EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR

Kaszányi Jennyfer

MIT REJT A „CORONA” MASZKJA?

Szakdolgozat
Matematika Bsc, elemző szakirány

Témavezető
Pröhle Tamás
Valószínűségelméleti és Statisztika Tanszék

Budapest, 2020.

NYILATKOZAT

Név: Kaszányi Jennyfer

ELTE Természettudományi Kar, szak: Matematika Bsc

NEPTUN azonosító: Y9W5W1

Szakdolgozat címe: Mit rejt a „Corona" maszkja?

A **szakdolgozat** szerzőjeként fegyelmi felelősségem tudatában kijelentem, hogy a dolgozatom önálló szellemi alkotásom, abban a hivatkozások és idézések standard szabályait következetesen alkalmaztam, mások által írt részeket a megfelelő idézés nélkül nem használtam fel.

Budapest, 2020. december 27.

Kaszányi Jennyfer

a hallgató aláírása

Bevezetés

Váratlan helyzetek, fordulatok folyamatosan előjönnek az emberek életében és a világban. A 2020-as év egy emlékezetes év lesz mindenki számára. Ki gondolta volna, hogy egyszer egy járvány uralni fogja a világot hónapokon át, hatással lesz az emberek mindennapjaira, és az országok gazdaságát is befolyásolni fogja hosszú hónapokon keresztül? Igaz a mondás, miszerint bármi bármikor megtörténhet. Szakdolgozatomat egy olyan témából szerettem volna írni, ami aktuális, hatással van a világra. Jelen helyzetben a koronavírus járvány az, ami nagy mértékben változásokat hoz a Földön.

Köszönetnyilvánítás

Szeretnék köszönetet mondani témavezetőmnek, Pröhle Tamásnak, aki támogatott és lehetővé tette számomra, hogy egy olyan témáról írjak a szakdolgozatomban, ami mai napig uralja és hatással van mindenre a világon. Hálás vagyok, hogy szakértelmével és tanácsaival végig segítette a szakdolgozatom megírását. Köszönet a sok konzultációkért is. Nélküle és tanított tanáraim nélkül ez a szakdolgozat nem jöhetett volna létre. Valamint szeretném megköszönni családtagjaimnak és barátaimnak a sok-sok türelmet és támogatást, amit az egyetemi éveim alatt kaptam tőlük.

Tartalomjegyzék

1. Mi az a korona vírus?

A COVID-19 egy újonnan felfedezett koronavírus, mely a kínai Wuhanból terjedt el 2019 decemberében. 2020.03.11-én a WHO (World Health Organization - Egészségügyi Világszervezet) pandémiává nyílvánította a vírust.

Magyarországon először az első koronavírussal fertőzött beteget 2020.03.04-én regisztrálták, az első koronavírus járványban elhunyt beteget 2020.03.15-én jelentették be.

Jelen helyzetben a második hullámmal veszi fel a harcot az emberiség. Az első hullám Magyarországon márciustól májusig tartott, míg a második hullám augusztusban robbant ki és a mai napig tart.

A járvány legfőképp az olyan idős emberekre jelent veszélyt, akik egészségügyi problémákkal élnek együtt, mint például cukorbetegséggel, rákkal, szív-és érrendszeri betegségekkel. Amennyiben valaki elkapja a koronavírus járványt, tünetként elsőkörben légzőszervi megbetegedést, lázat, fájdalmat és szaglás, ízlelés elvesztését tapasztalhatja. A kormány különféle intézkedéseket vezet be arra vonatkozóan, hogy megelőzzük a járvány terjedését. Ilyen például az, hogy folyamatosan értesülünk a COVID-19 vírusról, embertársaink és saját magunk megvédése az által, hogy egy nap többször is kezet mosunk, fertőtlenítő szereket használunk és maszkot hordunk.

A WHO - World Healt Organization honlapján nyomon lehet követni a napi aktuális számokat:

dátum	igazolt esetek	megerősített haláleset	érintett országok
november 8.	49 578 590	1 245 717	219
november 15.	53 507 282	1 305 164	220
november 22.	57 882 183	1 337 795	220
november 29.	61 869 330	1 448 896	220
december 06.	65 870 030	1 523 583	220
december 13.	70 228 447	1 595 187	220
december 14.	71 051 805	1 608 648	220
december 18.	73 275 943	1 650 348	222
december 21.	75 110 651	1 680 395	222
december 28.	79 232 555	1 754 493	222

2. Gini-féle együttható

A Gini-féle együttható egy közgazdasági mérőszám, mely a statisztikai eloszlás egyenlőtlenségét méri. A Gini-együtthatót egy olasz statisztikus, demográfus és szociológus, Corrado Gini fejlesztette ki az 1900-as évek elején.

2.1. Gini-féle távolság

Legyen az $x = (x_1, ..., x_n)$ tetszőleges, nem feltétlen különböző, nemnegatív valós számok sorozata, és legyen az $y(x) = y = (y_1, ..., y_n)$ sorozat az x sorozat elemeinek növekvő nagyságrend szerint rendezett változatából képzett sorozat. Továbbá legyen, a csak n-től függő konstansokból álló $\ell = (\ell_1, ..., \ell_n)$, szintén n elemű sorozat olyan, hogy az $\ell_i = (i - 0.5)/n - 0.5$, $i = 1, ..., n$-re.

Vegyük a következő két, a fent mondott n hosszú sorozatokon értelmezett függvényt:

$$G(x) = \frac{\sum_{j=1}^{n}\sum_{i=1}^{n} |x_i - x_j|}{2\sum_{j=1}^{n}\sum_{i=1}^{n} x_i}$$

$$G^*(y) = \frac{2\sum_{i=1}^{n} \ell_i \cdot y_i}{\sum_{i=1}^{n} y_i}$$

Definíció

A $G(x)$ valós számot az $x = (x_1, ..., x_n)$ számsor Gini-féle indexének nevezzük.

Állítás

A $G^*(y) = G(x)$, ha az y az x nagyság szerinti növekvő sorrendbe rendezett változata. Azaz, ha $y = (y_1, ..., y_n)$ növekvő nemnegatív számok sorozata, akkor $G^*(y) = G(x)$.

A Gini-féle index tipikus értelmezése, hogy az $x_1, x_2, \ldots, x_n$ számok például egy cég alkalmazottainak az egyéni keresete. Ekkor, mint majd látni fogjuk a Gini-féle index azt méri, hogy az adott cégen belüli keresetek mennyire térnek el az egyenletestől. Az ugyanis a $G(x)$ képletből közvetlenül látszik, hogy a $G(x) = 0$, ha minden alkalmazott ugyanakkora összeget keresne, és ha úgy alakulna, hogy a cégen belül mindenkinek 0 a keresete, egyet kivéve, akkor a $G(x) = 1$ lenne.

Másik tipikus alkalmazása a Gini-féle indexnek, ha az x értékek az egyes egyének vagyonát jelentik. Ez a fajta interpretáció azért hasznos, mert a Gini-féle index elvileg olyan esetekben is alkalmazható, amikor nem mindegyik x érték pozitív. Egy vagyon lehet negatív az adósságok miatt, ám a kereset esetén a negatív kereset szokatlan volna.

A szakdolgozatban a Gini-indexet annak mérésére fogjuk használni, hogy a ragály következtében elhunyt személyek életkora mennyiben tér el az egyenletestől, tehát attól, hogy az egyes korosztályok azonos mértékben volnának a ragály-halállal érintve. Ebben az értelmezésben a korosztályok veszik át az előző értelmezések szerinti alkalmazottak szerepét, és a vagyon az adott életkorban elhalálozottak száma lesz.

2.2. A Gini-féle index négyféle számolási módja

A Gini-féle indexet egy rendelkezésre álló x sorozatból sokféleképpen ki lehet számolni. Ezek közül négyet emelünk ki. Azt, amelyik szerint a Gini-index

- az abszolut differenciák összege
- egy ekvidisztáns sorozattól vett távolság
- saját magától vett reverz-távolság
- egy ekvidisztáns sorozattal vett korreláció

E számolási módoknak az érdekességen és a számolási prakticitáson túl az a jelentősége, hogy mindegyik a Gini-indexnek más-és-más tulajdonságára világít rá.

A számolási módok könnyebb bemutathatósága kedvéért néhány, a továbbiakban felhasznált jelölést vezetünk be.

Jelölések

Legyen $(x_1^+,\dots,x_n^+)$ az x sorozat növekvő nagyságrend szerint rendezett változata. Ezt a sorozatot korábban y-nal jelöltük. Most azért alkalmazunk újabb jelölést, mert szükségünk van az x sorozat csökkenő sorrendbe rendezett $(x_1^-,\dots,x_n^-)$ változatára is.

Ismételten szükségünk lesz a x^+ és az x^- tagjaiból képzett szukcesszív összegek sorozatára, azaz az

$$s_k^+ = \sum_{j=1}^{k} x_j^+ \qquad k = 1,\dots,n$$

és az

$$s_k^- = \sum_{j=1}^{k} x_j^- \qquad k = 1,\dots,n$$

továbbá az

$$s_k = \sum_{j=1}^{k} x_j \qquad k = 1,\dots,n$$

elemekből álló n elemű s^- és s^+, illetve s sorozatokra.

Szukcesszív összegek relatív értékét jelölje:

$$r_k^+ = \frac{s_i^+}{s_n} \qquad k = 1,\dots,n$$

illetve

$$r_k^- = \frac{s_i^-}{s_n} \qquad k = 1,\dots,n$$

Jelölje az x sorozat átlagát $\overline{x} = \dfrac{s_n}{n}$.

Megjegyzés

Nyilvánvaló, hogy $s_n^+ = s_n^- = s_n$, továbbá, hogy $s_1^+ = \min(x)$ és $s_1^- = \max(x)$,

és ezekből az is következik, hogy a rendezett sorozatok szukcesszív összegeinek relatív változata olyan, hogy az utolsó elemekre $r_n^- = r_n^+ = 1$.
Az x^- és x^+ rendezett sorozatok átlaga mindig egyenlő $\bar{x}$-al.

További jelölések

A definíciók során kétféle, n hosszú „létrát” fogunk alkalmazni. Ezek egymás konstanssal való eltoltjai.

Legyen

$$e_k = \frac{k}{n}, \qquad \text{ahol } k = 1, \dots, n$$

és legyen

$$d_k = \frac{k - 0.5}{n} - 0.5, \qquad \text{ahol } k = 1, \dots, n$$

Megjegyzés

Mint látható, mindkét létra $1/n$ lépésközű.
Közülük az e, az első 1-ig tartó, az utóbbi, a d a 0-ra szimmetrikus, az alapdefiníció D^* formában megadott variánsában is felhasználtuk, ℓ jelöléssel.

2.2.1. Abszolút differencia

A Gini-index megegyezik az abszolút differenciák összegével, ami

$$A(x) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j|}{2ns_n}$$

Ha a számlálóban a tagok nem csak abszolút értékben lennének, hanem négyzetre is lennének emelve, akkor a számlálóban lévő összeg egy a tapasztalati szórásnégyzettel arányos mennyiséget adna meg. Tehát azt a következtetést vonhatjuk le, hogy a Gini-féle index szóródási mérőszám is.

Ez az $A(x)$ képlet nyilvánvaló módon azonos eredményt ad a Gini-index definíciója szerinti $G(x)$ képlettel.

2.2.2. Ekvidisztánstól vett távolság

A Gini-index megegyezik azzal az $E(r^+)$ távolság differencia összeggel, ami az r^+ növekvő relatív keresetek szukcesszív összeg sorozata és az e, 1-ig tartó, ekvidisztáns, $\frac{1}{n}$-közű létra elemei közt észlelhető:

$$E(r^+) = \frac{2}{n}\sum_{k=1}^{n}(e_k - r_k^+)$$

Vegyük észre, hogy az utolsó különbség mindig 0. Továbbá, mivel az r^+ növekvő sorozat szukcesszív összege, a szummán belüli különbségek nemnegatívak. Tehát az abszolútérték jel a képletből nem hiányzik. Ebből az a következtetés is adódik, hogy a Gini-index dupla szummás definíciós képlete átrendezhető egy sokkal hatékonyabban számolható formára is. Ugyanakkor ennek a képletnek az alapján is közvetlenül látszik, hogy a Gini-index pontosan akkor 0, ha minden x_k (mindenki jövedelme) egyenlő lenne. Ugyanis pontosan ekkor lesz az r^+ is egy $1/n$ közű, 1-ig tartó létra.

A képlet azt is megmutatja, hogy miért méri jól a Gini-féle index a jövedelem tagozódást. Ugyanis a benne szereplő távolság összeg azt méri, hogy a szegényebb rétegek együttes jövedelme mennyivel marad el attól, amennyi kereset nekik egyenlő keresetek mellett jutna.

Nyilvánvaló, hogy ez a Gini-index felírás a Lorenz-görbe alapján vett – későbbiekben bemutatott – értelmezésének felel meg.

2.2.3. Reverz-távolság

Az ellentétesen rendezett normált szukcesszív összegek differenciájának átlaga:

$$D(r^+, r^-) = \frac{\sum_{k=1}^{n} (r_k^+ - r_k^-)}{n}$$

Következménye az előző képletnek.

Ez a képlet a Gini-index egy másfajta értelmezési lehetőségét jelenti.

Az $\left(r_k^+ - r_k^-\right)$ differencia minden $k = 1, \ldots, n$-re azt mutatja meg, hogy a k fő legszegényebb és a k fő leggazdagabb összkereset-hányada közt mekkora a különbség.

A reverz távolság szerinti felírás a Lorenz görbés interpretációjában is értelmezhető. Azt a területet kell venni, amelyet a Lorenz görbe és az a görbe határol, amelyet a Lorenz görbének az $\left(\frac{1}{2}, \frac{1}{2}\right)$ pontra vett centrális tükrözése ad. Ellentétben annak a területnek a kétszeresével, aminek a Gini-index az ekvivalens távolság szerint felel meg. Ami ugyanis a $(0,0) - (1,1)$ négyzet origón átmenő átlója és a Lorenz-görbe közti terület.

2.2.4. Ekvidisztánssal vett korreláció

A Gini-index az átlagos aránytól való növekvő eltérés sorozatnak és a 0-ra szimmetrikus $\frac{1}{n}$ közű d létrának a skalárszorzata:

$$R(x^+, d) = 2\sum_{k=1}^{n} \left(\frac{x_k^+}{s_n} - \frac{1}{n}\right) d_k$$

Vegyük észre, hogy a zárójelben szereplő $\frac{x_k^+}{s_n}$, $k = 1, ..., n$ sorozat szukcesszív összege az r^+ sorozatot adja. Továbbá, mivel $\sum_{k=1}^{n} d_k = 0$ a $\sum_{k=1}^{n} \frac{1}{n} d_k = 0$, a képlet egyszerűsíthető. Megkapjuk belőle az alapdefiníció D^* variánsát.

Azért vettük mégis ezt a bonyolultabb képletet, mert az $x_k^+ \big/ s_n$ hányados azt adja meg, hogy a k. alkalmazott az összjövedelem hányad részét kapja meg. És így a zárójelbeli összeg azt adja meg, hogy az alkalmazottak az átlagos $1\big/n$ jövedelem hányadtól mennyivel térnek el.

Ebben az interpretációban a Gini-index tehát annak a korrelációnak felel meg, ami a szimmetrikus létra és a jövedelmek átlagostól való eltérése között van.

A Gini-index $R(x^+, d)$ korrelációs kifejezése a

$$C(x) = \frac{2\sum\limits_{k=1}^{n} k x_k^+}{n s_n} - \frac{n+1}{n}$$

egyszerűbb formában is felírható, ekvivalens átalakításokkal.

2.3. A Gini-index diszkrét eloszlások esetén

A Gini-index akkor is értelmezhető, ha az $x_1,...,x_n$ sorozatról nem tesszük fel, hogy a tagjai különbözőek. Figyelembe véve a Gini-index skála invarianciáját nyilvánvalóan adódik, hogy a Gini-index definíciója természetes módon terjeszthető ki tetszőleges diszkrét eloszlásra is.

Legyen $D(P,X) = \big((p_1, \ldots, p_k, \ldots), (x_1, \ldots, x_k, \ldots)\big)$ egy tetszőleges diszkrét eloszlás, azaz legyen $\mathbb{P}(x_k) = p_k$, $k = 1, 2, ...$-re. Ennek az eloszlásnak a várható differenciája, definíció szerint:

$$D(P,X) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} p_i p_j |x_i - x_j|$$

A $D(P,X)$ nyilván megegyezik a $\mathbb{E}|\xi_1 - \xi_2|$ várhatóértékkel, ha a ξ_1, ξ_2 két egymástól független $D(P,X)$ eloszlású mennyiség. Ugyanakkor megegyezik azzal a differencia összeggel is, amivel a Gini-féle távolságot tetszőleges megfigyelés sorozatok esetén definiáltuk.

2.4. A Gini-index dinamikája

Alkalmazzuk a Gini-féle index eredeti definícióját, azaz a

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

képletet!

Tegyük fel, hogy a vizsgált x_k, $k = 1,...,n$ értékek közül valamely $i \neq j$-re olyan $x_i' = x_i + \varepsilon$, $x_j' = x_j - \varepsilon$ módosulás történik, aminek következtében az x-ek sorrendje nem változik. Könnyen látható, hogy ekkor a Gini-féle együttható változásának a mértéke

$$d_\varepsilon(i,j) = \frac{2\varepsilon}{\bar{x}n(n-1)}(j-i)$$

Vagyis a Gini-együttható változásának mértéke lineárisan függ attól, hogy az átcsoportosítás a vagyon szerint populáció milyen távoli tagjai közt történik.

Alkalmazzuk a Gini-féle index korrelációs definíciójából közvetlenül adódó

$$G = \frac{\sum_{k=1}^{n}(2k-(n-1)x_k)}{\bar{x}n(n-1)}$$

formuláját.

Tegyük fel, hogy a vizsgált értékek közül valamely ℓ-re csak egy x érték változik egy olyan kis ε-nal: $x_\ell' = x_\ell + \varepsilon$, hogy az x-ek sorrendje ettől nem változik meg. Ismét csak könnyen látható, hogy ekkor a Gini-féle együttható új értéke, céljainknak megfelelően ekvivalens módon átalakítva:

$$G' = \frac{\sum_{k=1}^{n}(2k-(n-1)) + \varepsilon(2\ell-(n-1))}{(n-1)(n\bar{x}+\varepsilon)}$$

Látható, hogy a megváltozott Gini értékben a megváltozott x-érték ℓ indexétől csak a számláló második tagja függ. Ez a tag az ℓ lineáris függvénye. Ha $\varepsilon > 0$, akkor ha az

$\ell < \frac{n-1}{2}$ akkor csökken a Gini-index értéke, ha $\ell > \frac{n-1}{2}$ akkor növekszik. Akkor csökken a Gini-index értéke legjobban, ha a legszegényebb vagyona nő, és akkor növekszik a legnagyobb mértékben a Gini-index, ha a leggazdagabb vagyona nő. Ugyanez az ε mértékű vagyon gyarapodás a medián $\ell = \frac{n-1}{2}$ személy esetén nem okoz változást a Gini-index mértékében.

2.5. A Gini-féle index karakterizációja

Négy olyan tulajdonságot definiálunk, ami egy megfigyelés sorozathoz rendelt $I(.)$ index karakterizáló tulajdonsága lehet.

1. **Tulajdonság:** skála invariancia

   Tetszőleges $x_1, x_2, \ldots, x_n \in \mathbb{R}$ megfigyelés sorozatra

   $$I(\lambda x_1, \ldots, \lambda x_n) = I(x_1, \ldots, x_n)$$

   tetszőleges $\lambda \in \mathbb{R}$ konstans esetén.

2. **Tulajdonság:** szimmetria

   Tetszőleges, n elemen értelmezett $\mathscr{P}$ permutációra az $I(x)$ invariáns. Azaz

   $$I(\mathscr{P}(x)) = I(x)$$

   tetszőleges $x \in \mathbb{R}^n$ esetén.

3. **Tulajdonság:** arányosság

   Az index értéke a konvex lineáris kombinációkkal felcserélhetők. Azaz

   $$I(E_k) = \frac{k}{n},$$

   ahol E_k a k. csúcspontot jelöli, tehát $E_k = (\underbrace{0, \ldots, 0}_{k}, \underbrace{1, \ldots, 1}_{n-k})$.

4. **Tulajdonság:** linearitás

   Az index értéke a konvex lineáris kombinációkkal felcserélhető. Azaz

$$I(\lambda u+(1-\lambda)v)=\lambda I(u)+(1-\lambda)I(v)$$

   ahol $\lambda \in [0,1]$ és az u és v két tetszőleges nemnegatív n elemű számsor, amire

$$\sum_{k=1}^{n} u_k = \sum_{k=1}^{n} v_k$$

2.5.1. Tétel

Az $x=(x_1',\dots,x_n') \in \mathbb{R}^n$ *megfigyelés sorozatokhoz tartozó*

$$G(x)=\left(n+1-2\cdot\frac{\sum\limits_{k=1}^{n}(n+1-k)x_k^*}{\sum\limits_{k=1}^{n}x_k^*}\right)$$

Gini-féle index, ami az egyetlen olyan $\mathbb{R}^n \longmapsto \mathbb{R}^n$ *függvény, amely bír a fenti 4 tulajdonsággal, azaz skála invariáns, szimmetrikus, a csúcsokon arányos és lineáris.*

Bizonyítás. Legyen

$$K=\left\{y \in \mathbb{R}_+^n \mid \sum_{k=1}^{n} y_k = 1 \text{ és } y_1 \le y_2 \le \dots \le y_n\right\}$$

Belátható, hogy szimmetrikus és skála invariáns esetben elég a problémát csak a rendezett, 1 összegű, azaz a K-beli sorozatokra vizsgálni. Az $y \in K$ sorozatok sorbafejthetőek azon $v^1,\dots,v^j,\dots,v^n$ sorozatok szerint, amelyeknek az $i<j$ elemei nullák, és a többi $v_i^j{=}\frac{1}{(n+1-j)}$. Ráadásul a sorfejtés $\alpha \in \mathbb{R}^n$ együtthatóira $\sum\limits_{i=1}^{n}\alpha_i=1$ és az $\alpha_i=(n+1-i)(y_i-y_{i-1})$, ahol $i=1,\dots,n$ is teljesül.

Ugyanakkor n szerinti indukcióval belátható, hogy n darab, K-beli, tetszőleges $w^1,\dots,w^i,\dots,w^n$ vektorra az $I\left(\sum\limits_{i=1}^{n}\beta_i w^i\right)$ súlyozott összeg az I leképezéssel felcserélhető, egyenlő $\sum\limits_{i=1}^{n}\beta_i I(w^i)$-vel.

Vagyis a skála invariancia és az arányosság a csúcspontokon, a konvexitás és a lineáris pedig a belső pontokon határozza meg a Gini-indexet. Viszont a mondottak alapján a tétel következik. □

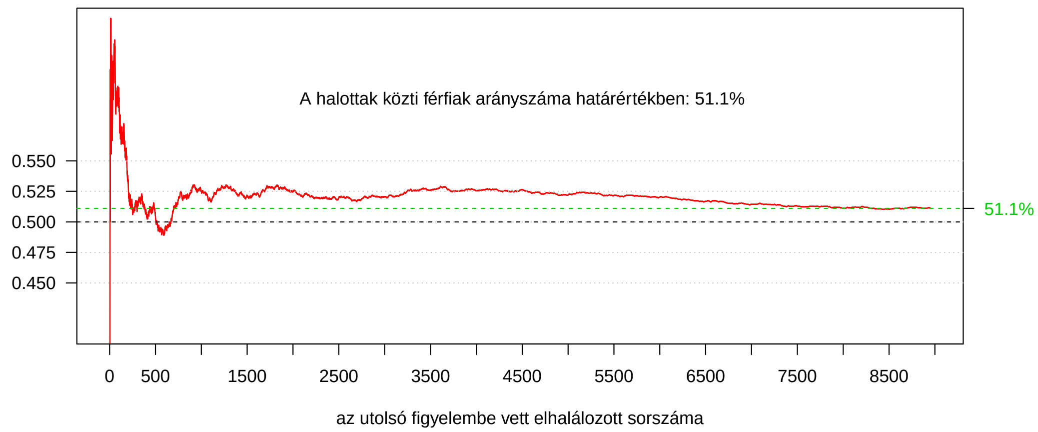
2.6. A Gini-index alkalmazása a 2020-as járvány adataira

A 2020-as magyarországi járvány halálozási adatait a kormány ".gov" portálja részletesen közli, feltüntetve az elhaltak nemét és korát. Közli az esetlegesen volt krónikus betegségüket is, de ez az elemzésünk szempontjából irreleváns.

Két adatsort használunk fel, az egyik a 2020. december 27-ig elhunyt 8951 fertőzött neme és kora, az elhalálozásuk sorrendjében. És azt, amelyik megmondja, hogy a járvány egyes, első 298 napján hány halott adódott.

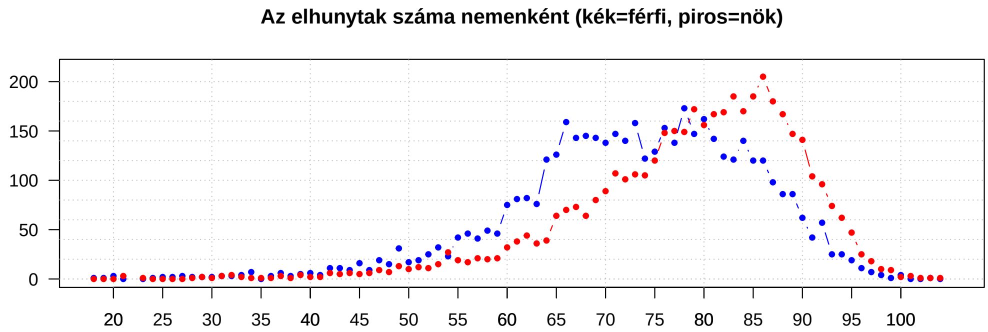
Vizsgálatainkat előbb 4, az adatokat közvetlenül leíró ábrával, majd 2, a koreloszlás Gini-féle mértékének időbeli változását bemutató ábrával dokumentáljuk.

Az első vizsgálat tárgya az elhaltak közti férfi+nő arány változása az első 8951 halott alapján.

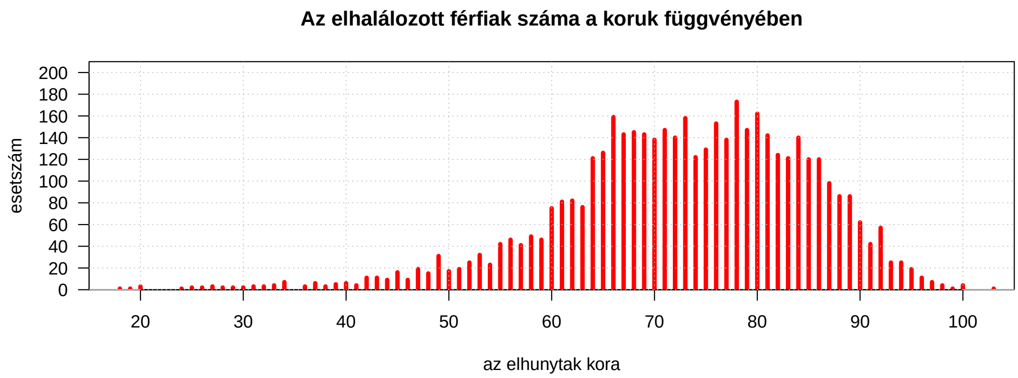


A közölt ábra azt mutatja, hogy a férfi halottak száma lényegében mindig magasabb volt, mint a női halottaké. Ez a megállapítás azért érdekes, mert a közhiedelem szerint a ragály általi halál inkább az időseket fenyegeti, és mint ismeretes a társadalmon belül sokkal több az idős hölgy, mint férfi. Vagyis ez utóbbi ellenére magasabb a férfi halottak száma, mint a nőké. Érdekes továbbá az, az ugyanezen ábrán látható további tény, hogy noha a férfi halálozási arány az 51.1%-hoz látszik konvergálni, a tényleges férfi halálozási arány lényegében mindig ez az arány felett volt.

Ezen megállapítások miatt érdemesnek látszik megvizsgálni, hogyan alakul az első 298 nap alapján a meghalt férfiak és nők koreloszlása.



Külön, csak a férfiak koreloszlása.



Külön, csak a nők koreloszlása.

Összevetve a női és férfi elhalálozottak koreloszlását, három szembeötlő dolgot állapíthatunk meg. Míg a női halálozás tipikusan a 80-90 éveseket érinti, addig a férfi halálozások tipikusan a jóval szélesebb és fiatalabb 65-80-as korosztályt érintik. Harmadik, társadalmilag nem nagyon számontartott, ugyanezekről az ábrákról leolvasható tény, hogy különösen a férfiaknál, de a nők esetén sem elhanyagolható a 60 évesnél fiatalabbak halálozással való érintettsége.

A dolgozat tárgya szempontjából a következő két ábra a legérdekesebb. Ezen ábrák segítségével azt vizsgáltuk, hogyan változik az elhunytak kora a járvány idejének múlásával.

A férfi és női elhalálozásokat külön értékelve az összes adatot figyelembe véve azt kaptuk, hogy a míg a férfi Gini-index 9.02%, addig a női 7.69%. Ezek viszonylag alacsony értékek, nyilván amiatt vannak, hogy a halálozások nem koncentrálódnak egy-egy korosztályra. De a két index közt kapott viszonylag nagy – 1.33 százalékpont – különbség megfelel azoknak a korábbi megjegyzéseinknek, amelyeket az eloszlások közvetlen vizsgálatakor tettünk.

Előzetes vizsgálatok alapján úgy döntöttünk, hogy a koreloszlás Gini-indexét egymásutáni 42 napokon vizsgáljuk az első naptól kezdve csúsztatva, az utolsó 42 napig. Így összesen idősoronként 298-42=256 időablakban vizsgáltuk meg a Gini-index értékét. Az eredményeket az alábbi két ábra mutatja.

A férfi koreloszlás Gini-indexének változása a járvány első 298 napja alatt,
42 napos szakaszokban értékelve.

A női koreloszlás Gini-indexének változása a járvány első 298 napja alatt,
42 napos szakaszokban értékelve.

Jól látható, a két görbe közti különbség: a férfiaké sokkal nagyobb ingadozásokat mutat és van egy szakasza, amikor a halálozás eloszlása összehúzódik, míg ugyanebben az időszakban a női halálozás inkább jobban szétterül a korosztályok közt.

3. Lorenz-görbe

A Lorenz-görbét tipikusan az egy gazdaságon belüli jövedelmek megoszlásának bemutatására használjuk. A Lorenz görbét Max O. Lorenz 1905-ben fejlesztette ki.

M. O. Lorenz 1905-ben a Pub. of ASA-ban „Methods of Measuring the Concentration of Wealth" címmel megjelent cikkében azt vizsgálta adózási adatokat elemezve, hogy mit is mondhatunk az egyes osztályokbeli jólét változásáról.

Számos korábbi szerző módszerét megvizsgálta, és arra a következtetésre jutott, elsősorban az USA-beli és porosz adatokat elemezve, hogy egyik, addig alkalmazott módszer sem megfelelő a helyzet bemutására.

Bevezeti az azóta róla elnevezett Lorenz-görbét és alkalmazza azt a porosz és mesterséges adatokra. Megállapította azt, hogy egyrészt a görbe egyértelműen választ ad arra, hogy Poroszországban a jövedelem koncentráció 1892 és 1901 között növekedett. Hiszen a görbe, melyet 1901-ben alkalmaztak, majorálja, azaz többre értékeli az 1892-es adatokat.

Az egyértelmű növekedés, az egyértlemű majorálás ténye azt is jelenti, hogy a megfelelő görbék nem metszik egymást. Ezzel kapcsolatban megmutatja a mesterséges adatokon azt, hogy a görbék metszésének a hiánya nem szükségszerű. Viszont azt láthatjuk, hogy a metsző görbék is alkalmasak az elemzést megalapozó statisztikának. Hisz, hogyha a görbék a jövedelem egy sávjában végig majorálják, illetve minorálják egymást, akkor az, az adott jövedelem sávon belül jövedelem koncentrációt vagy kiegyenlítődést jelent.

Megjegyezendő, hogy Lorenz ebben a cikkében „kumulált összjövedelem hányad x kumulált populációs hányad" ábrát alkalmaz, ami az xy-koordináta rendszerben szükségszerűen az origóhoz csatlakozó egységnégyzet átlója felett halad. Ennek megfelelően

Lorenz alulról konkáv, növekvő görbékkel dolgozik, ellenben a róla elnevezett, manapság használatos konvex függvényhez képest.

3.1. Lorenz-görbe definíciója

A Lorenz görbét maga M. O. Lorenz a következő módon adja meg: "*plot along one axis cumulated per cent of the population from the poorest to the richest, and along the other axis the percent of the total wealth held by these percents of the population*". Értelemszerű fordításban: "*rajzoljuk az egyik tengelyen az összesített populáció hányadot a legszegényebbtől a leggazdagabbig, a másik tengelyen pedig az összvagyonát ugyanezen populáció hányadnak*". Kicsit másképpen fogalmazva: Tekintsünk egy populációt, amelyre ismert az egyénenkénti vagyon mértéke. A Lorenz-görbe egy paraméteres görbe, amelynek paramétere az egyénenkénti vagyon lehetséges mértéke. A görbe két koordinátája pedig legyen egyfelől az a populációhányad, akiknek legfeljebb egy adott mértékű vagyona van. A másik pedig az az összvagyon hányad, amelyet a populációnak az a része birtokol, akiknek az egyenkénti személyes vagyona nem haladja meg a szóbanforgó paraméter értéket.

Tehát, ha a Lorenz-görbe például átmegy az $(\frac{1}{2}, \frac{1}{3})$ ponton, akkor az azt jelenti, hogy a populáció szegényebbik fele a vagyon egyharmadát, következés képpen a gazdagabbik fele a vagyon kétharmadát birtokolja. Ez az a tulajdonság, aminek alapján az eredeti definíció, az $\frac{1}{2}$ érték változtatása mellett a Lorenz-görbét megadja. Ugyanakkor a Lorenz-görbe semmilyen információt nem ad arról, hogy mi az a paraméter érték, azaz vagyon mérték, aminek alapján az adódott, hogy a populáció szegényebbik fele az összvagyon egyharmadát birtokolja. Ugyanis a Lorenz-görbe alapján nem tudjuk megmondani, hogy például a populáció szegényebbik felébe tartozóknak mennyi a ténylegesen maximális személyenkénti vagyona.

A Lorenz-görbe formális definíciója igen egyszerű. Ha a populációban a vagyon vagy más nemnegatív számmal jellemezhető mennyiség mértéke az F eloszlás szerinti, akkor ennek a vagyon-eloszlásnak Lorenz-függvénye egy a $[0,1]$ intervallumból a $[0,1]$ intervallumba képező függvény. Formális definíció a következő:

3.1.1. Definíció

Az F eloszláshoz tartozó $L(p) = L_F(p)$ Lorenz-görbe a

$$L(p) = \mu^{-1} \int_{t=0}^{p} F^{-1}(t)dt$$

képlet szerinti, minden $p \in [0,1]$ értékre, ahol a μ az F eloszlás várhatóértéke.

Megjegyzés

Nem feltétlen szükséges kikötni, hogy a vagyon nemnegatív szám legyen. Csak mint majd látni fogjuk, esetlegesen negatív értékű eloszlások esetén a Lorenz-görbe számos fontos tulajdonsága módosul.

Nyilvánvalóan ekvivalens, alternatív definíció a következő.

3.1.2. Definíció

Legyen $x = x_p$ az F eloszlás $p \in [0,1]$ valószínűséghez tartozó kvantilise, azaz legyen

$$p = F(x) = \int_{t=0}^{x} f(t)dt$$

Azaz p annak a részpopulációnak a hányada, akiknek a vagyona legfeljebb x. Az így megadható p értékre az $L(p)$ Lorenz-görbe az a függvény, amelyre:

$$L(p) = \int_{t=0}^{x_p} tf(t)dt \Big/ \mu$$

Megjegyzés

Látható, hogy a Lorenz-függvény tulajdonképpen a várhatóérték függvény transzformáltja, hiszen a $\int\limits_{t=0}^{x} tf(t)dt$ a várhatóértéknek az x-ig tartó részét adja meg, és például $\int\limits_{t=0}^{\infty} tf(t)dt = \mu$, az $f(x)$ eloszlás várhatóértéke.

3.2. Tulajdonságai

A Lorenz-görbe az alábbi tétel szerinti tulajdonságokkal rendelkezik:

3.2.1. Tétel

Ha az $L(p)$ Lorenz-görbe az F nemnegatív eloszláshoz tartozik, akkor az $L(p)$

- *konvex*
- *monoton növekedő*
- *minimuma 0, maximuma 1*
- *skála invariáns*
- *nem eltolás invariáns*
- *a $\mathscr{U}[0,\vartheta]$ eloszlásokhoz az $L(p) = p$ görbe tartozik.*

A Lorenz-görbék felsorolt tulajdonságai könnyen beláthatóak, a fenti definíciók alapján.

Megjegyzés

Eltolás esetén a következő formula érvényes. Az F eloszláshoz tartozó L_F és a c értékkel eltolt $F_c(x) = F(x-c)$ eloszláshoz tartozó $L_{F,c}$ Lorenz-görbéket vizsgálva a következő érvényes:

$$\mu_c\left(F_c(x) - L_{F,c}(F_c(x))\right) = \mu\left(F(x) - L_F(F(x))\right)$$

ahol μ az F, a $\mu_c = \mu + c$ pedig az F_c eloszlás várhatóértéke.

Az alábbiakban három példa, illetve ábra segítségével bemutatjuk, hogyan viselkednek a Lorenz-görbék a gyakorlatban.

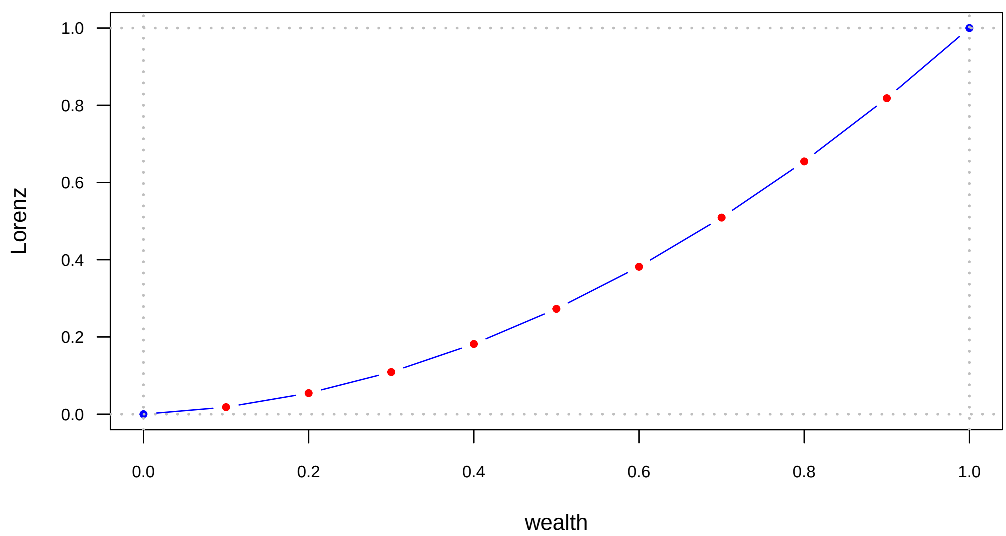
1. Példa

Először megmutatjuk, milyen annak a Lorenz-görbének a formája, amelyik ahhoz a teszt adatsorhoz tartozik, amely szerint a populáció 10 deciliséhez tartozó személyek vagyona a decilistől függő homogén lineáris érték. Az egyszerűség kedvéért úgy vesszük, hogy mindegyik decilishez mindössze 1 személy tartozik. Ennek megfelelően a tesztadatunk az alábbi táblázat szerinti.

réteg	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
n	1	1	1	1	1	1	1	1	1	1
y	1/10	2/10	3/10	4/10	5/10	6/10	7/10	8/10	9/10	10/10

tesztadatokhoz tartozik. Itt tehát az y az adott rétegbeliek vagyonát, az n pedig az adott rétegbeliek össz létszámát jelentik.

A Lorenz-görbe skálainvarianciája miatt az nyilván nem változtatna a kapott görbén, ha úgy vennénk, hogy minden egyes réteghez N személy tartozik, ahol N egy tetszőlegesen rögzített nemnegatív egész (vagy valós szám).



Az ábrát az

```
n <- rep(1,10) # gyakoriság mind a 10 rétegben 1 személy
y <- (1:10)/10 # a vagyon homogén lineárisan nő
# az ineq::Lc szerinti Lorenz fuggveny
L <- as.matrix(data.frame(Lc(y,n)[1:2]))
L
szin<-c("blue",rep("red",9),"blue")

plot(L,t="b",col=szin,pch=20,xlab="wealth",ylab="Lorenz")
abline(h=c(0,1),v=c(0,1),lty=3,lwd=2,col="gray")
```

R-programmal hoztuk létre. Az **R** rendszer `ineq` kiegészítő csomagját használtuk fel. Tulajdonképpen teljesen fölöslegesen, hiszen a felhasznált `Lc()` függvénye, a megadott (y,n) adatokon mindössze a következő 4 utasítással ekvivalens transzformációt hajtotta végre.

```
n <- n[order(y)]              ; y <- sort(y)
y <- c(0,cumsum(n*y)/sum(n*y)); n <- c(0,cumsum(n)/sum(n))
```

Ugyanakkor igaz, hogy az `ineq::Lc()` függvény a fentiekkel ekvivalens sorokon kívül számos formai ellenőrzést is tartalmaz. Ezért a továbbiakban is az `ineq::Lc()` függvényt — lévén az a fentieknél robusztusabb — fogjuk alkalmazni.

---

2. Példa

Az eltolás hatása a Lorenz-görbére. A fenti definíciót nemnegatív F eloszlások mellett adtuk meg. Ugyanakkor a Lorenz-görbe tetszőleges eloszlások mellett is ugyanúgy értelmezhető. Ez a megjegyzés azért szükséges, mert az alábbiakban azt fogjuk vizsgálni, hogyan változik meg a Lorenz-görbe, ha a megfelelő eloszlást egy adott c értékkel eltoljuk.

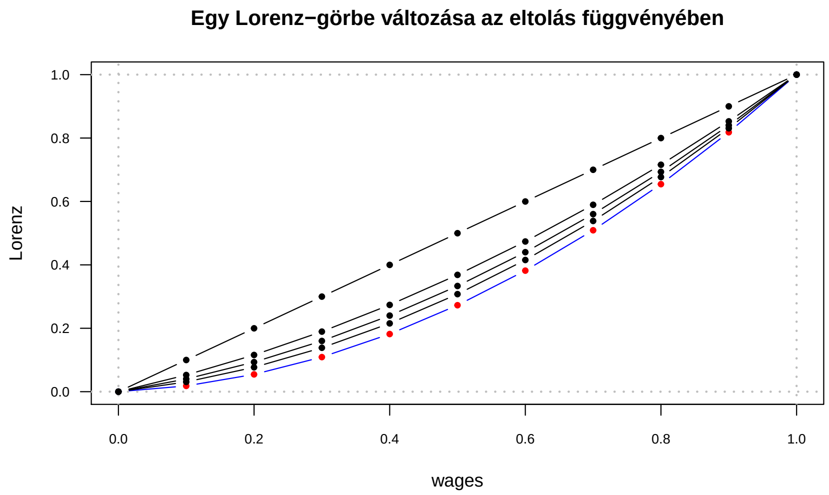
A mondott eltolás annak felel meg, hogy az $F(x)$-hez tartozó görbe helyett az $F_c(x) = F(x-c)$ függvényhez tartozó görbét vizsgáljuk. Ez az eltolás annak felel meg, hogyha ξ véletlen mennyiség az F eloszlás szerinti, akkor a $\xi + c$ véletlen mennyiség F_c szerinti. Azaz mintha a populáció mindenegyes tagjának a vagyona a c konstans értékkel módosulna.

```
LC <- function(w,f,plus=FALSE)
       { if(plus) LC0(w,f)
         else
  { plot(as.matrix(data.frame(Lc(w,f)[1:2])),
                 t="b",pch=20,col=szin,las=1,
                 cex.axis=.66,cex.axis=.75,
                 xlab="wages",ylab="Lorenz",main=cim)
            abline(h=c(0,1),v=c(0,1),
                 lty=3,lwd=2,col="gray")} }
LC0 <- function(w,f,szin="black",jel=20)
   points(as.matrix(data.frame(Lc(w,f)[1:2])),
           t="b",col=szin,pch=jel)
```

Az ábránkon azt fogjuk vizsgálni, hogy hogyan változik a Lorenz-görbe, ha a rétegenkén érvényes $1/10,...,10/10$ személyi vagyont rendre 1/10-del, 2/10-del, 4/10-del, illetve nagyon sokkal (azaz most éppen 1000-rel) megnöveljük?

```
cim<- "Egy Lorenz-görbe változása az eltolás függvényében"
LC(y,n)
LC(y+1/10,n,TRUE)
LC(y+2/10,n,TRUE)
LC(y+4/10,n,TRUE)
LC(y+1000,n,TRUE)
```

A fenti utasítások az alábbi ábrát hozzák létre:



Itt a legalsó görbe az eredeti adatok Lorenz-görbéje. Látható, hogy a további görbék monoton egymás fölött helyezkednek el. A legnagyobb, az 1000-es eltoláshoz tartozó görbe (kissé elkülönülve), csak szemmel nem is alig látható mértékben tér el a 45 fokos egyenestől. Hiszen az rétegenkénti $1000.1, 1000.2, \ldots, 1001$ vagyonérték is csak csekély mértékben tér el az egyenletestől.

Ha a fenti Lorenz-görbe sorozatot a hozzájuk tartozó Gini-együttható alapján vizsgáljuk, akkor nyilvánvaló, hogy a megfelelő Gini-együttható sorozat csökkenő értékű. Ami logikus viselkedés, hiszen fokozatos azonos mértékű vagyon növekedés relatív kiegyenlítődést eredményez. Ugyanakkor fontos észrevenni, hogy a Gini-féle együttható csökkenése a legnagyobb mértékben a középső osztályok vagyonnövekedése által következik be. Ott a legnagyobb a lecsípett görbe feletti területből.

---

3. Példa

Vizsgálatunk tárgya e példa során az, hogy a vagyon értékek előjelének és az átlag vagyon előjelének mi hatása a Lorenz-görbére.

Hat görbét rajzolunk fel. A hat görbéhez tartozó eloszlásnak, a hozzájuk tartozó Lorenz-görbe szempontjából a következők a fő jellemzői:

$+$ jelű: minden vagyon pozítív értékű

$-$ jelű: minden vagyon negatív értékű

$\Uparrow$ jelű: a vagyonok közt van negatív, de a vagyon átlag pozitív

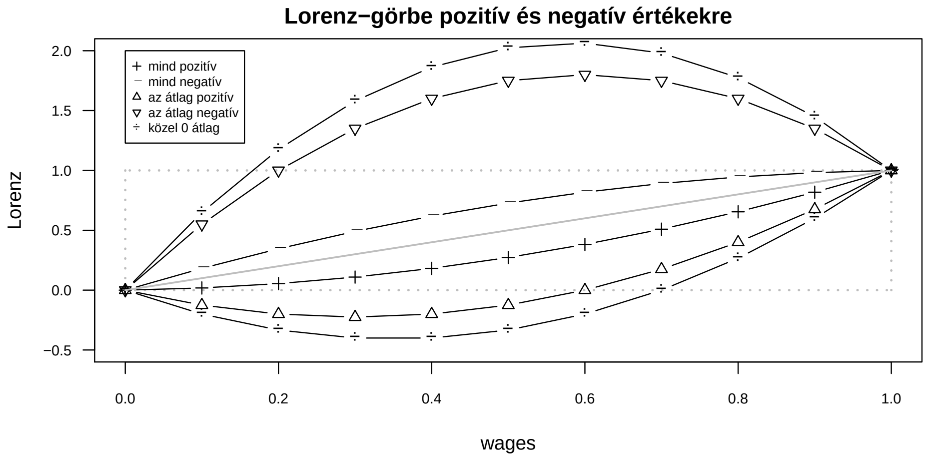
$\Downarrow$ jelű: a vagyonok közt van pozitív, de a vagyon átlag negatív

% jelű: a vagyonok vegyes előjelűek, az átlag 0-hoz közeli pozitív

% jelű: a vagyonok vegyes előjelűek, az átlag 0-hoz közeli negatív

A bemutatott görbék most speciálisan ugyanannak az eloszlásnak az transzformáltjai, de ennek a bemutatandó tulajdonság szempontjából nincsen jelentősége.

```
m <- mean(y*n) ; m # .55
plot(0,0,xlim=c(0,1),ylim=c(-.5,2),t="n",
         xlab="wages",ylab="Lorenz",main=cim)
segments(c(0,1,1,0),c(0,0,1,1),c(1,1,0,0),c(0,1,1,0))

LC0(y,n,jel=3) # mind pozitiv (plus jel)
LC0(-y,n,jel=175) # mind negativ (minus jel)

y0 <- y-m # a megfigyelések 0 átlagba tolva

# van pozitiv es negativ is
LC0(y0+.2,n,jel=24) # az atlag pozitiv (felnyil jel)
LC0(y0-.1,n,jel=25) # az atlag negativ (lenyil jel)

# kozel 0 átlag
LC0(y0+.15,n,jel=247) # pozitiv atlag (szazalek jel)
LC0(y0-.082,n,jel=247) # pozitiv atlag (szazalek jel)

segments(0,0,1,1,col="gray",lwd=1.5)

legend(0,2,c("mind pozitív","mind negatív",
             "az átlag pozitív","az átlag negatív",
             "közel 0 átlag"),
             pch=c(3,175,24,25,247),cex=2/3)
```



Az ábra alapján a Lorenz-görbére vonatkozóan az alábbi négy észrevételt tehetjük:

- Csak akkor (0,0)-(1,1) négyzetbeli a görbe, ha minden vagyon pozitív, vagy ha minden vagyon negatív. Pozitív vagyonok esetén a görbe monoton növekvő, konvex, a négyzet átlója alatti. Negatív vagyonok esetén (azaz ha csak adósók vannak, akkor) a görbe monoton növekvő, de konkáv, a négyzet átlója feletti.

- Ha vannak pozitív és negatív vagyonok, de az összvagyon (azaz az átlag vagyon) pozitív, akkor marad a konvexitás, de a monotonitás sérül, lefele kilép a görbe a (0,0)-(1,1) négyzetből.

- Ha vannak pozitív és negatív vagyonok, de az összvagyon (azaz az átlag vagyon) negatív, akkor a görbe konkáv és a monotonitás is sérül, a görbe felfele kilép a (0,0)-(1,1) négyzetből.

- Ha az összvagyon közel 0, akkor a görbe továbbra is a (0,0) pontból az (1,1) pontba fut, de a végtelenbe ugrik. Ha az összvagyon tényleges előjele pozitív, akkor a mínusz végtelenbe, ha pedig negatív, akkor a plusz végtelenbe.

3.3. Lorenz-mértékek - a Lorenz-görbe karakterizációja

3.3.1. Definíció

Legyen F egy nemnegatív eloszlás függvény μ várhatóértékkel és legyen L az F eloszlás Lorenz-görbéje, tehát legyen

$$L(p) = \frac{1}{\mu}\int_0^p F^{-1}(t)\,dt \qquad p \in [0,1]$$

az F eloszlás k. Lorenz momentumának a [0,1]-beli

$$D_F(k) = \frac{k+1}{k}\int_0^1 p^k dL(p) - \frac{1}{k} \qquad k = 1,2,3...$$

konstansoknak nevezzük.

Ezek a konstansok, nyilván skála invariánsak ugyanúgy, mint maga a Lorenz-görbe. Tehát ezek a momentumok akár egyenlőtlenségi mértéknek is nevezhetők. Szokásos nevük még: Lorenz-mérték.

3.3.2. Tétel

A $D_F(k)$, $k = 1,2,3,...$-ra a $D_F(k)$ Lorenz mértékek léteznek.

3.3.3. Tétel

Az $L_F(u)$ Lorenz-görbét a $D_F(k)$, $k = 1,2,3,....$ Lorenz mértékek meghatározzák.

3.3.4. Tétel

A $D_F(k)$, $k = 1,2,3,...$ Lorenz mértékek, a μ várhatóértékkel együtt meghatározzák az F eloszlásfüggvényt.

A Lorenz-mértékek

$$D_F(k) = (k+1)\int_0^1 p^{k-1}(p+L(p))\,dp$$

alternatív kifejezése mutatja, hogy $D_1(F)$ Lorenz-mérték a Gini-féle index-szel azonos. Tehát a Lorenz-mértékek a Gini-féle index egyfajta kiterjesztései.

A Lorenz-mértékek további tanulságos alternatív kifejezését kapjuk, a k-ad rendű jövedelem várható jövedelem differencia, a $g_k(x)$ szerint. Ami tehát

$$g_k(x) = x - \mathbb{E}\left(\max_{i\in\{1,\dots,k+1\}} X_i | X_i \leq x, \qquad i \in \{1,\dots,k+1\}\right)$$

vagyis a $g_k(x)$ jövedelem többlet annyi, mint amennyi egy tetszőleges x jövedelem esetén az x jövedelem, és az x-nél kisebb jövedelmek maximális várhatóértéke k jövedelemszerzőt figyelembe véve. E várható jövedelem többlet alapján a k. Lorenz-mérték:

$$D_k(F) = \frac{\mathbb{E}(g_k(X))}{\mu}$$

vagyis a k. rendű jövedelem többlet várható értéke.

4. Az R0 reprodukciós index értelmezése és becslése

Egy járvány intenzitásának alakulását a napi eset számokból (az ismertté vált friss megbetegedések számából vagy az adott vírusban elhalálozottak számából) álló nem-negatív egész számokat tartalmazó számsorral tudjuk leírni. A járvány egyik fontos jellemzője, hogy egy beteg várhatóan hány további személyt fertőz meg. Ezt az értéket szoktuk R0 értéknek nevezni. Ez az R0 érték nyilvánvalóan nem állandó a járvány ideje alatt, mivel változás áll be annak függvényében, hogy a hatóságok például időközben milyen korlátozó intézkedéseket vezetnek be az országban.

Ebből az R0 értékből meg tudjuk mondani, hogy a járvány várhatóan még terjedő vagy már lecsengő szakaszban van-e.

4.0.1. Definíció

Az R0 reprodukciós ráta, ha

$$R0 = \begin{cases} R0 > 1, & \text{a járvány várhatóan még terjedő szakaszban van} \\ R0 < 1, & \text{a járvány lecsengő szakaszban van} \end{cases}$$

Sokféleképpen lehet modellezni magát a járványt és a járványt jellemző R0 értéket is. Természetesen egy-egy konkrét járvány nem feltétlenül felel meg egy adott modell osztálynak. Ebben a részben nem egy konkrét járvány valódi modellezésére törekszem, hanem arra, hogy bemutassak egy konkrét járvány modellt és azt, hogy annak alapján hogyan lehet az R0 értéket becsülni.

A módszer alkalmazható a jelenleg folyó járvány, a korona vírus paraméterének megbecslésére is. De erre ennek a dolgozatnak a keretében aligha kaphatnánk ténylegesen érvényes eredményeket, hiszen terjedelmi okokból is számos olyan szempontot kéne elhanyagolnunk, ami az eredmény szempontjából nem mellékes. Például az általunk bemutatott modell zárt társadalmat feltételez. Vagyis olyat, ahol nincs időközbeni be- illetve kivándorlás. De ennek a modellnek az is fontos paramétere, hogy hogyan alakul a betegségük lefolyása alatt egyes fertőzöttek fertőző képessége. Valamint milyen eloszlással modellezhető az egy beteg által megfertőzöttek száma. (Az általunk keresett R0 ennek az eloszlásnak csak a várhatóértéke.) Jelenleg nem állnak rendelkezésre nyilvános adatok az utóbbi két eloszlás tényleges megbecslésére.

4.1. A járvány modellezése

A járvány lefolyását két típusú véletlen mennyiség segítségével írjuk le:

1. egy fertőzött a fertőződésétől számított hányadik napon fertőző
2. egy fertőzött ténylegesen hány további személyt fertőz meg

Tulajdonság

- az egy-egy megfertőzöttet leíró változók függetlenek
- a megfertőződések függetlenek egymástól
- diszjunkt egymástól az a kör, akit egy-egy fertőzött megfertőz

Azaz minden megfertőzötthöz tartozik egy másik, korábban megfertőződött személy, az, akitől az illető a vírust megkapta.

Feltételezzük, hogy egy fertőzött a megfertőződése után egyetlen napon fertőző, és feltételezzük, hogy ez a nap a megfertőződés napját egy gamma-eloszlás szerinti késedelemmel követi. Az egy fertőzött által megfertőzöttek számát Poisson eloszlásúnak tekintjük, valamilyen ismeretlen értékű, R0 paraméterrel.

4.2. Egy járványgörbe szimulációja

Ha olyan járványt szeretnénk szimulálni, ahol a betegség inkubációs ideje gamma eloszlású, akkor első lépésként diszkretizáljuk a megfelelő gamma eloszlást, mert ugyan az az időtartam, ami egy beteg megfertőződésétől a fertőzővé válásáig eltelik, az folytonos, azonban az, hogy ezt úgymond mikor, melyik nap vesszük észre, az már diszkrét.

Egy n hosszú járványgörbe szimulálásához felveszünk egy n elemű vektort. Ennek a vektornak a cellái eleinte azt fogják jelenteni, hogy az adott napon hány fertőzötté vált személy van. Később a vektor első cellájától az utolsóig haladva az egyes cellákat átírjuk, és végül a cellák tartalma azt fogják jelenteni, hogy az egyes napokon hány új fertőzött adódott.

Kezdőértékként a vektor első elemébe 1-et írunk, a többibe 0-át. Ez így tehát azt jelenti, hogy az első napon 1 fertőző beteg jelent meg, a többi napon egyelőre nem tudunk fertőzötté vált betegről.

Sorsoljunk ki egy $m = m_1$ egész számot az $R0$ paraméterű Poisson eloszlás szerint. Ez az m jelenti azt, hogy az első napi 1 fertőző forrás hány személyt fertőzött meg. Sorsoljunk ki továbbá a diszkretizált gamma eloszlás szerint m darab $s_1, ..., s_m$ nemnegatív egész számot. Ezek fogják azt jelenteni, hogy az első napon megfertőződöttek hány nap múlva válnak fertőzővé.

Írjuk tehát, a járványvektor első cellájába az 1 helyett az m-et, és az $1+s_1$. ... $1+s_m$. cellák értékét növeljük meg eggyel-eggyel.

Így tehát azt látjuk, hogy az első napon m új fertőzött van és már látjuk azt is, hogy bizonyos további napokon is meg fog jelenni, egy-egy napon esetleg több fertőzőforrás.

Lépjünk tovább a következő olyan cellához – naphoz – amikor az eddig generált járványvektor szerint van legalább egy fertőzőforrás. Ha a generátum azt mutatja, hogy e napon k fertőzőforrás van, akkor generáljunk egy $k \cdot R0$ paraméterű Poisson eloszlás szerinti $m = m_k$ számot, és a diszkretizált gamma eloszlás szerint m darab $s_1, ..., s_m$ nemnegatív egész számot.

Az m jelenti azt, hogy az adott napon m új fertőzött adódott. Írjuk tehát, a járvány-vektor aktuális cellájába a k helyett az m-et, és növeljük az aktuális cállától számított s_1. ... s_m. cellák értékét eggyel-eggyel. Hiszen akkor a generátum szerint eggyel-eggyel több fertőzőforrás fog megjelenni.

Lépjünk a következő nem-nulla cellához. És így tovább.

Ezzel leírtuk azt, hogy hogyan generálunk egy járványgörbét és egyben azt is, hogy mit feltételezünk az elemezni kivánt járványgörbéről.

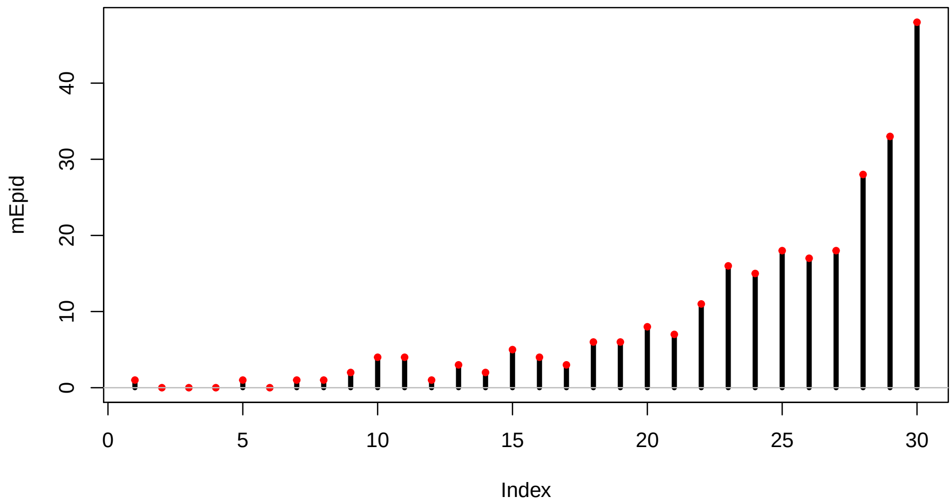
4.3. A módszer működése egy szimulált járványgörbe esetén

Feltételezzük, hogy az ágens $(3, 1.5)$ paraméterű gamma eloszlású idő multán $(5/3)$ paraméterű eloszlás szerinti számosságú újabb személyt fertőz meg. Ennek megfelelően generálunk egy 30 megfigyelés hosszú adatsort.

```
library(R0)
mGT <- generation.time("gamma", c(3,1.5))
str(mGT) # ez egy diszkretizalt eloszlás
sum(mGT$GT) # 1 $time=lehetséges értékek, $GT=valószínűség
sum(mGT$time*mGT$GT)-mGT$mean # $mean=várhatóérték
sum((mGT$time-mGT$mean)^2*mGT$GT)-mGT$sd^2 # $sd=szórás
set.seed(39) # a járványgörbe generálása
mEpid <- sim.epid(epid.nb=1, GT=mGT, epid.length=30,
               family="poisson", R0=1.67, peak.value=500)
mEpid <- mEpid[,1] # a generált esetszám
```

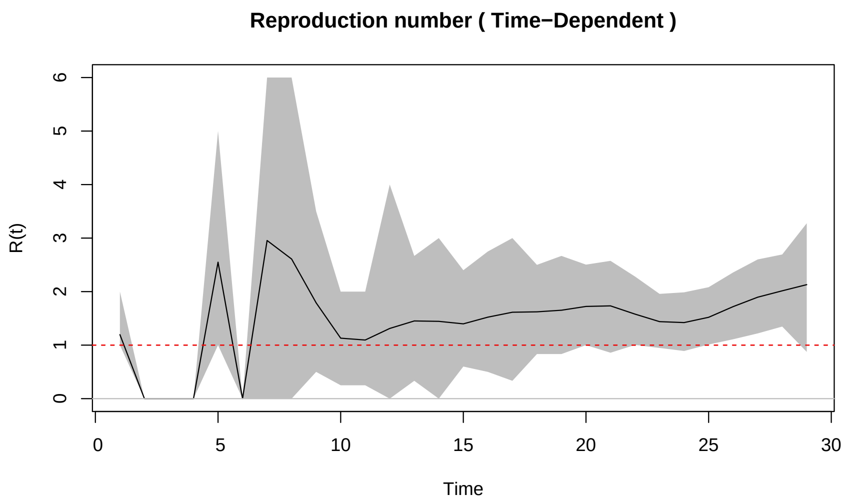
A generált járvány adatsor rajza

```
plot(mEpid,t="h",lwd=4)
points(mEpid,pch=20,col="red")
abline(h=0,col="gray")
```



Az időfüggő R0 becslése és kirajzolása.

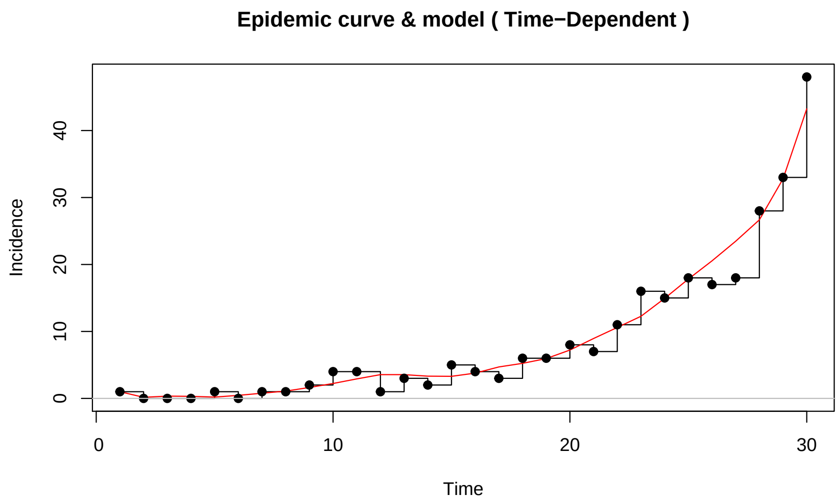
```
M <- estimate.R(epid=mEpid, GT=mGT, methods="TD")
plot(M)
abline(h=0,col="gray")
abline(h=1,lty=2,col="red")
```



Az $R0$ érték megbecslésére az $R0$ csomag `estimate.R()` eljárásának `TD` időfüggő verzióját használtuk fel. Az eredmény rajzán jól látható, hogy az adott járványgenerátum esetén az $R0$ érték folyamatosan az 1 határ fölött van. Ez a határ mutatja, hogy a járvány lecsengő vagy felfutó szakaszában van. Ugyanakkor az is érvényes, hogy a görbe értéke, a szürkén jelölt konfidencia tartománnyal együtt nem mond ellent annak, miszerint az adatokat egy $1+2/3$ $R0$ értéknek megfelelő módszerrel generáltuk.

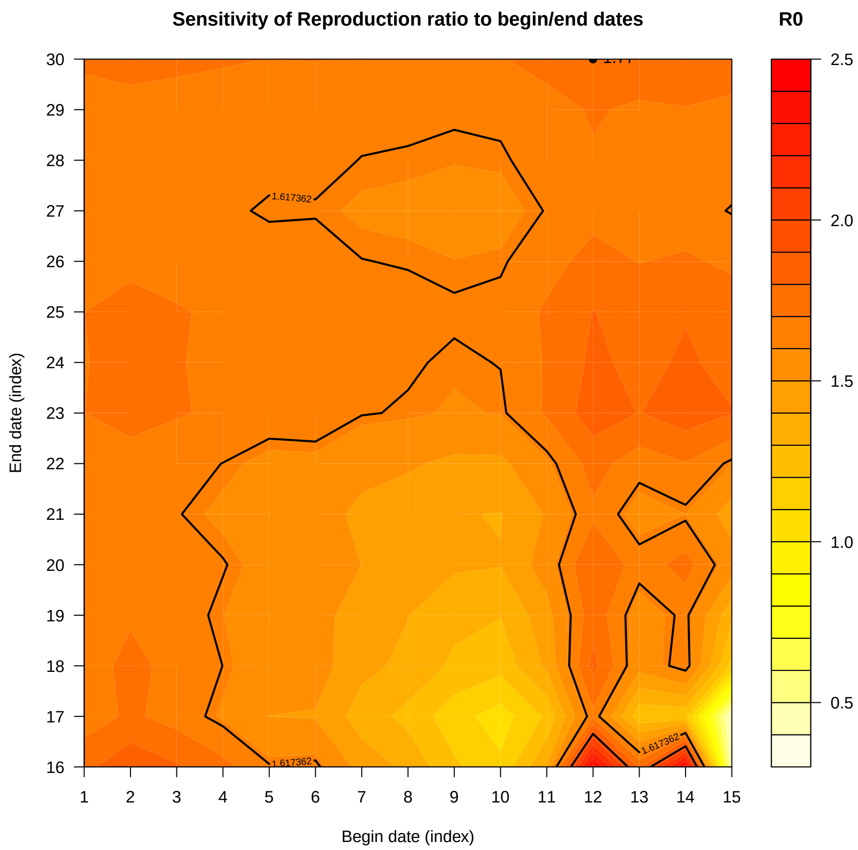
A járvány lefolyása a becsült R0 alapján.

```
plotfit(M)
abline(h=0,col="gray")
```



A fenti görbén az látható, hogy az a járványgörbe, amit az előzőleg becsült $R0$ értékeket felhasználva nyerhetünk és az ábrán piros színnel láthatunk, elég pontosan közelíti a járvány lefolyás valós értékeit.

```
M <- estimate.R(epid=mEpid, GT=mGT, methods="EG")
S <- sensitivity.analysis(res=M$estimates$EG,
                begin=1:15, end=16:30, sa.type="time")
pdf("abra3.pdf",width=5,height=5)
 plot(S)
dev.off()
```



Az ábráról az olvasható le, hogy a konstansként becsült $R0$ érték hogyan függ attól, hogy a járványidőszak melyik időablakát vesszük figyelembe. A kapott ábrán látható világosabb folt azt mutatja, hogy a járványidőszak középső harmadában az R0 érték viszonylag alacsony volt. A sötétebb foltok pedig azt, hogy a járvány második periódusában a reprodukciós ráta jelentősen romlott.

5. Összegzés

A járvány jellemzésének 3 eszközét vizsgáltam:

- a korosztály szerinti eloszlás szóródását Gini-féle együttható alapján
- a korosztály szerinti eloszlást az eloszlás Lorenz-görbéje alapján
- a reprodukciós ráta alakulását az idő függvényében

A dolgozatban az általam elvégzett vizsgálatoknak csak egy részét dokumentáltam. Ugyanakkor a tapasztalati adatok modellezéséhez az általam ténylegesen elemzett modelleknél sokkal részletesebbekre van szükség.

A jelen dolgozatban dokumentáltak a tapasztalati adatok vonatkozásában csak a modellek alkalmazhatóságának a vizsgálatára szorítkoznak. Ugyanakkor számos érdekes megfigyelést tartalmaznak, úgy az alkalmazott eszközök mint a tényleges modellezési eredmények tekintetében.

Hivatkozások

[1] J. Sanchez-Perez, L. Plata-Perez and F. Sanchez-Sanchez: An elementary characterization of the Gini index MPRA Paper No. 36328, posted 9. February 2012.

[2] R. Aaberge: Characterizations of Lorenz curves and income distributions, Soc. Choice Welfare Vol 17, pp. 639-653, 2000.

[3] Jacco Wallinga & Peter Teunis: „Differend Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures" American Journal of Epidemiology, Vol. 160, Num. 6, Sept. 15, pp. 509-516 2004.

[4] Achim Zeileis: ineq: Measuring Inequality, Concentration, and Poverty; R package version 0.2-13, url = https://CRAN.R-project.org/package=ineq, 2014.

[5] MA. Sanchez & SM. Blower: Uncertainty and Sensitivity Analysis of the Basic Reproductive Rate Am. J. Epidemiology, Vol. 145, No. 12, pp. 1127-1137, 1997.

[6] `who.int/emergencies/diseases/novel-coronavirus-2019`
(Legutóbbi elérés időpontja: 2020. december 15.)

[7] `hu.wikipedia.org/wiki/2020-as_COVID%E2%80%9319`
(Legutóbbi elérés időpontja: 2020. december 15.)

[8] `economicshelp.org/blog/glossary/lorenz-curve/`
(Legutóbbi elérés időpontja: 2020. december 15.)

[9] `investopedia.com/terms/g/gini-index.asp`
(Legutóbbi elérés időpontja: 2020. december 15.)

[10] `hu.wikipedia.org/wiki/2020-as_COVID-19-koronavírus-járvány_Magyarországon`
(Legutóbbi elérés időpontja: 2020. december 15.)