math.e

Hrvatski matematički elektronički časopis

O Ginijevu koeficijentu koncentracije

definicija **Ginijev koeficijent koncentracije** **Lorenzova krivulja** **osnovna svojstva**

Bojan Kovačić, Renata Opačić i Luka Marohnić

Tehničko veleučilište u Zagrebu, Zagreb, Vrbik 8a,
bojan.kovacic@tvz.hr renata.opacic@tvz.hr, luka.marohnic@tvz.hr

Sažetak

U opisnoj (deskriptivnoj) statistici mjere koncentracije izdvajaju se kao pokazatelji načina razdiobe zbroja svih vrijednosti numeričkoga obilježja[1], svih vrijednosti konačnoga numeričkog statističkog niza podataka[2] ili neke druge veličine prema elementima statističkoga skupa ili modalitetima statističkoga obilježja. Pritom se razlikuju apsolutne i relativne mjere koncentracije. U apsolutne mjere pripadaju *koncentracijski omjer* i *Herfindahl*[3]*–Hirschmanov*[4] *indeks*, dok se kao najčešća relativna mjera koncentracije navodi *Ginijev*[5] *koeficijent koncentracije* ili *Ginijev indeks*. U ovome ćemo članku opisati taj koeficijent, te dati statističke interpretacije nekih njegovih konkretnih vrijednosti koje se odnose na Hrvatsku.

Ključne riječi: Ginijev koeficijent koncentracije, Lorenzova krivulja

About Gini concentration coefficient

Sažetak

In descriptive statistics concentration measures stand apart showing distribution of *total*, all values of finite numerical statistical data sequence or some other statistical value distributed by elements of statistical set or variable modality. There are *absolute* and *relative* concentration measures. Absolute measures include *concentration ratio* and *Herfindahl-Hirschman index*, while the most frequently used relative measure is *Gini coefficient* or *Gini index*. In this article we will describe the latter and give statistical interpretations of its concrete values related to Republic of Croatia.

Keywords: Gini concentration coefficient, Lorenz curve

1. Uvod

Kako je već istaknuto u sažetku, Ginijev koeficijent koncentracije pripada u *relativne mjere koncentracije* ili *mjere nejednakosti* (*dispariteta*) statističkoga niza. Njihova vrijednost može biti bilo koji realan broj iz segmenta $[0, 1]$. Što je ta vrijednost bliža nuli, vrijednosti niza su ravnomjernije raspoređene, tj. na svaki modalitet otpada približno jednak udio u totalu. Obrnuto, što je ta vrijednost bliža jedinici, vrijednosti niza su neravnomjernije raspoređene: slobodno govoreći, većina totala otpada na jedan modalitet.

Okvirno pojasnimo ovu ideju na konkretnom primjeru.

Primjer 1. *U tablici 1 zadane su razdiobe uvoza i izvoza Kraljevine Ubananije*[6] *2011. godine.*

Zemlja	*Uvoz (mil. €)*	*Izvoz (mil. CHF)*
Kraljevina Niškoristija	*0*	*25*
Savezna Republika Vrijedlandija	*100*	*25*
Kneževina Drplandija	*0*	*25*
Republika Ljenčarija	*0*	*25*

Tablica 1: podaci preuzeti iz Statističkog ljetopisa Kraljevine Ubananije, 2011. (Primjer 1.)

Iz Tablice 1 se vidi da Kraljevina Ubananija uvozi isključivo iz SR Vrijedlandije, pa je pripadna relativna mjera koncentracije u ovom slučaju jednaka 1. Uvoz je ravnomjerno raspoređen prema svim četirima zemljama, pa je relativna mjera koncentracije u ovom slučaju jednaka 0.

2. Lorenzova krivulja

Da bismo precizno definirali Ginijev koeficijent, najprije moramo definirati *Lorenzovu*[7] *krivulju*. U tu svrhu pretpostavimo da statistički niz $x_1, x_2, \ldots, x_n$ ima sljedeća svojstva:

(S1)
$x_i \geq 0$, za svaki $i \in [n]$;

(S2)
Postoji barem jedan $i \in [n]$ takav da je $x_i > 0$.

Pritom je standardno:

- $[n] := \{1, 2, \ldots, n\}$ tj. skup prvih n prirodnih brojeva,
- $[n]_0 := \{0, 1, 2, \ldots, n\}$.

Budući da je svaki konačan podskup skupa realnih brojeva moguće uzlazno urediti, tj. poredati elemente toga podskupa od najmanjeg do najvećeg, bez smanjenja općenitosti možemo pretpostaviti da niz $x_1, x_2, \ldots, x_n$ ima svojstvo:

(S3)
$x_1 \leq x_2 \leq \cdots \leq x_n$,

tj. da je taj niz uzlazno uređen. Tada za svaki $i \in [n]$ definiramo:

$$p_i := \frac{i}{n}\,, \tag{1}$$

$$S_i := \sum_{k=1}^{i} x_k, \tag{2}$$

$$y_i := \frac{S_i}{S_n}\,. \tag{3}$$

Pojasnimo konkretno značenje ovih triju izraza. Izraz (1) definira točke koje dijele segment $[0, 1$ na n jednakih dijelova. Izraz (2) definira i–ti djelomični zbroj članova niza (analogno kao štc činimo prigodom definicije numeričkoga reda), pri čemu je S_n zbroj svih članova niza. Napokon, izraz (3) definira udio k–toga djelomičnog zbroja u ukupnom zbroju svih članova niza.

Za netom definirane vrijednosti definiramo niz ravninskih točaka T_i, $i = [n]_0$ s:

$$\begin{cases} T_0 & = (0,0); \\ T_i & = (p_i, y_i), \text{ za svaki } i \in [n]. \end{cases}$$

Ucrtamo li dobivene točke u pravokutni koordinatni sustav u ravnini, dobit ćemo tzv. *izlomljenu poligonalnu liniju* L. Ta linija je graf neprekidne, po dijelovima linearne funkcije. Linija L počinje u ishodištu, završava u točki $T_n = (1, 1)$ i naziva se *Lorenzova krivulja*. Istaknimo jedno svojstvc Lorenzove krivulje L.

Propozicija 2. *Sve točke krivulje L nalaze se ili ispod pravca $y = x$ ili na tom pravcu.*

Dokaz. Dovoljno je dokazati da za svaki $i \in [n]$ vrijedi nejednakost:

$$y_i \le p_i. \tag{4}$$

Definirajmo niz $(z_k)_{k\in[n]}$ na sljedeći način (uz upotrebu (2)):

$$z_i := \frac{S_i}{i} = \frac{1}{i}\sum_{k=1}^{i} x_k. \tag{5}$$

Tvrdimo da je niz $(z_k)_{k\in[n]}$ monotono rastući. Doista, za svaki $i \in [n-1]$ vrijedi:

$$\begin{aligned} z_{i+1} - z_i &= \frac{1}{i+1}\sum_{k=1}^{i+1} x_k - \frac{1}{i}\sum_{k=1}^{i} x_k = \\ = \frac{i\sum_{k=1}^{i+1} x_k - (i+1)\sum_{k=1}^{i} x_k}{i(i+1)} &= \frac{i\cdot x_{i+1} - \sum_{k=1}^{i} x_k}{i(i+1)}\,. \end{aligned} \tag{6}$$

Nazivnik posljednjeg razlomka je očito strogo pozitivan, a nenegativnost brojnika tog razlomka izravno slijedi zbrajanjem nejednakosti

$$\begin{aligned} x_1 &\le x_{i+1}\,, \\ x_2 &\le x_{i+1}\,, \\ &\dots \\ x_i &\le x_{i+1}\,. \end{aligned}$$

koje vrijede prema svojstvu (S3). Stoga iz (6) slijedi

$$z_i \leq z_n,$$

odnosno, prema (5),

$$\frac{1}{i}\sum_{k=1}^{i} x_k \leq \frac{1}{n}\sum_{k=1}^{n} x_k,$$

odnosno

$$\frac{\sum_{k=1}^{i} x_k}{\sum_{k=1}^{n} x_k} \leq \frac{i}{n}\,. \tag{7}$$

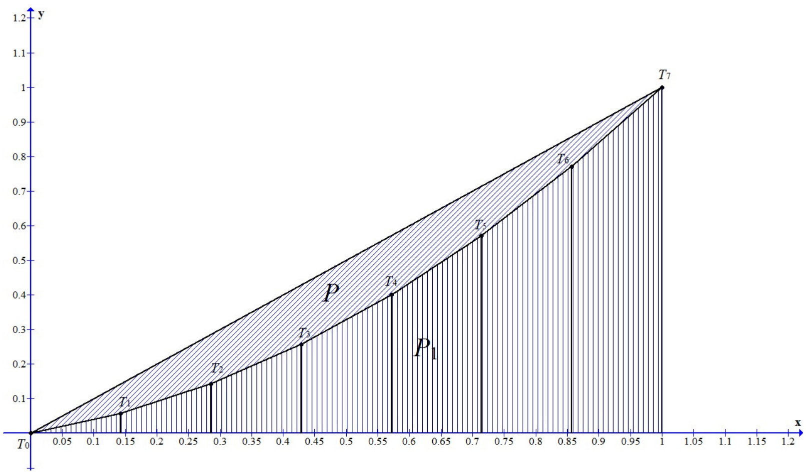
Nejednakost (7) je, prema (1), (2) i (3), ekvivalentna nejednakosti

$$y_i \leq p_i, \text{ za svaki } i \in [n],$$

koju smo i željeli dokazati. ∎

Neka su

- $P =$ površina ravninskoga lika koji krivulja L zatvara s pravcem $y = x$;
- $P_1 =$ površina ravninskoga lika koji krivulja L zatvara s osi x i pravcem $x = 1$;
- $P_2 =$ površina trokuta s vrhovima u točkama $T_0 = (0,0)$, $A = (1,0)$ i $T_n = (1,1)$.



Slika 1: Primjer Lorenzove krivulje

Očito vrijede jednakosti:

$$P_2 = \frac{1}{2}\,, \tag{8}$$

$$P = P_2 - P_1. \tag{9}$$

Površina P definirana u prethodnom paragrafu u opisnoj se statistici interpretira upravo kao mjera

3. Definicija i osnovna svojstva Ginijeva koeficije Izračun Ginijeva koeficijenta u slučaju negrupira podataka

koncentracije, odnosno, grubo rečeno, kao mjera „odstupanja" Lorenzove krivulje od pravca $y = x$. Omjer površina P i P_2 promatra se kao relativna jedinica mjere koncentracije. Taj se omjer naziva Ginijev koeficijent koncentracije. Formalno imamo sljedeću definiciju:

Definicija 3. *Ginijev koeficijent koncentracije statističkoga niza $(x_i)_{i\in[n]}$ je nenegativan realan broj G definiran s*

$$G := \frac{P}{P_2} = 2P = 1 - 2P_1. \tag{10}$$

U Definiciji 1 Ginijev koeficijent izrazili smo pomoću površine P_1 jer je tu površinu lako izračunati koristeći koordinate članova niza točaka $(T_i)_{i\in[n]_0}$. Doista, ravninski lik omeđen izlomljenom linijom L i pravcima $x = 1$ i $y = 0$ sastoji se od jednoga pravokutnoga trokuta i n–1 pravokutnih trapeza. Površina pravokutnoga trokuta jednaka je

$$P_\triangle = \frac{y_1}{2n}\,, \tag{11}$$

dok je površina i-toga trapeza jednaka

$$\tilde{P}_i = \frac{y_i + y_{i+1}}{2n}\,,\ \text{za svaki } i = [n-1]. \tag{12}$$

Iz (11) i (12) slijedi:

$$P_1 = P_\triangle + \sum_{i=1}^{n-1} \tilde{P}_i = \frac{2\sum_{i=1}^{n-1} y_i + y_n}{2n} = \frac{2\sum_{i=1}^{n-1} y_i + 1}{2n}\,, \tag{13}$$

pa uvrštavanjem u (10), koristeći (2) i (3), dobivamo

$$\begin{aligned}
G &= 1 - 2P_1 = 1 - \frac{2\sum_{i=1}^{n-1} y_i + 1}{n} = 1 - \frac{2\sum_{i=1}^{n-1} \frac{S_i}{S_n} + 1}{n} = \\
&= 1 - \frac{2\sum_{i=1}^{n-1} S_i + S_n}{nS_n} = 1 - \frac{2\sum_{i=1}^{n} S_i - S_n}{nS_n} = \\
&= 1 - \frac{2\sum_{i=1}^{n}\sum_{k=1}^{i} x_k}{nS_n} + \frac{1}{n} = 1 + \frac{1}{n} - \frac{2\sum_{i=1}^{n}(n-i+1)\cdot x_i}{nS_n} = \\
&= \frac{n+1}{n} - \frac{2(n+1)\sum_{i=1}^{n} x_i - 2\sum_{i=1}^{n} i\cdot x_i}{nS_n} = \\
&= \frac{n+1}{n} - \frac{2(n+1)\cdot S_n - 2\sum_{i=1}^{n} i\cdot x_i}{nS_n} = \\
&= \frac{2\sum_{i=1}^{n} i\cdot x_i}{n\sum_{i=1}^{n} x_i} + \frac{n+1}{n} - \frac{2(n+1)}{n} = \frac{2\sum_{i=1}^{n} i\cdot x_i}{n\sum_{i=1}^{n} x_i} - \frac{n+1}{n}\,.
\end{aligned}$$

Dobiveni rezultat iskažimo sljedećom propozicijom.

Propozicija 4. *Za numerički niz $(x_i)_{i\in[n]}$ Ginijev koeficijent koncentracije G jednak je*

$$G = \frac{2\sum_{i=1}^{n} i \cdot x_i}{n\sum_{i=1}^{n} x_i} - \frac{n+1}{n}\,. \tag{14}$$

Navedimo neka svojstva Ginijeva koeficijenta.

Propozicija 5. *Za Ginijev koeficijent koncentracije G vrijedi sljedeće:*

a)
$G \in [0, 1]$.

b)
$G = 0$ ako i samo ako je $x_1 = x_2 = \cdots = x_n$.

c)
$G = 1 - \frac{1}{n}$ ako i samo ako je $x_1 = x_2 = \cdots = x_{n-1} = 0$ i $x_n > 0$.

Dokaz. Iz Definicije 1 imamo $0 \leq G = 1 - 2P_1$. Zbog $P_1 \geq 0$ je $G \leq 1$, odakle slijedi **a)**.

Dokažimo **b)**. Pretpostavimo li da je $x_1 = x_2 = \cdots = x_n$, iz (14) slijedi $G = 0$. Obrat slijedi iz dokaza Propozicije 1 jer se jednakost u (5) postiže za $x_1 = x_2 = \cdots = x_n$.

Slično pokazujemo da vrijedi i **c)**. Pretpostavimo li da je $x_1 = x_2 = \cdots = x_{n-1} = 0$ i $x_n > 0$, iz (14) slijedi $G = 1 - \frac{1}{n}$. Obrnuto, pretpostavimo da je $G = 1 - \frac{1}{n}$. Iz Definicije 1 slijedi

$$1 - 2P_1 = 1 - \frac{1}{n}$$

tj.

$$P_1 = \frac{1}{2n}\,.$$

Iz (13) slijedi

$$\sum_{i=1}^{n-1} y_i = 0$$

pa iz (2), (3) i svojstva (**S2**) slijedi tvrdnja. ∎

Napomena 6. *Budući da za svaki$n \in N$ vrijedi stroga nejednakost*

$$1 - \frac{1}{n} < 1,$$

tvrdnju Propozicije 3.a) možemo „pooštriti" na $G \in [0, 1\rangle$.}

Napomena 7. *U slučaju iz Propozicije 3.b) govorimo o savršeno ravnomjernoj koncentraciji. To znači da na svaki pojedini modalitet otpada jednak dio ukupnoga zbroja svih članova niza.*

Napomena 8. *Radi prirodnoga proširenja na slučaj beskonačnih numeričkih nizova, u slučaju iz Propozicije 3.c) (i samo u tom slučaju!) dogovorno se definira* $G := 1$ *i govori o savršeno neravnomjernoj koncentraciji. To znači da ukupan zbroj svih članova niza otpada na točno jedan modalitet.*

4. Izračun Ginijeva koeficijenta u slučaju podal grupiranih u razrede

U gospodarskoj praksi česti su slučajevi kad se numerički podaci, zbog svoje brojnosti i radi preglednosti, grupiraju u prave razrede. Podsjetimo, pravi razredi su (konačni) nizovi intervala $[x_0, x_1\rangle$, $[x_1, x_2\rangle$, $\ldots$, $[x_{n-2}, x_{n-1}\rangle$, $[x_{n-1}, x_n]$, pri čemu je obično x_0 najmanja vrijednost numeričkog niza podataka, a x_n najveća vrijednost tog niza. Razredna sredina i–toga pravog razreda je broj

$$(\bar{x})_i = \frac{x_{i-1} + x_i}{2}\ , \text{ za svaki } i \in [n].$$

Pretpostavimo da je, za svaki $i \in [n]$, i–tom pravom razredu pridružena apsolutna frekvencija f_i, što znači da ukupno f_i članova numeričkog niza podataka pripada dotičnom razredu. Umjesto apsolutne frekvencije, mogu se razmatrati i relativne frekvencije r_i koje označavaju udio broja dotičnog razreda u ukupnom broju svih članova niza.

Uz upravo uvedene oznake, definicijske relacije (1), (2) i (3) prelaze u:

$$p_i := \frac{f_i}{\sum_{k=1}^{n} f_k}\ , \tag{15}$$

$$S_i := \sum_{k=1}^{i} f_k \cdot (\bar{x})_k, \tag{16}$$

$$y_i := \frac{S_i}{S_n}\ . \tag{17}$$

Lorenzova krivulja L ponovno se dobiva iz točaka $T_0 = (0, 0)$ i $T_i = (p_i, y_i)$ za $i \in [n]$. Definiramo li dogovorno $y_0 := 0$, onda iz razmatranja analognoga dokazu Propozicije 2 lako dobivamo:

$$G = 1 - \sum_{i=1}^{n} p_i \cdot (y_{i-1} + y_i).$$

5. Normirani Ginijev koeficijent koncentracije

U Napomeni 1 ustvrdili smo da je najveća moguća vrijednost Ginijeva koeficijenta jednaka

$G_{max} = 1 - \frac{1}{n}$. Stoga se u praksi za „male" vrijednosti n (obično za $n \leq 20$) dodatno računa omjer vrijednosti G i G_{max}. Taj se omjer naziva *normirani Ginijev koeficijent koncentracije.* Formalna definicija je:

Definicija 9. *Normirani Ginijev koeficijent koncentracije G^* za numerički niz $(x_i)_{i\in[n]}$ definiran je s:*

$$G^* = \frac{G}{G_{max}} = \frac{n}{n-1} \cdot G. \tag{18}$$

Iz (14) i (18) izravno slijedi

$$G^* = \frac{2\sum_{i=1}^{n} i \cdot x_i}{(n-1)\sum_{i=1}^{n} x_i} - \frac{n+1}{n-1}.$$

Lako se vidi da za normirani Ginijev koeficijent G^* vrijedi sljedeća varijanta Propozicije 3.

Propozicija 10. *Za normirani Ginijev koeficijent koncentracije G^* vrijedi:*

a)
$G^* \in [0, 1]$.

b)
$G^* = 0$ ako i samo ako je $x_1 = x_2 = \cdots = x_n$.

c)
$G^* = 1$ ako i samo ako je $x_1 = x_2 = \cdots = x_{n-1} = 0$ i $x_n > 0$.

Napomena 11. *Nejednakost$x_n > 0$ u Propozicijama 3.c) i 4.c) slijedi iz uvjeta $x_1 = x_2 = \cdots = x_{n-1} = 0$ i svojstva (**S2**), pa je zapravo nismo trebali posebno navoditi. To smo ipak učinili kako bismo opetovano naglasili da niz $(x_n)_{n\in[n]}$ nužno sadrži najmanje jednu strogo pozitivnu vrijednost.*

6. Kontinuirana varijanta formule (10)

Pretpostavimo da se numerički podaci odnose na neko kvantitativno kontinuirano statističko obilježje (npr. visina, masa itd.) i da tvore segment $[a, b]$. U tom ih slučaju ne možemo poredati u (beskonačan) niz, što znači da pripadnu Lorenzovu krivulju možemo shvatiti kao graf neke realne funkcije $f(x)$ na segmentu $[0, 1]$. Tada definicijska relacija (10) prelazi u:

$$G := 1 - 2\int_0^1 f(x)\,\mathrm{d}x. \tag{19}$$

Formulu (19), dakako, ima smisla primijeniti znamo li propis funkcije f. Funkcija f ne može biti bilo kakva realna funkcija, nego mora imati sljedeća svojstva:

- $f(0) = 0$,
- $f(1) = 1$,
- $0 \leq f(x) \leq x$ za svaki $x \in [0, 1]$,

- f je monotono rastuća i Riemann/Lebesgue integrabilna na segmentu $[0, 1]$.

Varijanta Propozicije 3 za ovaj slučaj glasi:

Propozicija 12. *Za Ginijev koeficijent koncentracije G vrijedi sljedeće:*

a) $G \in [0, 1]$,

b) $G = 0$ ako i samo ako je $f(x) = x$, osim za konačno mnogo $x \in [0, 1]$ za koje je $0 < f(x) \leq x$,

c) $G = 1$ ako i samo ako je $f(x) = 0$, osim za konačno mnogo $x \in [0, 1]$ za koje je $0 < f(x) \leq x$.

Napomena 13. *U prvi mah moglo bi se zaključiti da je varijanta Propozicije 3.c) ekvivalencija jednakosti$G = 1$ i jednakosti*

$$f(x) = \begin{cases} 0, & \text{za } x \neq 1, \\ 1, & \text{za } x = 1. \end{cases}$$

No, to je pogrešno. Naime, iz matematičke je analize poznato da je integral nenegativne Riemann/Lebesgue-integrabilne realne funkcije f jednak nuli ako i samo ako je $f(x) = 0$, osim možda za konačno mnogo vrijednosti varijable x. U ovom slučaju postoji jedna takva vrijednost (to je $x = 1$), pa zato u iskazu Propozicije 5.c) treba navesti „osim za konačno mnogo (...)". Integral u formuli (19) je „obični" jednostruki Riemannov integral, pa prebrojivost skupa točaka u kojima je $f(x) \neq 0$ povlači da f nije integrabilna u Riemannovom smislu i tada ne možemo primijeniti citiranu formulu.

7. Primjeri

U posljednjem dijelu pokazat ćemo izračun Ginijevog koeficijenta, kao i izradu Lorenzove krivulje na nekoliko jednostavnih primjera. Potrebni podaci dobiveni su korištenjem službeno objavljenih podataka Državnog zavoda za statistiku Republike Hrvatske. Kako je već spomenuto, Ginijev koeficijent je mjera nejednakosti statističkog niza. U svakidašnjem životu vjerojatno je najzanimljivija primjena izračuna Ginijevog koeficijenta te analiza nejednakosti pri raspodjeli dohotka čime ćemo se pozabaviti u prva dva primjera. U Tablici 2 i Tablici 5 dani su podaci prikupljeni Anketom o potrošnji kućanstva iz 2010. godine. Njome se prikupljaju podatci o ukupnom dohotku i različitim oblicima potrošnje kućanstava i ona je glavni izvor podataka za izračune nejednakosti i siromaštva u RH.

Primjer 14. *Prvo ćemo se koncentrirati na komponente dohotka: raspoloživi dohodak obuhvaća dohodak od nesamostalnog rada, samostalne djelatnosti, mirovine te transfera i ostalih primitaka. Podaci su navedeni u Tablici 2.*

	Dohodak, kune
Dohodak od nesamostalnog rada	*47 346*
Dohodak od samostalne djelatnosti	*12 981*
Dohodak od imovine (nije uključena imputirana najamnina)	*762*
Mirovine	*12 284*
Transferi i ostali primici	*10 474*
Ukupno	*86 847*

Tablica 2: Raspoloživi dohodak prema izvorima stjecanja, prosjek po kućanstvu u 2009. (Primjer 2.)

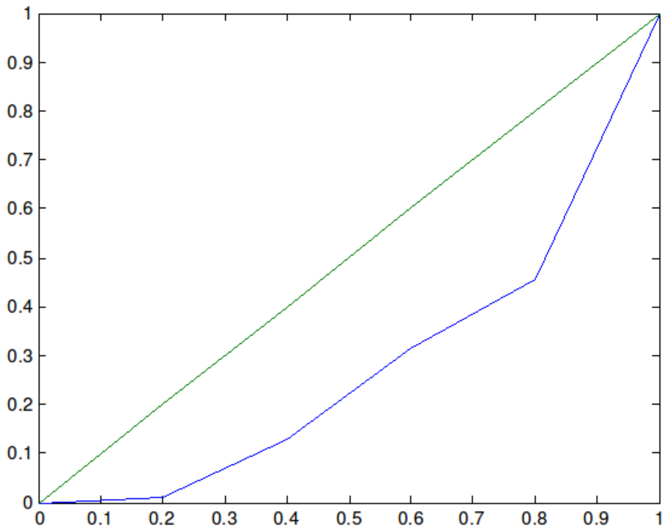
Počnimo s konstrukcijom Lorenzove krivulje: statistički niz je očito sastavljen od nenegativnih vrijednosti takvih da je najmanje jedna od njih strogo veća od nule (tj. zadovoljeni su uvjeti (S1) i (S2)). Međutim, trebamo ga još uzlazno urediti (uvjet (S3)), tj. poredati elemente niza od najmanjeg prema najvećem:

$$(x_1, x_2, x_3, x_4, x_5) = (762, 10474, 12981, 15284, 47346).$$

Lorenzovu krivulju dobivamo spajanjem točaka $(0, 0)$, (p_1, y_1), (p_2, y_2), \dots, $\big(p_5, y_5\big) = (1, 1)$. Izračunate vrijednosti p_i i y_i za $i \in [5]$ nalaze se u Tablici 3, dok je Lorenzova krivulja prikazana na Slici 2.

	dohodak, kn	*kumulativ podtotala*	p_i	y_i
Dohodak od nesamostalnog rada	762	762	0.2	0.0088
Dohodak od samostalne djelatnosti	10 474	11 236	0.4	0.1294
Dohodak od imovine	12 981	24 217	0.6	0.3134
Mirovine	15 284	39 501	0.8	0.4548
Transferi i ostali primici	47 346	86 847	1	1
Ukupno	86 847	-	-	-

Tablica 3: pomoćna tablica za Lorenzovu krivulju (Primjer 2.)



Slika 1: Lorenzova krivulja za Primjer 2.

Ginijev koeficijent računamo prema formuli (14), koristeći već izračunate vrijednosti navedene u Tablici 4.

	dohodak, kn	k	$k \cdot x_k$
Dohodak od nesamostalnog rada	762	1	762
Dohodak od samostalne djelatnosti	10 474	2	20 948
Dohodak od imovine	12 981	3	38 943
Mirovine	15 284	4	236 780
Transferi i ostali primici	47 346	5	358 569
Ukupno	86 847	-	-

Tablica 4: pomoćna tablica za Ginijev koeficijent (Primjer 2.)

Dobivamo:

$$G = \frac{2 \cdot 358519}{5 \cdot 86847} - \frac{6}{5} = 0.45.$$

Sada računamo i normiran Ginijev koeficijent

$$G^* = \frac{n}{n-1} \cdot G = 0.45 \cdot \frac{5}{4} = 0.56.$$

Iz Lorenzove krivulje i dobivenih vrijednosti Ginijevog koeficijenta, možemo zaključiti da su komponente dohotka neravnomjerno raspoređene. Naime, već Lorenzova krivulja pokazuje da npr. dohodak od imovine čini samo 0.88% ukupnoga raspoloživog dohotka, dok na dohodak od nesamostalnog rada otpada više od 50% ukupnog dohotka. Nadalje, otprije znamo da, što je vrijednost Ginijevog koeficijenta bliža nuli, vrijednosti niza su ravnomjernije raspoređene.

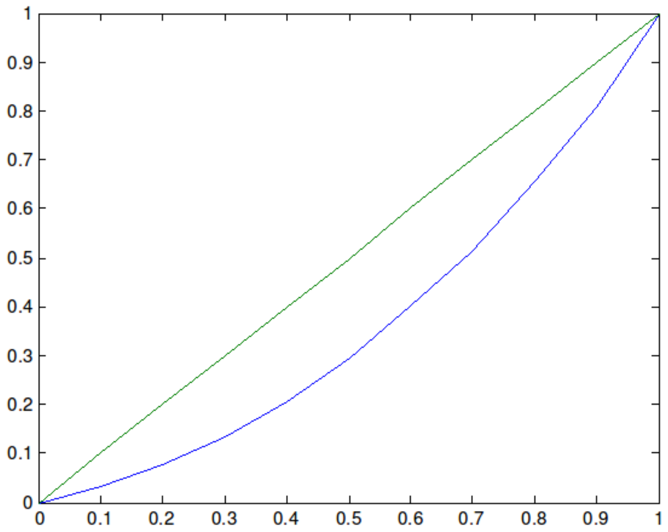
Analogno, približavanje jedinici znači i porast nejednakosti među elementima niza, tj. koncentraciju na samo jedan modalitet, što bi u ovom slučaju bio dohodak od nesamostalnog rada.

Primjer 15. *U ovome primjeru analizirat ćemo raspodjelu dohotka prema dohodovnim decilima u 2009. godini. Decili se oblikuju podjelom osnovnog skupa na deset jednakih dijelova izračunavanjem prosječnog neto-dohotka po kućanstvu, razvrstavanjem kućanstva prema dohotku od najnižega prema najvišemu te svrstavanjem svakoga pojedinog kućanstva u pripadajući decil. U prvom decilu su kućanstva s najnižim godišnjim neto-dohotkom. U drugom decilu su kućanstva čiji je godišnji neto-dohodak veći od kućanstava iz prvog decila, a manji od kućanstava iz trećeg decila itd. Kućanstva iz 10. decila imaju najviši godišnji neto-dohodak. Podaci su prikazani u Tablici 5.*

decili	*Izdaci za potrošnju - ukupno, kune*
1	*24 393*
2	*33 717*
3	*43 292*
4	*53 722*
5	*67 530*
6	*79 902*
7	*85 326*
8	*106 789*
9	*115 107*
10	*144 202*
	Ukupno: 753 980

Tablica 5: struktura izdataka za potrošnju prema dohodovnim decilima u 2009. (Primjer 3.)

Do rezultata dolazimo na isti način kao u prethodnom primjeru. Najprije pogledajmo Lorenzovu krivulju, prikazanu na Slici 3.



Slika 2: Lorenzova krivulja za Primjer 3.

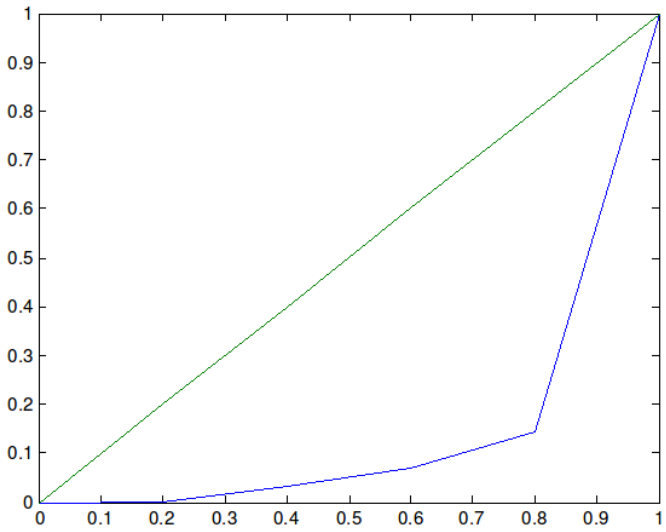
Vidimo da su vrijednosti uglavnom ravnomjerno raspoređene, odnosno niti jedan se modalitet ne ističe, već na svaki od njih otpada približno jednak udio u totalu pa očekujemo da će to potvrditi i sam Ginijev koeficijent. Izračunavanjem Ginijeva koeficijenta dobivamo da je njegova vrijednost jednaka 0.27 ($G^* = 0.31$), što potvrđuje zaključak o ravnomjernoj raspodjeli.

Primjer 16. *U posljednjem primjeru pogledajmo Tablicu 6 koja sadrži podatke o izvozu/uvozu RH u 2009. godini. (Podatci su preuzeti iz Statističkog ljetopisa 2010.)*

	Izvoz, tisuće kuna	*Uvoz, tisuće kuna*
Europa	*47 333 172*	*89 601 154*
Azija	*4 070 337*	*16 181 219*
Afrika	*1 951 041*	*661 073*
Sjeverna, Srednja i Južna Amerika	*1 782 383*	*5 244 855*
Australija i Oceanija	*135 266*	*62 797*
Ukupno	*55 242 198*	*111 751 089*

Tablica 6: Izvoz i uvoz po zemljama namjene/podrijetla (Primjer 4.)

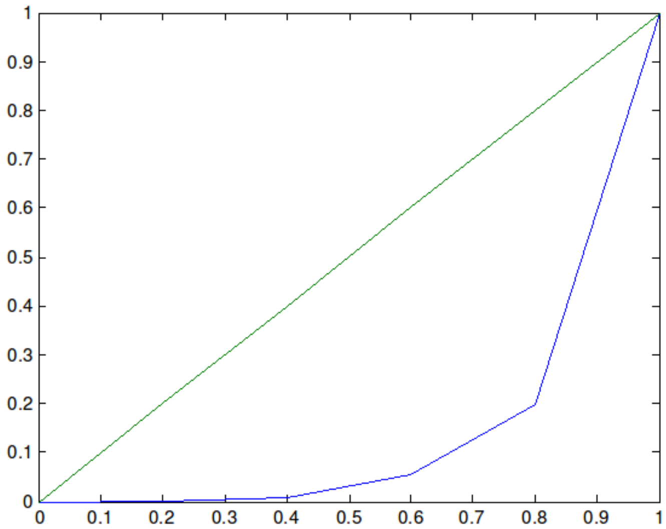
Lorenzova krivulja za izvoz prikazana je na Slici 4.



Slika 3: Lorenzova krivulja za izvoz (Primjer 4.)

Primjećujemo da jedan modalitet nosi većinu totala. Vrijednost Ginijevog koeficijenta je 0.70 ($G^* = 0.88$), što potvrđuje visok stupanj neravnomjernosti u raspodjeli vrijednosti niza. Konkretno, možemo zaključiti da modalitet „Europa" odnosi većinu totala, odnosno najveći dio ukupnog izvoza odlazi u Europu.

Pogledajmo sada Lorenzovu krivulju za uvoz (Slika 5).



Slika 4: Lorenzova krivulja za uvoz (Primjer 4.)

Već je iz slike razvidno da je situacija s uvozom veoma slična. Ginijev koeficijent iznosi 0.71 ($G^* = 0.89$), te opet možemo zaključiti da je raspodjela neravnomjerna i da najveći dio uvoza dolazi iz europskih zemalja.

Bibliografija

[1] I. Šošić, *Primijenjena statistika*, Školska knjiga, Zagreb, 2004.

[2] http://en.wikipedia.org/wiki/Gini_coefficient (15. 9. 2012.)

[3] http://en.wikipedia.org/wiki/Lorenz_curve (15. 9. 2012.)

[4] S. Kurepa, *Matematička analiza 2*, Školska knjiga, Zagreb, 1987.

[5] Statistički ljetopis Republike Hrvatske za 2009. godinu, Državni zavod za statistiku RH, Zagreb, 2010.

[6] Rezultati ankete o potrošnji kućanstava u 2009., statistička izvješća, Državni zavod za statistiku RH, Zagreb, 2011.

---

[1] Zbroj svih vrijednosti statističke varijable uobičajeno se naziva total, pa u nastavku teksta koristimo taj termin

[2] Ne istaknemo li drugačije, pretpostavljamo da je statistički niz konačan

[3] Orris Clemens Herfindahl (1918. – 1972.), američki ekonomist

[4] Albert Otto Hirschman (1915.), američki ekonomist porijeklom iz Njemačke

[5] Corrado Gini (1884. – 1965.), talijanski statističar, demograf i sociolog. Poznat je i kao jedan od vodećih talijanskih teoretičara fašizma i fašističke ideologije

[6] Osim u slučajevima vezanima za Republiku Hrvatsku, svi toponimi i numeričke vrijednosti navedeni u primjerima potpuno su izmišljeni i služe isključivo za ilustraciju osnovnih pojmova i ideja iz teksta

[7] Max Otto Lorenz (1876. – 1959.), američki statističar i ekonomist

ISSN 1334-6083