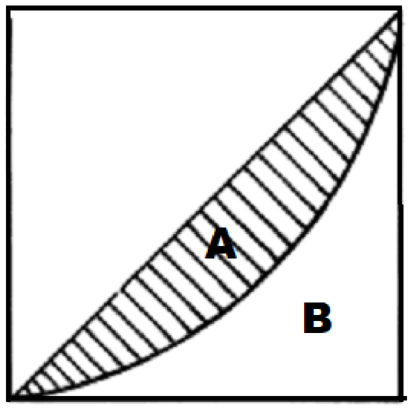
1.4.1 Интерпретација теоријског Ђинијевог коефицијента

Нека је X обележје популације, позитивна величина, нпр. присвојени удео у националном дохотку. Под претпоставком да X има функцију расподеле $F(x) = p$, и Лоренцову криву представљену функцијом *L(p)*, која је монотоно-растућа и интеграбилна на сегменту $[0, 1]$ и за коју важи: $L(0) = 0$, $L(1) = 1$, $0 \leq L(p_i) \leq p_i$ за свако $i \in [1, n]$, може се извести теоријска формула Ђинијевог кофицијента са следећим особинама:
– $G \in [0, 1]$,
– $G = 0$ акко $L(p) = p$, осим за коначно много $p \in [0, 1]$ за које је $0 \leq L(p) \leq p$,
– $G = 1$ акко $L(p) = 0$, осим за коначно много $p \in [0, 1]$ за које је $0 \leq L(p) \leq p$.

[3] Corrado Gini (1884-1965), италијански статистичар, демограф и социолог

1.4.2 Ђинијев коефицијент као површина

Ђинијев коефицијент је ненегативан реалан број, одређен односом површина на графику Лоренцове криве. Ако је са A означена површина између линије потпуне једнакости и Лоренцове криве, а са B површина испод Лоренцове криве, Ђинијев коефицијент се тада дефинише као $G = \frac{A}{A+B}$. Лоренцова крива је смештена унутар јединичног квадрата, па је површина доњег троугла $A + B = \frac{1}{2}$. Стога је Ђинијев коефицијент $G = 2A = 1 - 2B$.



Слика 7

Коефицијент тиме остаје садржан у вредностима између 0 и 1, а претходна дефиниција се може интерпретирати следећом интегралном једнакошћу:

$$G = 2\int_0^1 (p - L(p))dp = 1 - 2\int_0^1 L(p)dp.$$

Постоје многе формуле, тумачења и интерпретације Ђинијевог коефицијента, што је све у крајњој линији еквивалентно и даје кориснику слободу одабира.

Екстремне вредности:

Вредност коефицијента се теоријски креће у распону од 0 до 1. Што је Лоренцова линија закривљенија, то је Ђинијев коефицијент већи. Уколико Лоренцова линија узме екстремну вредност и поклопи се са хоризонталном осом од (0,0) до (100,0), па затим оде до (100,100), цео доходак присваја само један појединац и вредност Ђинијевог коефицијента је 1, што је потпуна неједнакост. Уколико се, пак, Лоренцова крива поклопи са линијом једнаких доходака, тада је вредност Ђинијевог коефицијента једнака 0. Претходно је разматрано да се границе остварују само у теорији, док се у реалним околностима не могу очекивати. Интервал коме припадају стварне величине овог показатеља је знатно ужи, обично између 0.20 и 0.50.

1.4.3 Ђинијев коефицијент као коваријација

Ђинијев коефицијент као коваријација се директно изводи из његове интегралне формуле, користећи парцијалну интеграцију са $du = 1$ и $v = L(p)$.

$$G = 1 - 2\int_0^1 L(p)dp = 1 - 2[pL(p)]_0^1 + 2\int_0^1 pL'(p)dp = -1 + 2\int_0^1 pL'(p)dp.$$

Увођењем смене промењливе $p = F(x)$ и применом особине $L'(p) = \frac{x}{\mu}$, добија се

$$G = \frac{2}{\mu}\int_0^\infty xF(x)f(x)dx - 1 = \frac{2}{\mu}[\int_0^\infty xF(x)f(x)dx - \frac{\mu}{2}].$$

Коначно се јавља могућност интерпретирања Ђинијевог коефицијента помоћу коваријације $Cov(x, F(x)) = E(xF(x)) - E(x)E(F(x))$. Одакле следи,

$$G = \frac{2}{\mu}Cov(x, F(x)).$$

Следи закључак да је Ђинијев коефицијент пропорционалан коваријацији између промењливе и њеног ранга у расподели дохотка.

1.4.4 Интерпретација емпиријског Ђинијевог коефицијента

За посматрану популацију, обележје X је ненегативна величина. Нека је $x_1, \dots, x_n$ низ случајних величина, такав да $x_i \geq 0$ за свако $0 \leq i \leq n$, и постоји барем једно i такво да је $x_i > 0$. Без умањења општости може се претпоставити и да је низ неопадајући, тј. $x_1 \leq x_2 \leq \dots \leq x_n$.

Ђинијев приступ мерењу неједнакости заснива се на поређењу појединачних парова доходака и сабирању апсолутних вредности измерених разлика. Тако се неједнакост за целу расподелу представља као збир разлика у паровима доходака:

$$\frac{1}{2n^2\bar{x}}\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|,$$

где су x_i и x_j дохоци i-тог и j-тог појединца, $\bar{x}$ је просечан доходак, а n укупан број прималаца дохотка.

Пошто се свака апсолутна разлика $|x_i - x_j|$ рачуна два пута (други пут као $|x_j - x_i|$, кад i и j замене места), цео израз се дели са 2. Уз то, да би се добио показатељ задовољавајућих својстава, сума апсолутних доходака се дели квадратом броја становника, тј. учесника у расподели (који одговара броју посматраних парова), као и просечним дохотком.

Формула показује да је Ђинијев коефицијент заправо једнак половини аритметичке средине апсолутних вредности разлика између свих парова доходака, посматране у односу на просечан доходак.

Пошто упоређује сваки пар доходака појединаца, Ђинијев коефицијент је врло директна мера неједнакости. Овај показатељ на синтетички начин изражава диспаритете у дохоцима преко целе расподеле.

1.4.5 Ђинијеве средње разлике

Претходно смо дефинисали емпиријски Ђинијев коефицијент као средњу вредност апсолутних разлика, подељену двоструком средњом вредношћу дохотка. Ако су x и y две случајне променљиве из исте расподеле F, из теоријске дефиниције следи

$$I_G = \frac{1}{2\mu}\int_0^\infty \int_0^\infty |x-y|dF(x)F(y).$$

Такође је познато да су *F(x)* и *1 – F(x)* проценти које становници индивидуално остварују у дохотку испод и изнад вредности x, па интеграција кроз све могуће вредности x даје поново Ђинијев коефицијент, у својој форми $\frac{1}{\mu}\int F(x)[1-F(x)]dx$.

Ако се извођењу приступа корак по корак, прво треба приметити $|x-y| = (x+y) - 2\min(x,y)$, па је очекивање ове апсолутне разлике $\Delta = E|x-y| = 2\mu - 2E(\min(x,y))$.

Да би се израчунало последње очекивање неопходна је расподела статистике минимума за 2 произвољне променљиве из исте расподеле. Оно што се може показати је да је тражено очекивање једнако $1-(1-F(y))^2$, док је извод овог очекивања $-2d(1-F(y))$. Одатле,

$$\Delta = 2\mu + 2\int_0^\infty yd(1-F(y))^2.$$

Последњи интеграл може се трансформисати парцијалном интеграцијом са $u = y$ и $v = (1-F(y))^2$:

$$\int_0^\infty yd(1-F(y))^2 = [y(1-F(y))^2]_0^\infty - \int [1-F(y)]^2 dy.$$

Коначно, како је $[y(1-F(y))^2]_0^\infty = 0$, добија се интегрална форма Ђинијевог коефицијента

$$I_G = \frac{\Delta}{2\mu} = 1 - \frac{1}{\mu}\int [1-F(x)]^2 dx.$$

Ова репрезентација показује да Ђинијев коефицијент мери *релативну неједнакост*, јер представља однос мере дисперзије, средње разлике, и средње вредности μ.

1.4.6 Примери израчунавања у R-у

Статистички софтвер R омогућава одређивање емпиријског Ђинијевог коефицијента за дате податке из истраживања. С тим у вези, омогућено је и генерисање одговарајуће емпиријске Лоренцове криве.

ineq је функција из истоименог пакета у R-у која служи за израчунавање мере неједнакости у вектору података.

Употреба:
ineq(x, parameter = NULL, type = c("Gini", "RS", "Atkinson", "Theil", "Kolm", "var", "square.var", "entropy"), na.rm = TRUE)
Gini(x, corr = FALSE, na.rm = TRUE)
RS(x, na.rm = TRUE)
Atkinson(x, parameter = 0.5, na.rm = TRUE)
Theil(x, parameter = 0, na.rm = TRUE)

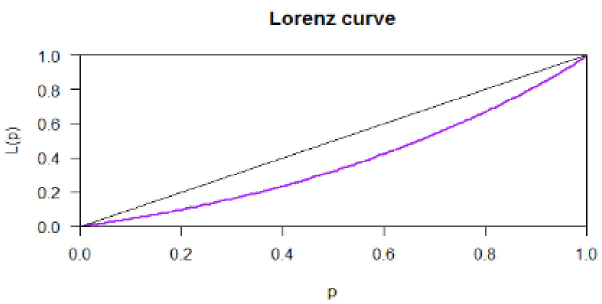
Аргументи:
x – вектор који садржи ненегативне вредности популације;

parameter – параметар као мера неједнакости (ако је *NULL* користи се подразумевани параметар одговарајуће мере);
type – стринг који представља меру која се користи за израчунавање неједнакости. Мора бити један од стрингова у подразумеваном аргументу (први карактер је довољан). Подразумевана вредност је "Gini".
corr – логичка вредност. Аргумент функције Gini прецизира да ли треба применити коначну корекцију узорка.
na.rm – логичка вредност која указује треба ли недостајуће вредности уклонити пре прорачуна.

Пример 3:

AirPassengers је база која садржи податке о месечном броју међународних авио путника, изражене у милионима, од 1949. до 1960., и приказана је у *Додатку 3.* Ова база се може искористити за анализу концентрације и одређивање мере неједнакости за посматрани период.

Израчунавање Ђинијевог коефицијента и генерисање Лоренцове криве, може се извршити претходно описаним функцијама *ineq* и *Lc*. Одговарајући *R*-код употребљен у примеру се може пронаћи у *Додатку 3.* Као резултат, добијен је податак о Ђинијевом коефицијенту од приближно 0.24, уз одговарајућу Лоренцову криву, приказану на *Слици 8.*



Слика 8

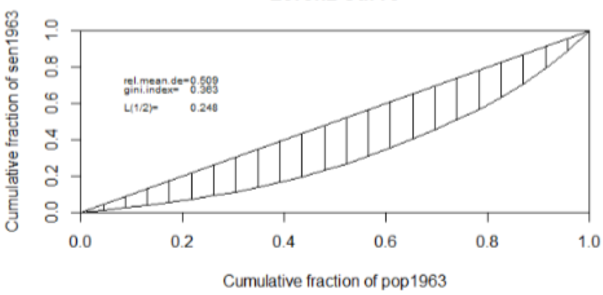
Коментар: Може се закључити да је у питању слаба неједнакост у расподели путника по месецима у години. Иако их највише путује у летњим месецима, Ђинијев коефицијент од 0.24 указује на прилично равномерну расподелу у посматраних 11 година.

Лоренцова крива и Ђинијев коефцијент могу се генерисати и позивом функције **lorenz.curve** из пакета **lawstat**:

Употреба:
lorenz.curve(data, weight=NULL, mul=FALSE, plot.it=TRUE, main=NULL, xlab=NULL, ylab=NULL, xlim=c(0,1), ylim=c(0,1),...)

Аргументи:
data – улазни подаци. Ако је аргумент низ, матрица, база података или листа, онда се прва колона третира као вектор са подацима, а друга колона као вектор вредности на у оси.
weight – Уколико је већ укључен у data аргумент, тј. ако постоји друга колона у data, онда се игнорише.
mul – логичка вредност која указује да ли се исцртава Лоренцова крива са више пондера.
plot.it – логичка вредност која говори да ли ће Лоренцова крива бити исцртана на екрану.
...
Пример 4:

data1963 је база R-овог пакета *lawstat* која садржи податке из 1963. године о односу броја сенатора и броја становника у 13 америчких округа. *lorenc.curve* функција генерише Лоренцову криву за посматрану базу података и на графику даје вредности:
Gini – Ђинијев коефицијент за улазне податке;
relative mean deviation – релативно средње одступање од улазних података;
L(1/2) – медијана кумулативног дела збира података;
> data(data1963)
> lorenz.curve(data1963)



Слика 9

1.5 Показатељи и позитивне расподеле

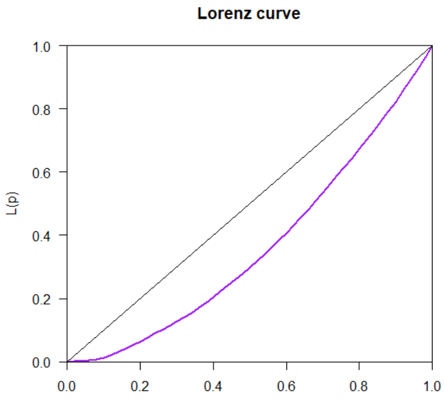
Томпсон, 1976. године, доказује наведено важно својство Лоренцове криве да свака расподела која има коначно математичко очекивање јединствено одређује своју Лоренцову криву до на трансформацију скалирања. Као показатељ изведен из Лоренцове криве, исто важи и за Ђинијев коефицијент. Нека је X ненегативна случајна променљива која представља обележје популације, нпр. присвојени удео у националном дохотку. Њена функција расподеле $F(x) = p$, има густину $f(x)$ и коначно математичко очекивање.

Претходно дефинисаним формулама показатеља неједнакости, а на основу директне везе са расподелом обележја, у овом поглављу биће приказани теоријска Лоренцова крива са графиком и Ђинијев коефицијент за неке позитивне расподеле.

Униформна расподела, $U[a,b]$: $f(x) = \frac{1}{b-a}, a \leq x \leq b,$

$$L_U(x) = \frac{\int_0^x tf(t)dt}{\int_0^1 tf(t)dt} = \frac{\int_0^x \frac{tdt}{b-a}}{\frac{a+b}{2}} = \frac{x^2 - a^2}{b^2 - a^2},$$

$$G_U = \frac{b-a}{3(b+a)}.$$

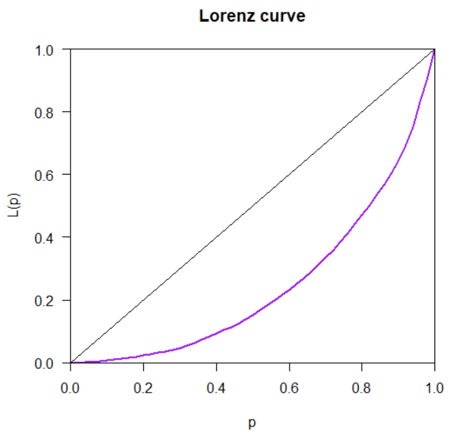


Слика 10

Експоненцијална расподела, $\mathcal{E}(\lambda)$: $f(x) = \lambda e^{-\lambda x}, x \geq 0,$

$$L_E(x) = 1 - e^{-\lambda x}(1 + \lambda x),$$

$$G_E(x) = \frac{1}{2}.$$



Слика 11

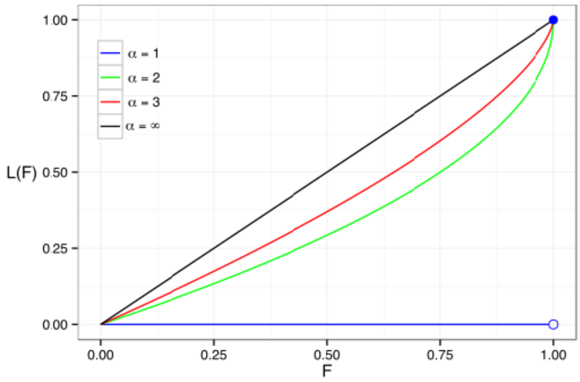
Паретова расподела: $f(x) = (\alpha - 1)(x+1)^{-\alpha}, 0 < x,$

$$L_P(x) = 1 - (1-x)^{1-\frac{1}{\alpha}},$$

$$G_P(x) = \frac{\alpha - 1}{2\alpha - 3}.$$

Код Паретове расподеле захтева се да важи да је $\alpha > 2$, како би Ђинијев коефицијент био употребљив податак. Може се приметити да што је α ближе броју 2, то је Ђинијев коефицијент ближи јединици, а самим тим неједнакост је већа. За $\alpha = 2$ Лоренцова крива одговара потпуној неједнакости, са $G = 1$. Ова чињеница наново показује директну везу између параметара функција расподела са Лоренцовом кривом и Ђинијевим коефицијентом, конкретно указујући да Паретове расподеле уз што већи параметар α имају и већу једанкост.

Такође важи, када $\alpha \to \infty \Rightarrow G \to \frac{1}{2}$. То се објашњава чињеницом да са повећањем параметра α Паретова расподела конвергира ка експоненционалној.



Слика 12

Бета расподела $B(\alpha,\beta)$: $F(x)=\int_0^x \frac{t^{\alpha-1}(1-t)^{\beta-1}dt}{B(\alpha,\beta)}$,

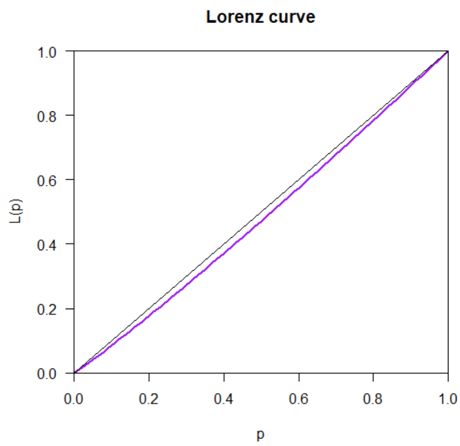
$$f(x)=\frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}, 0<t<1.$$

Може се доказати да важе следеће једнакости:

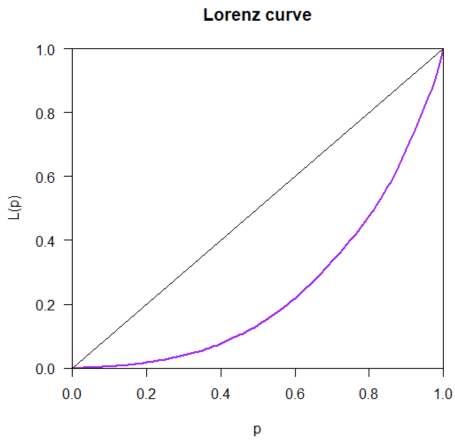
$$L_B(x)=F_{\alpha+1,\beta}(x)=\int_0^x \frac{t^{\alpha}(1-t)^{\beta-1}dt}{B(\alpha+1,\beta)},$$

$$G_B(x)=\frac{B(2\alpha,2\beta)}{\alpha(B(\alpha,\beta))^2}.$$

Лоренцове криве за $B(n,1)$ и $B(1,n)$, дате су редом графицима:



Слика 13 - $B(n,1)$



Слика 14 - $B(1, n)$

Користећи се особином Бета функције $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$, лако се доказује за:
- $B(n, 1)$, Ђинијев коефицијент од $G = \frac{1}{2n+1}$,
- $B(1, n)$, Ђинијев коефицијент од $G = \frac{n}{2n+1}$.

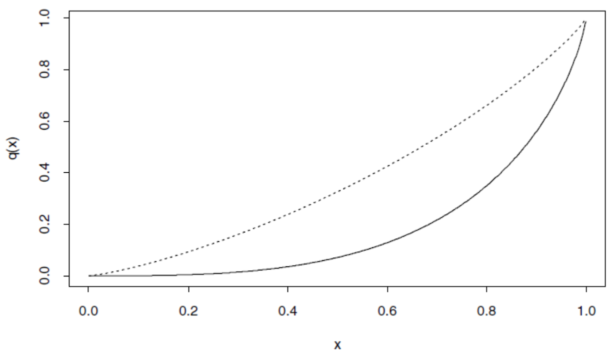
Слично Паретовој расподели, Бета расподела са параметрима $\alpha = 1$ и $\beta = n$ конвергира ка експоненционалној расподели при условима: $n \to \infty$ и $G \downarrow \frac{1}{2}$.

Гама расподела $G(\alpha)$: $f(x) = \frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)}, x \geq 0$

Након компликованијег извођења, добија се следећа формула Ђинијевог коефицијента за једнопараметарску Гама расподелу:

$$G_G = \frac{1}{4^{\alpha}\alpha B(\alpha, \alpha+1)}.$$

И сама Лоренцова крива одређена је вредностима које узима α, па је могуће пронаћи и зависност квантила од Ђинијевог коефицијента. Примера ради, Лоренцова крива Гама расподеле са параметрима **0.5** и **5** дата је на графику који следи. Лоренцова крива Гама расподеле са параметром $\alpha = 0.5$ представљена је пуном линијом, док је за $\alpha = 5$, испрекиданом.



Слика 15

Ђинијев коефицијент Гама расподеле са параметром **0.5** износи **0.64**, док за Гама расподелу са параметром **5**, Ђинијев коефицијент износи **0.25**. Ако параметар Гама расподеле тежи бесконачно, Ђинијев коефицијент тежи нули, одакле се може закључити да за што је већи параметар Гама расподеле, мања је неједнакост у расподели, поново доказујући узрочно-последичну везу између расподеле и њој одговарајућег показатеља неједнакости.

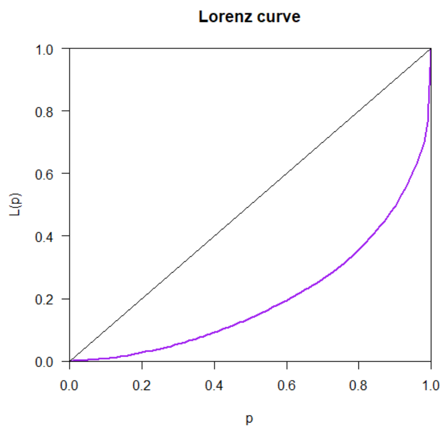
Лог-нормална расподела $LN(\mu, \sigma^2)$:

Густина: $f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$,

Математичко очекивање: $\mu = \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} dx = e^{\mu + \frac{\sigma^2}{2}}$,

Лоренцова крива: $L_{LN}(x) = \Phi(\frac{\ln x - \mu - \sigma^2}{\sigma})$,

Ђинијев коефицијент: $G_{LN} = 2\Phi(\frac{\sigma}{\sqrt{2}}) - 1$.

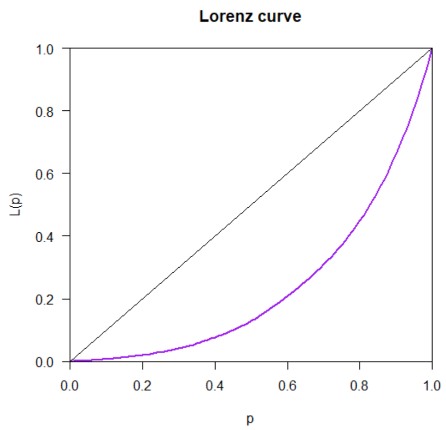


Слика 16

Вејбулова расподела:

$f(x) = \frac{k}{\lambda}(\frac{x}{\lambda})^{k-1}e^{-(\frac{x}{\lambda})^k}$,

Ђинијев коефицијент: $G = 1 - 2^{-\frac{1}{k}}$.



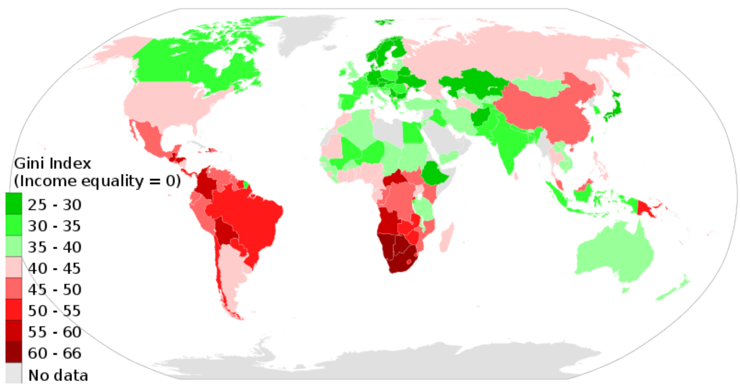
Слика 17

1.6 Значај, предности и недостаци

Ђинијев коефицијент налази широку примену у разним областима као што су социологија, економија, медицина, екологија, агрикултура и друге. Иако се најчешће употребљава у економији, Ђинијев коефицијент се може применити у било којој науци која проучава расподеле.

Ђинијев коефицијент у образовању оцењује неједнакости у степену образовања одређене популације. На основу научне студије из 1990.године, чији је предмет било 85 земаља, добијена је информација да је држава Мали имала највиши Ђинијев коефицијент образовања у износу од 0.92 што је указивало на јако високу неједнакост у степену образовања. У истој студији Сједињене Америчке Државе оствариле су најмањи Ђинијев коефицијент од 0.14. У периоду између 1960. и 1990. године, Северна Кореја, Кина и Индија забележиле су најбржи пад Ђинијевог коефицијента неједнакости у образовању. Такође, истраживања су показала да се Ђинијев коефицијент у Сједињеним Америчким Државама благо повећао у периоду од 1980. до 1990. године. [11]

У екологији се Ђинијев коефицијент користи за мерење биодиверзитета (биолошке разноврсности). У здравству се може применити као мера неједнакости здравља у зависности од квалитета живота популације. Налази примену и у хемији, изражавајући селективност протеина киназе. Области примене су, дакле, небројене.



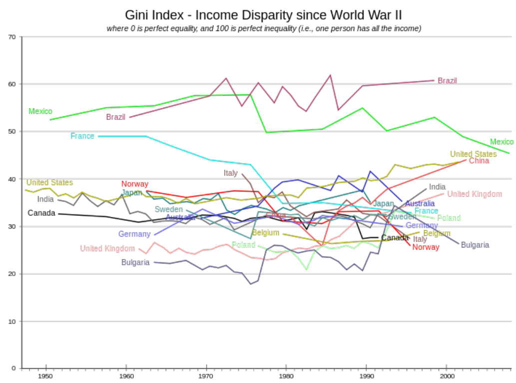
Слика 18 - Ђинијев коефицијент расподеле дохотка у свету [11]

Ђинијев коефицијент расподеле богатства у свету:

Посматрано на нивоу света, Ђинијев коефицијент се креће у интервалу доње границе до око 25% (скандинавске и неке средњеевропске државе) и горње границе до око 60% (углавном афричке државе и Латинска Америка). Према званичним информацијама Светске банке, оцењени Ђинијеви коефицијенти варирају од 0.195 (Словачка) до 0.629 (Сиера Леоне). Израчунат Ђинијев коефицијент за Канаду је 0.315, док је за САД то вредност 0.401. Државе у Африци имају највише Ђинијеве коефицијенте, са Јужноафричком републиком на првом месту са 70%. Вредност Ђинијевог коефицијента Србије налази се на нивоу просека држава региона.[11]

На основу података из 2014. године о Ђинијевом коефицијенту као мери неједнакости у расподели доходака различитих земаља, на *Слици 18* дат је визуелни приказ о стању у свету. Црвеном бојом су обележене државе са високом, а зеленом бојом са ниском неједнакошћу у расподели дохотка. Ђинијев коефицијент читавог света у целини, оцењује се на основу различитих становишта и износи између 0.61 и 0.68.

График на *Слици 19* приказује вредности и кретање Ђинијевог коефицијента одређених земаља, почев од завршетка Другог светског рата.



Слика 19 - Ђинијев коефицијент након Другог светског рата [11]

Предности и ограничења:

Мерење неједнакости у расподели је сложен статистички захват, па је сигурно да не постоји потпуно задовољавајући показатељ неједнакости.

Корисне карактеристике Ђинијевог коефицијента су:

1. једноставна интерпретација – најзаслужнија чињеница за популарност овог показатеља;
2. анонимност – небитно је ко има ниске, а ко високе дохотке;
3. скала независности – Ђинијев коефицијент не узима у обзир, на пример, да ли је у питању просечно сиромашна или богата држава;

4. независност становништва – неважно је колика је популација одређене државе;
5. принцип трансфера – уколико се доходак пренесе са богате на сиромашну особу повећава се равномерност расподеле.

Приписују му се и извесни недостаци:
1. Осетљив је на све сегменте расподеле дохотка подједнако, односно не придаје већу важност доњем делу доходне лествице, на коме се налазе сиромашнији појединци;
2. Није адитиван, тј. збир Ђинијевих коефицијената подгрупа становништва није једнак коефицијенту за целу популацију;
3. Игнорише промене у друштву као што су раст популације и промене популације услед емиграције и имиграције;
4. Различите расподеле могу имати исти Ђинијев коефицијент, што је последица чињенице да се две Лоренцове криве могу сећи. У таквој ситуацији, свака расподела дохотка (представљена одговарајућом Лоренцовом кривом), у неком сегменту има већи степен неједнакости од друге расподеле са којом се пореди;
5. Не распознаје ефекте структурних промена у популацији, тј. игнорише промене у животном веку у популацији. Људи у младости имају ниже приходе и богатсво него у свом старијем добу, па пропорција младог и старог становништва значајно утиче на економске неједнакости.

1.6.1 Неке алтернативне мере неједнакости

Узимајући у обзир ограничења овог коефицијента, треба поменути и неке статистичке методе које омогућавају уградњу пристрасности у корист сиромашнијих или неких других појединаца. Неизбежно је у том случају поменути Тилов и Аткинсонов индекс као мере ентропије.

Тилов индекс

$$T = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i}{\bar{x}}\ln\frac{x_i}{\bar{x}}\right)$$

Ако сваки појединац присваја једнак доходак, тада је Тилов коефицијент једнак 0. Са друге стране, ако појединац присваја читав доходак, тада коефицијент износи $\ln N$. Дељење коефицијента са $\ln N$ нормализује вредности на распон од 0 до 1.

Аткинсонов индекс

За $\mathcal{E} = 1$: $A_{\mathcal{E}}(y_1, \ldots, y_N) = 1 - \frac{1}{\mu}(\prod_{i=1}^{N} y_i)^{\frac{1}{N}}$

За $0 \leq \mathcal{E} \neq 1$: $A_{\mathcal{E}}(y_1, \ldots, y_N) = 1 - \frac{1}{\mu}(\frac{1}{N} \prod_{i=1}^{N} y_i^{1-\mathcal{E}})^{\frac{1}{1-\mathcal{E}}}$

Када је Аткинсонов параметар $\mathcal{E}$ једнак нули и Аткинсонов индекс износи 0, указујући на одсуство неједнакости у расподели. За $\mathcal{E} = \infty$, Аткинсонов индекс биће 1 и сматра се да је извесна велика неједнакост.

Поглавље 2

Оцене показатеља неједнакости и тестови

Чест циљ статистичког истраживања јесте оцењивање параметара који зависе од функције расподеле испитиваних података. Такви се параметри могу назвати функционалним параметрима. На пример, математичко очекивање случајне променљиве са функцијом расподеле F, $\mu(F) = \int_{-\infty}^{\infty} x dF(x)$.

Лоренцова крива, а тиме и Ђинијев коефицијент, представљају нешто сложеније функционалне параметре, о чијем је значају и применама у друштву речено у претходном поглављу. Очигледно је да је вредност *L(F)* у зависности од функције расподеле F, па се стога и може размишљати о *L(F)* као о реалном функционалу на простору функција расподела. Аналогно важи за Ђинијев коефицијент.

2.1 Оцена Лоренцове криве и тест сагласности с експоненцијалном расподелом

Као што је претходно дефинисано, Лоренцова крива јесте функционални параметар у зависности од функције расподеле F. За функцију расподеле F која је строго растућа, где је G њен јединствени инверз, теоријска Лоренцова крива дефинисана је са

$$L(p) = \mu^{-1} \int_0^{G(p)} x dF(x) = \mu^{-1} \int_0^p G(t) dt = \frac{\nu(p)}{\mu}.$$

Под претпоставком да је $X_1, \dots, X_n$ низ позитивних случајних промењљивих са статистикама поретка $X_{(1)}, \dots, X_{(n)}$, нека је $r = [np]$ највећи цео број мањи или једнак од np. Тада је **оцена Лоренцове криве** базирана на узорку дефинисана са

$$L_n(p) = \frac{\sum_{i=1}^{r=[np]} X_{(i)}}{\sum_{i=1}^{n} X_{(i)}}.$$

Како свака расподела са коначним очекивањем јединствено одређује своју Лоренцову криву и обрнуто, природно се намеће идеја да се оцена Лоренцове криве може искористити за тест сагласности с одређеном расподелом.

Тест заснован на овој идеји објављен је у [2], са **нултом хипотезом**:

$$H_0 : F(x) = 1 - e^{-\lambda x}, x \geq 0$$

и $\lambda > 0$ непознатим параметаром. За дату претпоставку о функцији расподеле, теоријска Лоренцова крива биће

$$L(p) = \lambda \int_0^p [-\lambda^{-1} \ln(1-t)] dt = p + (1-p) \ln(1-p),$$

а **тест статистика**, оцена Лоренцове криве:

$$L_n(p) = \frac{\sum_{i=1}^{r=[np]} X_{(i)}}{\sum_{i=1}^{n} X_{(i)}}.$$

У свом раду([2]), Гејл испитује различита својства и особине ове оцене, узимајући у обзир различите вредности p и доказује да се најбољи резултати постижу управо за $p = 0.5$.

Расподела за $L_n(p)$ под H_0: F је функција експоненцијалне расподеле

При хипотези H_0, да је $F(x) = 1 - e^{-\lambda x}$, за $x \geq 0$, могуће је одредити расподелу тест статистике $L_n(p)$. Конкретно за $p = 0.5$, $L_n(0.5)$ као најбоља могућа оцена има следећу расподелу, при услову $0 < l < \frac{r}{n}$:

$$P\{L_n(0.5) \leq l\} = 1 - \sum_{j=0}^{\psi} \frac{n!(r-j-l(n-j))^{n-1}(-1)^j}{(n-j)(n-r)^{r-1}(r-j)^{n-r-1} j!(r-j)!(n-r)!},$$

где је $r = [np]$ и $\psi = \frac{r-ln}{1-l}$. На основу табличних вредности (*Табела 1*), доноси се одлука о прихватању хипотезе.

		Probability level							
n	[.5*n*]	*0.01*	*0.025*	*0.05*	*0.1*	*0.9*	*0.95*	*0.975*	*0.99*
2	1	0.00500	0.01250	0.02500	0.05000	0.45000	0.47500	0.48750	0.49500
3	1	0.00167	0.00419	0.00844	0.01711	0.22792	0.25880	0.28063	0.29999
4	2	0.02405	0.03853	0.05532	0.08007	0.33909	0.37228	0.39862	0.42529
5	2	0.01329	0.02140	0.03093	0.04522	0.22389	0.25198	0.27553	0.30103
6	3	0.03902	0.05422	0.07008	0.09152	0.29618	0.32504	0.34869	0.37439
7	3	0.02600	0.03630	0.04716	0.06204	0.21944	0.24438	0.26548	0.28906
8	4	0.04950	0.06420	0.07880	0.09779	0.27185	0.29742	0.31888	0.34274
9	4	0.03640	0.04739	0.05841	0.07289	0.21501	0.23751	0.25677	0.27860
10	5	0.05725	0.07124	0.08477	0.10203	0.25606	0.27914	0.29873	0.32080
11	5	0.04472	0.05584	0.06668	0.08060	0.21117	0.23179	0.24955	0.26984
12	6	0.06326	0.07659	0.08927	0.10522	0.24487	0.26599	0.28406	0.30463
13	6	0.05149	0.06252	0.07308	0.08644	0.20789	0.22702	0.24356	0.26257
14	7	0.06814	0.08088	0.09286	0.10779	0.23644	0.25601	0.27283	0.29208
15	7	0.05713	0.06798	0.07823	0.09108	0.20509	0.22299	0.23850	0.25641
16	8	0.07221	0.08444	0.09584	0.10993	0.22984	0.24812	0.26390	0.28204
17	8	0.06191	0.07254	0.08250	0.09487	0.20266	0.21953	0.23419	0.25114
18	9	0.07570	0.08746	0.09836	0.11176	0.22449	0.24171	0.25659	0.27377
19	9	0.06604	0.07644	0.08611	0.09805	0.20054	0.21653	0.23045	0.24658
20	10	0.07873	0.09009	0.10055	0.11336	0.22006	0.23636	0.25049	0.26685
21	10	0.06964	0.07982	0.08922	0.10077	0.19867	0.21390	0.22717	0.24259
22	11	0.08140	0.09240	0.10248	0.11476	0.21632	0.23183	0.24530	0.26090
23	11	0.07283	0.08279	0.09193	0.10312	0.19700	0.21157	0.22428	0.23905
24	12	0.08378	0.09445	0.10419	0.11602	0.21310	0.22793	0.24081	0.25576
25	12	0.07567	0.08542	0.09433	0.10519	0.19550	0.20949	0.22169	0.23590
26	13	0.08594	0.09630	0.10573	0.11715	0.21030	0.22452	0.23689	0.25126
27	13	0.07823	0.08777	0.09647	0.10703	0.19415	0.20761	0.21937	0.23306
28	14	0.08788	0.09797	0.10712	0.11818	0.20784	0.22152	0.23343	0.24727
29	14	0.08055	0.08990	0.09839	0.10867	0.19292	0.20591	0.21726	0.23049
30	15	0.08966	0.09949	0.10838	0.11911	0.20566	0.21884	0.23034	0.24371
31	15	0.08267	0.09183	0.10013	0.11015	0.19180	0.20436	0.21534	0.22815
32	16	0.09128	0.10088	0.10955	0.11997	0.20370	0.21645	0.22756	0.24051
33	16	0.08461	0.09359	0.10171	0.11150	0.19076	0.20294	0.21359	0.22601
34	17	0.09279	0.10217	0.11062	0.12077	0.20193	0.21428	0.22506	0.23761
35	17	0.08640	0.09522	0.10316	0.11272	0.18981	0.20164	0.21198	0.22405
36	18	0.09418	0.10336	0.11161	0.12151	0.20033	0.21232	0.22278	0.23497
37	18	0.08805	0.09671	0.10449	0.11385	0.18893	0.20043	0.21048	0.22223
38	19	0.09548	0.10447	0.11253	0.12219	0.19887	0.21052	0.22069	0.23254
39	19	0.08959	0.09809	0.10573	0.11489	0.18810	0.19930	0.20909	0.22054
40	20	0.09669	0.10550	0.11339	0.12283	0.19753	0.20887	0.21877	0.23032

Табела 1 [2]

Пример 5:

Нека узорак {1,2,3,4,5,6,7,11,14,18,22,28,33,35,36,39,41,42,51,64} даје податке о дужини живота у месецима након дијагностиковане болести, за 20 случајно одабраних пацијената оболелих од рака. Тестира се претпоставка да случајна величина X, која представља дужину живота након дијагностиковања болести, има експоненцијалну расподелу. Како се из графичке репрезентације не може доћи до поузданог закључка, може се прибећи тесту сагласности помоћу оцене Лоренцове криве на датом узорку, $L_{20}(0.5)$.

$$L_{20}(0.5) = \frac{\sum_{i=1}^{r=[10]} X_{(i)}}{\sum_{i=1}^{20} X_{(i)}} = \frac{71}{462} = 0.1319$$

За $\alpha = 0.05$, помоћу таблице посматра се критична вредност. На основу тога се може донети закључак да су подаци сагласни са хипотезом о експоненцијалној расподели, тј. H_0 се прихвата.

Пример 6:

Узмимо у разматрање случајни узорак из гама расподеле $\gamma(3.1)$, обима 20. Нека је H_0: $F(x) = 1 - e^{-\lambda x}$, тј. претпоставља се да је расподела експоненцијална.

$X = \{0.47, 0.66, 0.80, 1.11, 1.14, 1.24, 1.25, 1.28, 1.34, 1.40, 1.48, 1.49, 1.54, 1.57, 1.73, 1.74, 2.16, 3.76, 4.37, 5.11\}$

Оцена Лоренцове криве овог узорка рачуна се по истом принципу.

$$L_{20}(0.5) = \frac{\sum_{i=1}^{r=[10]} X_{(i)}}{\sum_{i=1}^{20} X_{(i)}} = \frac{10.69}{35.64} = 0.2999,$$

што је веће од одговарајуће критичне вредности 0,23636. H_0 се одбацује, и закључујемо да и без претходног знања да су подаци из гама расподеле, може се одбацити идеја о експоненцијалној расподели.

Монте Карло оцена емпиријске мере и моћи:

Посматрано на 10 000 случајно одабраних узорака, може се одредити емпиријска мера и моћ Лоренцове статистике за тест сагласности с експоненцијалном расподелом.

За одређивање емпиријске мере, 10 000 узорака обима 20 генерисано је из експоненцијалне расподеле. При прагу значајности $\alpha = 0.05$, а помоћу R-кода приказаног у *Додатку 5*, добија се пропорција узорака при којима се H_0 одбацује. Добијени проценат износи 0.0454, дајући грешку прве врсте и указујући да у случајно генерисаних 10 000 узорака из експоненцијалне расподеле, 454 одбацују хипотезу експоненцијалности, тестом спроведеним преко Лоренцове статистике.

С друге стране, за емпиријску моћ посматра се 10 000 узорака из $\gamma(3,1)$ расподеле у односу на H_1 претпоставку. Добијена емпиријска моћ теста у односу на алтернативну хипотезу, при датим условима, износи 0.7139.

2.2 Оцена Ђинијевог коефицијента

Када је функција расподеле F позната може се доћи до теоријске вредности Ђинијевог коефицијента. То углавном није случај или је сам интеграл компликован за извођење. Приступ који се чешће спроводи у мерењу неједнакости, јесте оцена Ђинијевог коефицијента на одговарајућем узорку из популације.

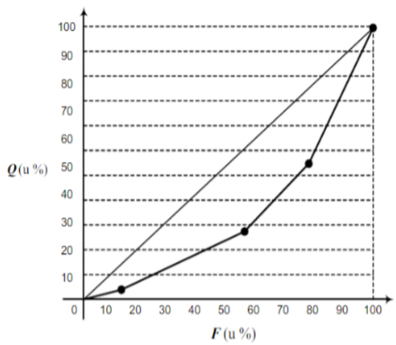
Оцена: Уобичајена метода оцене базира се на идеји да су подаци груписани у одговарајуће доходне класе, сегменте, а средња вредност дохотка у сваком сегменту је позната. Ако се за узорак X посматрају средње вредности по препознатим сегментима, добија се оцена Ђинијевог коефицијента за основну функцију расподеле - **уопштена оцена Ђинијевих средњих разлика**[1].

Ако је f_j фреквенција интервала чија је средња вредност (центар) x_{C_j}: кумулативне фреквенције F_j, на апсциси, и кумулативне глобалне вредности Q_j, на ординати, одређене су на следећи начин:

$$F_j = \sum_{x \leq x_j} p_j; Q_j = \sum_{x \leq x_j} q_j;$$

$$p_j = \frac{f_j}{n}; q_j = \frac{x_{C_j} f_j}{\sum_{j=1}^{n} x_{C_j} f_j};$$

$$\sum_{j=1}^{n} p_j = \sum_{j=1}^{n} q_j = 1$$



Слика 18

Примена статистичког софтвера R при оцени Ђинијевог коефицијента и формирања одговарајуће Лоренцове криве узорка: Оцена Ђинијевог коефицијента може се добити помоћу функције **gini** из пакета **reldist**.

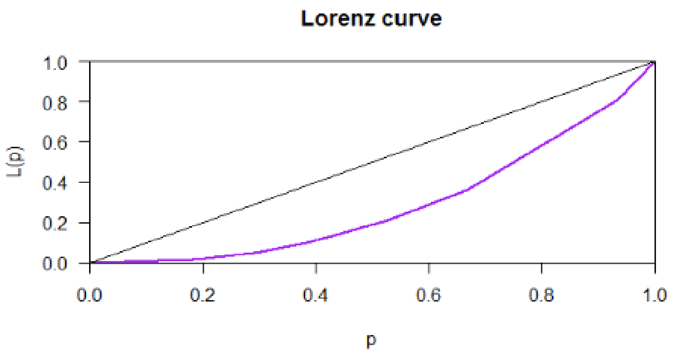
[1] Gini mean difference

Употреба:
gini(x, weights=rep(1,length=length(x)))

Аргументи:
x – вектор који садржи ненегативне елеменате;
weights – опциони вектор чије вредности се наносе на y осу;

Пример 7:
Нека је дата популација од 22 269 становника, за коју је потребно одредити степен неједнакости у расподели дохотка. Из тог разлога, може се прибећи оцени Ђинијевог коефицијента на основу узорка. Нека вектор X позитивних вредности, садржи редом 10 пресека становништва, посматрано од најсиромашнијих ка најбогатијим.
$X = \{548, 1763, 2495, 3438, 4257, 5481, 6392, 8304, 12904, 22269\}$

Претпоставимо да узорсак Y из униформне расподеле даје податке о одговарајућим присвојеним дохоцима, који се налазе на y-оси Ђинијевог квадрата.



Слика 19

Оцена Ђинијевог коефицијента у овом примеру износи 42%, што указује на приличан степен неједнакости у расподели дохотка за дати узорак униформне расподеле, чији теоријски коефицијент износи $\frac{1}{3}$. R-код употребљен за одређивање ове оцене, и генерисање одговарајуће Лоренцове криве (*Слика 19*), приказан је у *Додатку 4*.

У наставку ће бити описане неке напредније методе оцењивања Ђинијевог коефицијента и примена тих оцена.

2.3 Оцена Ђинијевог коефицијента базирана на средњим вредностима

Претходно је илустровано уопштење ове оцене преко средњих вредности одабраних сегмената. У овом поглављу биће детаљније описан поступак оцене и дефинисање граничних вредности Ђинијевог коефицијента.

Нека је $L(p)$ Лоренцова крива популације. Стандардна метода оцењивања Ђинијевог коефицијента базира се на апроксимацији поља концентрације - А (између линије једнакости и Лоренцове криве), што се врши одабиром k сегмената одређених тачкама: $0 = p_0 < p_1 < \cdots < p_k < p_{k+1} = 1$ и анализом изломљене линије задате редом тачкама $(0,0),(p_1,L(p_1)),\ldots,(p_k,L(p_k)),(1,1)$ и поља које обухвата.

Ова процедура дефинише доњу границу за површину А, али и Ђинијев коефицијент G јер права линија која повезује $(p_i,L(p_i))$ и $(p_{i+1},L(p_{i+1}))$ лежи изнад $L(p)$. Стога, доња граница за G биће: $G \geq 1 - \sum_{i=0}^{k}(p_{i+1}-p_i)[L(p_i)+L(p_{i+1})]$.

Ради утврђивања тачности претходне границе потребна је и горња граница од G. За ову потребу може се употребити интерпретација Ђинијевог коефицијента преко средњих разлика $G = \frac{\Delta}{2\mu}$. Нека је дато n бројева поређаних растуће и груписаних у $(k+1)$ подгрупа $x_1,\ldots,x_{m_1};x_{m_1+1},\ldots,x_{m_2};\ldots;x_{m_k+1},\ldots,x_n$, где су $m_1 = np_1, m_2 = np_2,\ldots,m_k = np_k$ и $0 < p_1 < \cdots < p_k < p_{k+1} = 1$. Тада је оцена за суму средњих разлика Δ:

$$\Delta^* = (\frac{n}{2})^{-1}\sum_{i<j}\sum_{i<j}\|x_i - x_j\| = \sum_{i\neq j}\sum_{i\neq j}\gamma_i\gamma_j\|\mu_i - \mu_j\| + \sum_{i=1}^{k+1}\gamma_i^2\Delta_i^*,$$

где μ_i представља средњу вредност i-те подгрупе, Δ_i^* такође одговара i-тој групи, а γ_i представља пропорцију опсервација захваћених i-том групом (нпр. $\gamma_1 = p_1$, $\gamma_2 = p_2 - p_1$, ... , $\gamma_{k+1} = 1 - p_k$).

Занемарујући разлике у дохоцима унутар група, Ђинијев коефицијент је ограничен са горње стране статистиком која се назива "корекција груписања": $G \leq (2\mu)^{-1}\sum\gamma_i^2\Delta_i^*$. [4]

Овим је омогућена оцена Ђинијевог коефицијента без познавања расподеле популације. Без обзира која је густина расподеле, Ђинијев коефицијент оцењен на основу узорка имаће вредност у датим границама. У свом раду из 1972.,Гаствирт показује да разлика између овако дефинисаних граница износи око 0.006 чиме добијамо прецизну оцену.

2.4 Оцена Ђинијевог коефицијента методом замене

Нека је $X_1, \dots, X_n$ низ независних једнако-расподељених случајних променљивих са функцијом расподеле F и нека је G Ђинијев коефицијент, функционални параметар у зависности од F.

Веза функције расподеле са одговарајућим Ђинијевим коефицијентом G, указује на могућност оцене овог параметра тако што ће се пронаћи одговарајућа оцена за F, $\widehat{F}$, а затим том оценом заменити F у формули параметра и дати његову оцену: $\widehat{G} = G(\widehat{F})$.

За оцену функције расподеле F најчешће се узима емпиријска функција расподеле тог узорка $\widehat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x)$. У случају када емпиријска функција не производи задовољавајућу оцену, користи се "глаткија" оцена:

$$\tilde{F}(x) = \int_{-\infty}^{\infty} R(x-y)d\widehat{F}(y) = \frac{1}{n}\sum_{i=1}^{n} R(x - X_i),$$

где је R непрекидна функција расподеле концентрисана блиско око нуле. Оцене $\widehat{F}$ и $\tilde{F}$ су јако блиске у смислу да је $|\widehat{F} - \tilde{F}|$ јако мало.

У овом случају нема потребе за напреднијом оценом од емпиријске функције, па се према њој формира одговарајућа оцена Ђинијевог коефицијента методом замене:

$$\widehat{G} = G(\widehat{F_n}) = (\sum_{i=1}^{n} X_i)^{-1} \sum_{i=1}^{n} (\frac{2i-1}{n} - 1)X_{(i)}.[1]$$

Пример 8:

Претпоставимо да подаци у *Табели 2* представљају узорак од 30 пореских прихода, добијених из низа случајних једнако-расподељених променљивих $X_1, \dots, X_{30}$, чија је функција расподеле F непозната.

3841	7084	7254	15228	18042	19089
22588	23972	25694	27592	27927	31576
32528	32921	33724	36887	37776	37992
39464	40506	44516	46538	51088	51955
54339	57935	75137	82612	83381	84741

Табела 2[1]

Употребом ових података може се извршити оцена Ђинијевог коефицијента $G(F) = 1 - 2\int_0^1 L(t)dt$ методом замене, иако је функција расподеле непозната.

Оцена поступком замене за *G(F)* биће:

$$\widehat{G} = G(\widehat{F}) = (\sum_{i=1}^{30} X_i)^{-1} \sum_{i=1}^{30} (\frac{2i-1}{30} - 1) X_{(i)},$$

где су $X_{(1)} \leq \cdots \leq X_{(30)}$ статистике поретка низа случајних величина. За дате податке, оцена Ђинијевог коефицијента *G(F)* износи **0.311** указујући на средњи степен неједнакости у расподели дохотка.

Стандардна грешка оцене $\widehat{G}$ може бити оцењена Џекнајф[2] методом и износиће

$$\widehat{se}(\widehat{G}) = \frac{29}{30} \sum_{i=1}^{30} (\widehat{G_{-i}} - \widehat{G_\bullet})^2 = 0.0398,$$

где $\widehat{G_\bullet} = 0.310$ представља средњу вредност за низ $\widehat{G_{-1}}, \ldots, \widehat{G_{-30}}$. *Табела 3* приказује вредности оцена $\widehat{G_{-i}}$ за $G(F)$, добијених изостављањем одговарајуће i-те вредности низа приликом израчунавања.

0.2912	0.2948	0.2950	0.3028	0.3055	0.3064
0.3092	0.3103	0.3115	0.3127	0.3129	0.3148
0.3153	0.3154	0.3157	0.3166	0.3168	0.3168
0.3170	0.3170	0.3169	0.3167	0.3161	0.3159
0.3152	0.3140	0.3069	0.3033	0.3028	0.3020

Табела 3 [1]

2.5 Оцена Ђинијевог коефицијента методом максималне веродостојности

Претходно описане методе оцене Ђинијевог коефицијента спадају у непараметарске методе, које се не заснивају на претпостављеној или познатој расподели из које узорак потиче. Супротно, параметарске методе увек крећу од претпоставке о расподели, била она тачна или погрешна.

[2] Jackknife

У овој методи, полази се од претпостављене расподеле $F(x, \Theta)$ чији се параметри Θ оцењују методом максималне веродостојности. На тај начин, уврштивањем оцене параметра у израз Ђинијевог коефицијента G, добија се одговарајућа Ђинијева оцена $\widehat{G} = G(\widehat{\Theta})$.

Наредно поглавље приказаће неке примере оцењивања овом методом, као и поређење са непараметарском оценом и теоријским вредностима.

2.6 Оцена Ђинијевог коефицијента различитим методама

Оцењивање се врши на основу низа независних једнако–расподељених случајних величина $X_1, \ldots, X_n$, који представља случајан узорак из исте популације. Посматрајмо у наставку узорке из две расподеле, лог-нормалне и Паретове. Пронашавши одговарајуће оцене њихових Ђинијевих коефицијената, могу се донети закључци о оценама добијеним непараметарским методама, и параметарским методама са познатом и непознатом расподелом.

Лог-нормална расподела:

Нека су дати узорак и густина лог–нормалне расподеле, са $\sigma = 1$:
$X_{LN} = \{0.20, 0.39, 0.41, 0.46, 0.47, 0.64, 1.12, 1.14, 1.19, 1.32, 1.46, 1.50, 1.54, 1.97, 2.1, 2.62, 3.25, 3.31, 3.65, 3.93\}$,

$f_{LN}(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma x} e^{(-\frac{(\ln(x)-\mu)^2}{2\sigma^2})}$ за $x > 0$,

где $\ln(X_i) \in N(\mu, \sigma^2)$.

Теоријски Ђинијев коефицијент лог–нормалне расподеле зависи само од σ и дат је са

$$G_{LN}(\sigma) = 2\Phi(\frac{\sigma}{\sqrt{2}}) - 1,$$

где је $\Phi(x)$ функција нормалне расподеле $N(0, 1)$.

Оцена максималне веродостојности за σ је

$\widehat{\sigma_n} = (\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y_n})^2)^{\frac{1}{2}}$, где је $Y_i = \ln(X_i)$.

Стога је одговарајућа оцена Ђинијевог коефицијента $G_{LN}(\widehat{\sigma_n})$.

За дати узорак X_{LN} оцена параметра износиће:

$$\widehat{\sigma_{20}} = (\frac{1}{20}\sum_{i=1}^{20}(Y_i - \bar{Y}_{20})^2)^{\frac{1}{2}} = (\frac{13,8302}{20})^{\frac{1}{2}} = 0.83.$$

Одатле, оцена Ђинијевог коефицијента је $\widehat{G} = G(\widehat{\sigma}) = 2\Phi(\frac{0,83}{\sqrt{2}}) - 1 = 0.44$, указујући на приличан степен неједнакости.

Узмимо у обзир неке теоријске вредности Ђинијевог коефицијента за лог-нормалну расподелу. За фиксиране вредности $\sigma = 0.5$, $\sigma = 1$ и $\sigma = 3$, редом се добијају $G(0.5) = 0.27$, $G(1) = 0.52$ и $G(3) = 0.96$.

За непараметарску методу, као што је *метода средњих разлика*, наведени узорак са $\sigma = 1$ даје непараметарску оцену $G_{20} = 0.39$. У овом случају се могла применити и оцена методом замене.

Паретова расподела:
Посматрајмо сада узорак и густину Паретове расподеле са $\alpha = 5$:
$X_P = \{6, 8, 12, 15, 22, 27, 30, 34, 45, 65, 72, 80, 109, 112, 200, 315, 400, 621, 700, 789\}$

$f_P(x; \alpha) = (\alpha - 1)(x + 1)^{-\alpha}$ за $x > 0$,

Теоријски Ђинијев коефицијент Паретове расподеле зависи од α:

$$G_P(\alpha) = \frac{\alpha - 1}{2\alpha - 3}.$$

Оцена максималне веродостојности за α је

$$\widehat{\alpha_n} = \frac{n}{\sum_{i=1}^{n} \log \frac{X_i}{X_0}}.$$

Аналогно, одатле је оцена Ђинијевог коефицијента $G(\widehat{\alpha_n}) = 0.48$.

Поново се могу одредити неке теоријске вредности показатеља за фиксиране α. На пример, за $\alpha = 5$, $\alpha = 10$ и $\alpha = 50$, добијају се вредности $G(5) = 0.57$, $G(10) = 0.53$ и $G(50) = 0.50$. С друге стране, непараметарски метод преко средњих разлика, такође даје приближну оцену $G_{20} = 0.69$.

Додатно, могу се проверити и оцене за погрешно претпостављену расподелу, при H_1. Наиме, Ђинијев коефицијент узорка X_{LN} оценимо са $\widehat{G_P} = G(\widehat{\alpha})$ што одговара Паретовој расподели, и обрнуто, за X_P, одредићемо $\widehat{G_{LN}} = G(\widehat{\sigma})$.

Посматрајмо најпре податке из X_{LN} и при H_1 за Паретову расподелу, одредимо оцену $\widehat{G_P} = G(\widehat{\alpha})$. По претходно објашњеном поступку, добијају се следеће вредности:

$$\widehat{\alpha_{20}} = \frac{20}{\sum_{i=1}^{20} \log \frac{X_i}{0.20}} = \frac{20}{36.09} = 0.55$$

и $\widehat{G} = \frac{0.45}{1.9} = 0.23$.

Слично, за податке X_P, под погрешном претпоставком о лог-нормалној расподели, Ђинијев коефицијент се оцењује са $\widehat{G_{LN}} = G(\widehat{\sigma})$. При томе се долази до резултата:

$$\widehat{\sigma_{20}} = (\frac{1}{20}\sum_{i=1}^{20}(Y_i - \bar{Y}_{20})^2)^{\frac{1}{2}} = (\frac{55.69}{20})^{\frac{1}{2}} = 1.67,$$

тиме следи, $\widehat{G} = 0.76$.

Табела 5 која следи, сумира примере наведене у овом поглављу:

Узорак	Параметар	Теоријска вр.	Непарам. оцена	$\widehat{G_{LN}}$	$\widehat{G_P}$
X_{LN}	$\sigma = 0,5$	0.27	0.33	0.30	0.15
X_{LN}	$\sigma = 1$	0.52	0.39	0.44	0.23
X_{LN}	$\sigma = 3$	0.96	0.51	0.71	0.40
X_P	$\alpha = 5$	0.57	0.69	0.76	0.48
X_P	$\alpha = 10$	0.53	0.63	0.79	0.45
X_P	$\alpha = 50$	0.50	0.60	0.76	0.43

Табела 5

Може се донети закључак да, иако скоро никада није најбоља, исто тако ни најгора, оцена добијена непараметарским методама се намеће као најсигурнији избор приликом оцењивања Ђинијевог коефицијента. Ова метода искључује ризике лоше одабраног узорка и не зависи од претпоставки.

2.7 Тест сагласности с експоненцијалном расподелом базиран на Ђинијевом коефицијенту

Након формирања тест статистике, идеја и поступак су слични процедури представљеној у тесту сагласности базираном на оцени Лоренцове криве. Нека је обележје X позитивна случајна величина.

Нулта хипотеза: Претпоставка је да је узорак X_n из експоненцијалне расподеле.

$$H_0 : F(x) = 1 - e^{-\lambda x}, x \geq 0.$$

Теоријски Ђинијев коефицијент експоненцијалне расподеле не зависи од параметра λ и износи $\frac{1}{2}$.

Тест статистика: Као полазна тачка формирања Ђини тест статистике, узима се емпиријски Ђинијев коефицијент:

$$G_n = \frac{\sum_{i,j} |X_i - X_j|}{2n(n-1)\bar{X}}.$$

Ова формула може се трансформисати у следећи запис, који су у свом истраживању увели Гејл и Гаствирт:

$$G_n = \frac{\sum_{i=1}^{n-1} i(n-1)(X_{(i+1)} - X_{(i)})}{(n-1)\sum_{i=1}^{n} X_i}.$$

Они такође уводе, нормализована растојања $\delta_{i+1} = (X_{(i+1)} - X_{(i)})(n-i)$ која су једнако расподељене случајне величине из експоненцијалне расподеле, што даље омогућава следећи запис:

$$G_n = \sum_{i=1}^{n-1} \frac{i\delta_{i+1}}{(n-1)T},$$

где су $T = \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} \delta_i$, и $D_i = \frac{\delta_i}{T}$, n Дирихлеових промењивих из $\Gamma(n)$, $D_i \geq 0$ и $\sum D_i = 1$. Претходни запис даје тражену **тест статистику** која се може записати и као:

$$G_n = \sum_{i=1}^{n-1} \frac{i}{n-1} D_i = \sum_{j=1}^{n} \frac{n-j}{n-1} D_j.$$

Расподелу за G_n под нултом хипотезом, 1969. године увели су Демпстер и Клајл:

$$P(G_n \leq x) = x^{n-1}\{\prod_{i=1}^{n-1} c_i\}^{-1} - \sum_{j=r+1}^{n-1}(x-c_j)^{n-1}\{c_j \prod_{k\neq j}^{n-1}(c_k - c_j)\}^{-1},$$

где је $c_j = \frac{n-j}{n-1}$, а r је највећи индекс такав да $x \leq c_r$. За велике вредности x (а тиме мале вредности r), знатно згоднији запис је

$$P(G_n \leq x) = 1 - \sum_{j=1}^{r}(c_j - x)^{n-1}\{c_j \prod_{k\neq j}^{n-1}(c_j - c_k)\}^{-1}.$$

Табела 6 садржи познате p-вредности за G_n, $n = 3, 4, \ldots, 20$, на основу којих се доноси одлука о прихватању хипотезе.

	Probability level				*Probability level*		
n	**0·950**	**0·975**	**0·990**	***n***	**0·950**	**0·975**	**0·990**
3	**0·84189**	**0·88818**	**0·92932**	**12**	**0·64337**	**0·66992**	**0·70020**
4	**0·77686**	**0·82288**	**0·86951**	**13**	**0·63725**	**0·66275**	**0·69183**
5	**0·73834**	**0·77997**	**0·82501**	**14**	**0·63185**	**0·65641**	**0·68448**
6	**0·71307**	**0·75079**	**0·79260**	**15**	**0·62704**	**0·65076**	**0·67792**
7	**0·69439**	**0·72931**	**0·76831**	**16**	**0·62273**	**0·64567**	**0·67197**
8	**0·67988**	**0·71252**	**0·74921**	**17**	**0·61882**	**0·64107**	**0·66659**
9	**0·66821**	**0·69896**	**0·73370**	**18**	**0·61527**	**0·63688**	**0·66168**
10	**0·65855**	**0·68768**	**0·72070**	**19**	**0·61201**	**0·63304**	**0·65723**
11	**0·65039**	**0·67816**	**0·70972**	**20**	**0·60902**	**0·62952**	**0·65308**

Табела 6 [3]

Претходна табела даје вредности за обим узорка до 20. У случају већих узорака, може се прибећи нормалној апроксимацији, која је адекватна чак и за мање обиме узорка. Хефдинг у свом раду доказује, да само при услову $EX^2 < \infty$, важи да $\frac{G_n - E(G_n)}{var(G_n)^{\frac{1}{2}}}$ у расподели тежи нормалној расподели, $N(0,1)$.

Узимајући у обзир нулту хипотезу о експоненцијалној расподели, очекивана средња вредност и варијација могу се израчунати из изведене тест статистике преко Дирихлеових променљивих и њихових момената.

Заиста, $EG_n = 0.5$ за све n и $var(G_n) = \frac{1}{12(n-1)}$, па тако $V_n = (G_n - 0.5)(12(n-1))^{\frac{1}{2}}$ тежи стандардној нормалној расподели.

Пример 9:

Да би илустровали случај теста сагласности с експоненцијалном расподелом, посматра се случајно одабран узорак из експоненцијалне расподеле обима 20. Нека је H_0: $F(x) = 1 - e^{-\lambda x}$, тј. претпоставља се да је расподела експоненцијална.

$X = \{0.04, 0.09, 0.11, 0.21, 0.27, 0.35, 0.44, 0.54, 0.58, 0.72, 0.77, 0.90, 0.98,$
$1.21, 1.40, 1.57, 1.95, 2.43, 2.45, 4.69\}$

Тест статистика је оцена Ђинијевог коефицијента:

$$G_{20} = \sum_{i=1}^{19} \frac{i\delta_{i+1}}{19T},$$

где се D_i и δ_i вредности могу израчунати према познатим формулама. Израчунавањем се добија оцена,

$$G_{20} = \frac{206.8}{412.3} = 0.50157.$$

Познато је да је теоријски Ђинијев коефицијент експоненцијалне расподеле једнак 0.5, па се може наслутити да ће хипотеза бити прихваћена услед јако малог одступања. При прагу значајности $\alpha = 0.05$, провера одговарајуће табличне вредности 0.60902 уверава нас да се H_0 прихвата, тј. расподела узорка јесте експоненцијална.

Пример 10:

У овом примеру узмимо у разматрање случајни узорак из гама расподеле $\gamma(1.5, 1)$ обима 50, да би избегли грешку друге врсте, на чијем је разматрању писано у [2]. Претпоставимо поново да је узорак из експоненцијалне расподеле, H_0: $F(x) = 1 - e^{-\lambda x}$.

$X = \{0.47, 0.66, 0.80, 1.11, 1.14, 1.24, 1.25, 1.28, 1.34, 1.40, 1.48, 1.49, 1.54,$
$1.57, 1.73, 1.74, ..., 2.16, 3.76, 4.37, 5.11\}$

Оцена Ђинијевог коефицијента у овом случају биће:

$$G_{20} = \sum_{i=1}^{19} \frac{i\delta_{i+1}}{19T} = 0.61708,$$

где се D_i и δ_i вредности могу израчунати према познатим формулама. При прагу значајности $\alpha = 0.05$ и провером одговарајуће табличне вредности, долазимо до закључка да се H_0 одбације, тј. уважава се H_1 претпоставка да расподела није експоненцијална.

Монте Карло оцена емпиријске мере и моћи:

Као у случају теста сагласности помоћу Лоренцове статистике, на случајно одабраних 10 000 узорака, може се оценити и емпиријска мера и моћ теста сагласности помоћу Ђинијеве статистике.

При оцени мере, 10 000 узорака обима 20 генерише се из експоненцијалне расподеле. При прагу значајности $\alpha = 0.05$, као што је приказано *Додатку 6*, добија се пропорција узорака при којима се H_0 одбацује. Добијени проценат износи 0.3209, што говори да у случајно генерисаних 10 000 узорака, њих 3209, одбацују хипотезу експоненцијалности, дајући грешку прве врсте.

Супротно, у оцени моћи теста, генерише се 10.000 узорака из $\gamma(2, 1)$ расподеле. Посматрајући у односу на алтернативну хипотезу H_1, добија се емпиријска моћ од 0.142.

Поглавље 3

Закључак

Лоренцова крива и Ђинијев коефицијент, као мере неједнакости у расподели, нашли су своју примену у многим научним областима. Њихова узрочно-последична веза са расподелом олакшава и омогућава доношење различитих закључака: о степену неједнакости у популацији на основу познате расподеле, или о самој расподели на основу оцењених мера неједнакости.

Оцене Лоренцове криве и Ђинијевог коефицијента могу бити непараметарске уз потпуно незнање о расподели посматраних опсервација, али и параметарске уз познату или непознату, тј. погрешно претпостављену расподелу. Оно о чему је било речи у раду, а покривено је и кроз различиту литературу о овим показатељима, је да у већини случајева најбоља је параметарска оцена са тачном расподелом. Супротно, параметарска оцена са погрешном расподелом, даје најлошије резултате. Као најсигурнији избор при оцењивању мера неједнакости, намећу се непараметарске методе, које нити су најбоље, нити најгоре, а дају оцене задовољавајућих својстава.

Управо из овог разлога, непараметарске оцене су искоришћене у опису тестова сагласности с експоненцијалном расподелом, а базираних на Лоренцовој кривој и Ђинијевом коефицијенту. Ђинијев коефицијент, као нумерички показатељ изведен из Лоренцове криве, јесте прецизнија мера која је за разлику од Лоренцове криве употребљива у било ком поређењу неједнакости.

Приказани тестови показали су различиту емпиријску моћ у односу на алтернативну хипотезу о гама расподелама. У наведеној литератури могу се пронаћи анализе моћи и за неке друге алтернативне расподеле, Вејбулову, Паретову и униформну. При алтернативној хипотези о Вејбуловој и униформној расподели, већу осетљивост показао је тест сагласности помоћу Ђинијеве статистике. Док за Паретову и гама расподелу, већу осетљивост има тест помоћу Лоренцове статистике.

Простор за додатна истраживања је широк, почев од унапређења Лоренцове криве, и мера неједнакости уопште, на обележја негативних вредности, до разбијања различитих митова о осетљивости Ђинијевог коефицијента на структуру и промене у структури популације.

Поглавље 4

Додаци

1) *R*-код, на основу ког су произведени резултати из *Примера 1*:

```
> x <- c(25,68,93,123,179,211,297,364,400,470)
> v1<-runif(length(x))
> v2<-100*v1
> par(mfrow=c(1,2))
> plot(Lc(x,v1),col=”purple”,lwd=2)
> plot(Lc(x,v2),col=”purple”,lwd=2)
> par(mfrow=c(1,1))
```

2) *R*-код, на основу ког су произведени резултати из *Примера 2*:

```
> x<-c(25,68,93,123,179,211,297,364,400,470)
> v1<-runif(length(x),10,30)
> v2<-runif(length(x),0,100)
> plot(Lc(x,v1),col=”purple”,lwd=2)
> par(new=TRUE)
> plot(Lc(x,v2),axes=FALSE,col=”coral”,lwd=2)
```

3) База AirPassengers садржи податке о обиму путника у авио саобраћају за период од 1949. до 1960. године. База је употребљена у *Примеру 3* за анализу неједнакости у расподели.

```
> AirPassengers
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
1949 112 118 132 129 121 135 148 148 136 119 104 118
1950 115 126 141 135 125 149 170 170 158 133 114 140
1951 145 150 178 163 172 178 199 199 184 162 146 166
1952 171 180 193 181 183 218 230 242 209 191 172 194
1953 196 196 236 235 229 243 264 272 237 211 180 201
```

1954 204 188 235 227 234 264 302 293 259 229 203 229
1955 242 233 267 269 270 315 364 347 312 274 237 278
1956 284 277 317 313 318 374 413 405 355 306 271 306
1957 315 301 356 348 355 422 465 467 404 347 305 336
1958 340 318 362 348 363 435 491 505 404 359 310 337
1959 360 342 406 396 420 472 548 559 463 407 362 405
1960 417 391 419 461 472 535 622 606 508 461 390 432

R-код, на основу ког су произведени резултати базирани на претходним подацима, а описани у *Примеру 3*:

```
> ineq(AirPassengers,type="Gini") # Djinijev koeficijent
[1] 0.2407563
> plot(Lc(AirPassengers),col="purple",lwd=2) # Lorencova kriva
```

4) *R*-код, на основу ког су произведени резултати из *Примера 7*:

```
> library(reldist)
> x <- c(548, 1763, 2495, 3438, 4257, 5481, 6392, 8304, 12904, 22269)

> y <- runif(n=length(x))
> plot(Lc(x,y),col="purple",lwd=2)
> gini(x,y)
[1] 0.4234095
```

5) Лоренцова тест статистика - емпиријска мера и моћ

```
> brojac = 0
> n = 10000

> for(i in 1:n){
+ uzorak < − rexp (20, 1)
+ sortuz < − sort(uzorak)
+ ocena < − sum(sortuz[1]:sortuz[10])/sum(sortuz[1]:sortuz[20])
+ if (ocena > 0.10055) brojac < − brojac + 1
+ }
> m < − brojac/n
```

6) Ђинијева тест статистика - емпиријска мера и моћ

```
> brojac = 0
> n = 10000
```

```
> for(i in 1:n){
+ uzorak < − rexp (20, 1)
+ sortuz < − sort(uzorak)
+ delilac = 19*sum(sortuz)
+ for(j in 1 : 19){
+ clan = j ∗ (sortuz[j + 1] − sortuz[j]) ∗ (19 − j)
+ deljenik < − deljenik + clan
+ }
+ ocena < − deljenik/delilac
+ if (ocena > 0.39098) brojac < − brojac + 1
+ }
> m < − brojac/n
```

Литература

[1] K. Knight, Mathematical Statistics, Chapman & Hall, (2000)

[2] M. H. Gail and J. L. Gastwirth, A Scale-Free Goddness-of-Fit Test for the Exponential Distribution Based on the Lorenz Curve, American Statistical Association, (1978)

[3] M. H. Gail and J. L. Gastwirth, A Scale-Free Goodness-of-Fit Test for the Exponential Distribution Based on the Gini Statistic, Royal Statistical Society, (1978)

[4] J. L. Gastwirth, The Estimation of the Lorenz Curve and Gini Index, The MIT press, (1972)

[5] M. Lubrano, The econometric of inequality and poverity, Centre de la Vieille Charite, (2017)

[6] C. Dagum, Encyclopedia of Statistical Sciences, John Wiley & Sons, Inc., (2004)

[7] В. Јевремовић, Једна лекција - стотину статистичких појмова, Универзитетски центар за примењену статистику Нови Сад

[8] С. Деветаквић, Б. Јовановић-Гавриловић, Г. Рикаловић, Национална економија, Центар за издавачку делатност Економског факултета у Београду, (2012)

[9] М Јакшић, Основи макроекономије, Центар за издавачку делатност Економског факултета у Београду, (2011)

[10] Б. Ковачић, Р. Опачић, Л. Марохнић, О Гинијеву коефицијенту концентрације, Хрватски математички електронски часопис, (2009)

[11] Ђинијев коефицијент, https://sr.wikipedia.org/Gini_coefficient

[12] Лоренцова крива, https://en.wikipedia.org/wiki/Lorenz_curve

Биографија:

Јована Дубљанин, рођена је 20. априла 1992. године у Ужицу, где је завршила основну и средњу школу као носилац Вукове дипломе. Године 2011. уписује Математички факултет Универзитета у Београду, на смеру Математика - статистика, актуарска и финансијска математика, на ком и дипломира 2015. Исте године, школовање наставља на Математичком факултету, уписавши мастер студије на истом смеру. Са почетком мастер студија, започиње и каријеру у области Информационих технологија као пословни аналитичар и у области контроле квалитета софтвера.