

Szakdolgozat

Kofrán Dániel

2020

Budapesti Corvinus Egyetem
Közgazdaságtudományi Kar

A COVID-19 által okozott halálozások vizsgálata Olaszországban

Egy ökonometriai megközelítés

Készítette: Kofrán Dániel
Alkalmazott Közgazdaságtan
2020

Szakszemináriumvezető: Dr. Chikán Attila

Tartalomjegyzék

1.	Bevezetés.....	1
2.	A SARS-CoV-2 és a COVID-19 bemutatása.....	2
2.1.	A vírusról általánosan.....	2
2.2.	A vírus eredete.....	4
2.3.	A világjárvány halálozási mutatója	4
3.	A COVID-19 terjedése Olaszországban.....	7
3.1.	A járványhelyzet kialakulása.....	7
3.2.	A járvány terjedését könnyítő politikai tényezők	7
3.3.	A járvány okozta halálozások mögötti általános tényezők.....	8
3.4.	A járvány okozta halálozások mögötti demográfiai tényezők.....	9
4.	Adatbázis és felhasznált módszertan	12
4.1.	Adatbázisépítés.....	12
4.2.	Módszertan	14
4.3.	Korlátok.....	15
5.	Elemzés	16
5.1.	Az eredményváltozó leíró statisztikai elemzése	16
5.2.	A magyarázó változók jellemzése	18
5.3.	Az alapmodell kiválasztása	20
5.4.	Modellszűkítés.....	22
5.5.	A végleges lineáris modellek.....	24
5.6.	Térbeli autokorreláció	26
5.7.	Végleges lineáris modellek és a Térbeli autoregresszív modellek	29
5.8.	A végleges modell értelmezése	31
6.	Konklúzió	34

Ábrajegyzék

1. ábra: A COVID-19 halálozási rátája 2020. februárja és májusa között	5
2. ábra: Az olasz népesség becsült koreloszlása korcsoport és nem szerint, 2020-ban.....	10
3. ábra: A megfigyelt halálozási ráta az olasz koronavírusos betegek körében, korcsoport és nem szerint	11
4. ábra: Olaszország regionális és tartományi felosztása	Hiba! A könyvjelző nem létezik.
5. ábra: A COVID-19 által okozott halálozások hisztogramja Olaszországban, regionális szinten.....	17
6. ábra: A COVID-19 által okozott halálozások logaritmusának hisztogramja Olaszországban, regionális szinten.....	18
7. ábra: A leggyakoribb térbeli súlyok applikálása Olaszországra.....	27
8. ábra: A leggyakoribb térbeli súlyok melletti Moran-féle I-indexek, a halálozások számának logaritmusával számolva	28

1. Bevezetés

Jelen dolgozat a 2019 decemberében megjelenő COVID-19 által okozott halálozásokat vizsgálja Olaszországon belül. A COVID-19 2020 elején érkezett meg Olaszországba, ami hamar a megbetegedések európai epicentruma lett. Az országon belül azonban nagyon éles eltérések vannak abban, hogy a vírus mely országrészen mekkora halálozási számmal rendelkezik. Jelen dolgozat arra a kutatási kérdésre keresi a választ, hogy Olaszországon belül milyen tényezők befolyásolták döntően a COVID-19 által okozott halálozások rendkívül egyenetlen eloszlását, 2020 áprilisának végéig.

A témaválasztást a téma aktualitása indokolta; az egyetemi Statisztika és Ökonometria tárgyak során elsajátított ismeretek alapján szeretném elemezni a kialakult viszonyokat. Az elemzés a kutatási kérdésre a felhasznált adatbázisok típusából adódóan egy keresztmetszeti ökonometriai elemzésén keresztül keresi a választ. Az elemzés kérdésfelvetéséből adódóan a hagyományos keresztmetszeti elemzési módszerek alkalmazása mellett térbeli autoregresszióval és térbeli autokorrelációs modellekkel is kiegészül a módszertan.

A dolgozat további 5 fő részre tagolható. Az ezt követő fejezetben a jelenlegi világjárványt okozó vírust és magát a járványt mutatja be részletesebben a dolgozat. Kitér a vírus eredetére, illetve a halálozási rátája mögött meghúzódó tulajdonságaira, majd bemutatja a halálozási ráta növekedésén keresztül a vírus egyéb karakterisztikáit.

A második fejezet már kizárólag a COVID-19 és Olaszország kapcsolatára fókuszál. A vírus megjelenése és a fokozatosan szigorodó kormányzati rendeletek idővonalának bemutatása után a fejezet az Olaszországi halálozások potenciális okait keresi. Ezt először az olasz kormány korai válságkezelésén és demokratikus berendezkedésén keresztül, majd eggyel általánosabb szinten, végül pedig az Olaszország-specifikus demográfiai mutatók tükrében vizsgálja.

Az ezt követő fejezet az adatbázisépítés folyamatát, a módszertant és az elemzés korlátait ismerteti. Az ötödik fejezet a keresztmetszeti ökonometriai elemzés részein megy végig. Az elemzés először a felhasznált szakirodalom alapján összerakott alapmodellt jellemzi, majd kétféle modellszelekciós eljárás segítségével szűkíti azt. Ezután az itt kapott lineáris modellek mellett térbeli autoregresszív modellek relevanciáját is teszteli, végül a kiválasztott legjobb végleges modellt jellemzi. Az utolsó fejezet összegzi az elemzés eredményeit.

2. A SARS-CoV-2 és a COVID-19 bemutatása

2.1. A víusról általánosan

A jelenlegi vilájárványt okozó vírus hivatalos angol neve Severe Acute Respiratory Syndrome Coronavirus 2 [magyarul Súlyos Akut Légzőszervi Szindróma Koronavírus 2 (Magyarország Kormánya, 2020a)], azaz rövidítve, SARS-CoV-2. A COVID-19 a Coronavirus Disease 2019 rövidítése, ami a 2019-ben észlelt koronavírusos megbetegedést jelenti. Ezt a vírust a köznyelv csak koronavírusként emlegeti, amely a nevének csak egy részben indokolt rövidítése. Amíg a SARS-CoV-2 valóban a koronavírusok családjába tartozik, csupán egy a hét, emberre is áttérjedt koronavírus közül (Andersen et al. 2020).

Ez a hét vírus az angol nyelvű rövidítéseik szerint: az HCoV-HKU1 (HCov = Human Coronavirus, magyarul emberi koronavírus), az HCoV-NL63, az HCoV-OC43, az HCoV-229E, a SARS-CoV, a MERS-CoV, illetve mostmár, a SARS-CoV-2 is. Ezek között találhatók pusztán enyhe tüneteket produkáló vírusok (a felsoroltak közül az első 4), de akadnak közöttük olyanok is, amelyek képesek súlyos járványokat vagy akár vilájárványokat is okozni [a felsoroltak közül az utóbbi 3 (Andersen et al. 2020)].

A jelenlegi koronavírus a korábban emberekben is diagnosztizált koronavírusok közül a fent említett két csoportból a SARS-CoV-ra és a MERS-CoV-ra hasonlít tulajdonságaiban a legjobban. A SARS-CoV okozta a SARS-járvány kitörését, amely az évszázad első nagyobb volumenű járványügyi helyzetét teremtette meg (Coronavirus Today, 2020b). A járványt 5 kontinens 24 országában azonosították, 8096 esetből 774 halottal, azaz egy közepesen magas, 9,6 százalékos halálozási rátával (World Health Organization, 2003). Mivel a SARS a megfertőzött betegek közel egytizedével végzett, és mivel rövid volt a tünetmentes lappangási periódus, az egészségügyi hatóságok gyorsan reagálva, hamar meg tudták fékezni a terjedését, így vilájárványnak sosem nyilvánították ki.

A MERS-CoV által okozott MERS [Middle East Respiratory Syndrome, magyarul Közel-Keleti Légzőszervi Szindróma (Magyarország Kormánya, 2020a)] pedig egy, a COVID-19-nél és a SARS-nál is szignifikánsan magasabb, több, mint 30 százalékos halálozási aránnyal rendelkezik. Vilájárványi státuszba sosem került, amelyben nagy szerepet játszott, hogy a betegséget többnyire csak a közel-keleti régióban észlelték, ahol a fertőzöttek száma a 3000-et sem lépte át (Coronavirus Today, 2020a).

Ez a két betegség több aspektusában hasonlít a ma tomboló világjárványra. Párat a legfontosabbak közül kiemelve: eredetében, halálozási arányaiban és az elhalálozásra legesélyesebb korcsoportjaiban. Mindhárom megbetegedést okozó vírusnál zoonózisról beszélhetünk, azaz állatról emberre terjedt vírusokról. A MERS-CoV a dromedárokon keresztül találta meg az első emberi áldozatokat (Nowotny és Kolodziejek, 2014), míg a SARS-CoV, és feltételezhetően a SARS-CoV-2 is denevérekben volt először felfedezhető (Andersen et al. 2020, Zhou et al, 2020). Emellett, a 3 betegség mind jelentősen magasabb halálozási aránnyal rendelkezik, mint a hagyományos járványok (például az influenza), és mindhárom megbetegedés elsősorban az idősebb generációkra életveszélyes (Andersen et al. 2020, Coronavirus Today, 2020a, Coronavirus Today, 2020b).

Ezek a fajta koronavírusok sokkal kevésbé tudtak elterjedni a másik négyhez képest, éppen azon tulajdonságok mentén, amelyek a legjobban elkülönítik a két csoportot. Mivel a SARS-CoV-2 típusú koronavírusok sokkal veszélyesebbek az emberre, a HCoV-229E típusú koronavírusok sokkal elterjedtebbé tudtak válni [a közönséges megfázások 15-30 százalékáért például a HCoV-229E típusú koronavírusok a felelősek (Mesel-Lemoine et al, 2012)].

A SARS-CoV-2 által okozott megbetegedés, a COVID-19 többek között azért tudott mégis a SARS-szal és a MERS-szel ellentétben helyi járvány helyett világjárvánnyá válni, mert a leggyakoribb tünetei nagyban átfedtek a közönséges megfázás, vagy a hagyományos influenza tüneteivel, mint például a láz, a fáradtság és a száraz köhögés (Magyarország Kormánya, 2020b), és az enyhe tünetek gyakran sokkal kisebb feltűnést is keltenek. Továbbá fontos tulajdonsága a jelenlegi koronavírusnak, hogy a tünetmentes fertőzöttek is tovább tudnak fertőzni, amely a másik két vírusra egyáltalán nem volt jellemző.

Ezekben a tulajdonságaiban sokkal inkább a HCoV-229E típusú koronavírusokra hasonlít, azonban a halálozási rátája kiemeli ezek közül. A COVID-19 2020. május 1-re 3.175.207 hivatalos esetből 224.172 halottal, azaz 7,06 százalékos halálozási aránnyal rendelkezett (World Health Organization, 2020c). Összegezve, a COVID-19 egyszerre rendelkezik a HCoV-229E típusú vírusos megbetegedésekre hajazó terjedési sebességgel, és a SARS-ra vagy MERS-re hajazó halálozási aránnyal. Ezek a tulajdonságok tudták a jelenlegi járványt világjárvánnyá tenni.

2.2. A vírus eredete

Mielőtt az egész világra elterjedt, emberben először 2019 decemberében diagnosztizálták a COVID-19-et, a kínai Hubei tartomány Vuhan nevű városában (Lin et al, 2020). Ekkor még a kínai hatóságok nem tulajdonítottak akkora jelentőséget a megbetegedéseknek, így nem is hozták meg a megfelelő súlyú intézkedéseket. Mikor mind a kínai hatóságok, mind az Egészségügyi Világszervezet (World Health Organization, röviden WHO) számára egyértelművé vált a vírus fajsúlya, már nem lehetett Hubei tartományon belül rekeszteni.

A Kínai kormány ezután határozott és kiemelkedően szigorú intézkedések sorát vezette be (mint például a tömegközlekedés teljes felszámolása Vuhanban vagy részleges felszámolása Hubei tartomány-szerte), mellyel a terjedést kívánta megakadályozni (Lin et al, 2020). Az elsődleges céljuk a népesség olyan magas fokú izolálása volt, hogy a vírus minél kevesebb új gazdatestre találhasson rá, és idővel eltűnjön az érintett régiókból.

A kínai intézkedések hatékonysága megkérdőjelezhetetlen lett, azonban az időzítésével elkéstek. „Az Egészségügyi Világszervezet (WHO) január 30-án kihirdette az nemzetközi jelentőségű közegészségügyi vészhelyzetet (Public Health Emergency of International Concern)” (Jakab és Kemenesi, 2020). Ez még nem határozott és azonnali hatályú intézkedéseket jelentett, mint a határok teljes, vagy részleges lezárása, csupán felruházta a WHO-t egy megnövekedett, rendkívüli állapotra szabott mozgástérrel (Jakab és Kemenesi, 2020).

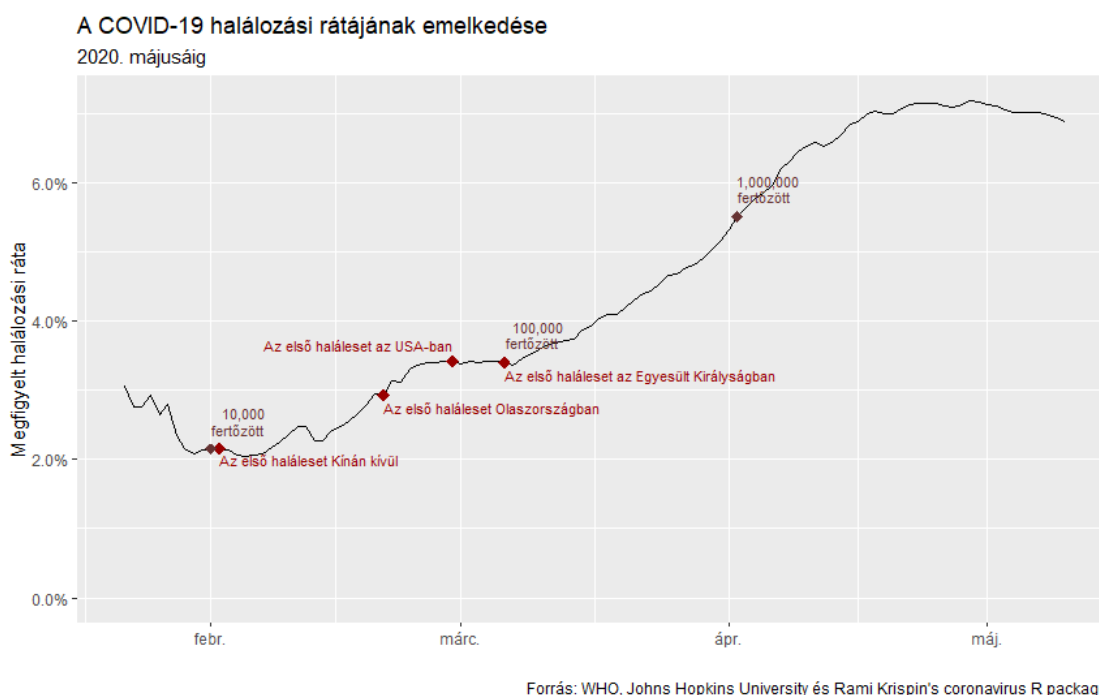
Ezután, mikor már több mint 100.000 bejelentett fertőzöttről és megközelítőleg 5.000 halottról tudtak összesen 114 országban, majdnem másfél hónapra rá, március 11-én jelentette be a WHO főigazgatója, Tedros Adhanom, hogy innentől világjárványról beszélünk (World Health Organization, 2020d). Ekkor azonban már Kínán kívül több új epicentruma is keletkezett a vírusnak, mint például Dél-Korea (7.755 eset, 60 halott), Irán (8.042 eset, 291 halott), vagy éppen Olaszország [10.147 eset, 631 halott (World Health Organization, 2020b)].

2.3. A világjárvány halálozási mutatója

Habár a SARS-CoV-2 2020 januárjában kezdett el Kínán kívül elterjedni, csupán 2020 februárjának vége felé, tavaszának elején kezdett el a vírus az európai és észak-amerikai országokban is halálos áldozatokat szedni. Az Amerikai Egyesült Államok és az Egyesült Királyság, ebben a sorrendben rendelkezik jelenleg (2020 május közepén) a legtöbb

COVID-19-ben elhunytal, többek között Olaszországot is megelőzve (Olaszország az Egyesült Királyság mögött szorosan a harmadik ezen a listán).

Az 1. ábra azt mutatja meg, hogy január második fele óta, hogyan alakult a COVID-19 halálozási rátája. Ez a ráta az összes ország összes korcsoportját tartalmazza. Az ábráról jól kivehető, hogy a 7 százalékos halálozási arányt már április második felében átlépte a járvány, majd május elején újra ez alá csökkent. A betegség megfigyelt halálozási aránya az idősebb generációk körében ennél a szintnél jóval magasabb, miközben a fiatalabbak számára többnyire veszélytelen.



1. ábra, A COVID-19 halálozási rátája 2020. februárja és májusa között. Saját ábra, források az ábrán feltüntetve (2020).

Az 1. ábra továbbá négy meghatározó eseményt és három fontos mérföldkövet jelöl, a vírus általános terjedése, és a COVID-19 halálozási arányával kapcsolatban. Az ábra tartalmazza az első Kínán kívüli halálestet, illetve az első halálestet a jelenleg legtöbb halottat számláló országokban. Érdekes megfigyelni, hogy amíg halálesteket csak Kínán belül jegyezték fel - feltételezhetően a már említett hirtelen és határozott intézkedéseknek köszönhetően – a halálozási ráta fokozatosan csökkenni kezdett. Azonban, amikor a vírus már elkezdett a kínai területeken kívül felbukkanni, és nem sokra rá megjelentek az első COVID-19-hez köthető halálozások Kínán kívül is, a halálozási ráta rövid átfutási idővel ismét elkezdett növekedni.

A vírus a megjelenése után eleinte lassan, majd fokozatosan gyorsuló tempóban kezd el embereket fertőzni, és elhalálozásokat okozni. Ebből adódóan az első megbetegedések és az első halálozások után el kell teljen egy bizonyos idő a halálozási ráta berobbanásához. Emellett, eleinte az elhunytaknál átlagosan 17,8 nap telt el aközött, hogy valakin megjelentek a vírus tünetei és aközött, hogy belehalt (Verity et al, 2020), tovább bővítve a két időpont között eltelt időt. Ez a vírus előrehaladtával egyre rövidült, májusra már csak átlagosan 11-15 nap kellett ehhez, attól rövidülve annak függvényében, hogy volt-e lehetőség intenzív osztályon kezelni az illetőt (Istituto Superiore di Sanità, 2020).

Az első kiemelt eseményt követően a halálozási ráta növekedésnek indult, azonban majdnem két hónapig a 4 százalékos határt sem lépte át a mutató. Az ábráról jól kivehető, hogy a fent említett időbeli eltolódásokat beleszámolva jelentős mértékben azután kezdett el nőni a halálozási arány, miután Olaszországban, az Amerikai Egyesült Államokban és az Egyesült Királyságban is elkezdett növekedni a halottak száma. Az ábra továbbá szemléletesen az esetszám rohamos növekedési ütemét február, március és április alatt. Az esetszám alig több mint egy hónap alatt nőtt 10.000-ről 100.000 fölé, majd ezután kevesebb, mint egy hónap alatt 1.000.000 fölé.

Ez jól érzékelteti a COVID-19 2020 eleji terjedésének ütemét, mely a helyi egészségügyeket világszerte lebénította, tovább növelve a koronavírus áldozatainak számát. A fertőzöttek száma 1 hónappal az egymillió küszöb átlépése után már csak alig több, mint megháromszorozta önmagát, ami többek között a világ legtöbb országában bevezetett szigorú intézkedéseknek, és az egyre sikeresebb válságkezelésnek tudható be. Ez a jelenség Olaszországban is megfigyelhető.

3. A COVID-19 terjedése Olaszországban

3.1. A járványhelyzet kialakulása

2020 január 30-án diagnosztizálták COVID-19-cel az első betegeket Olaszországban, két kínai turistát (Rezza et al, 2020). A koronavírus beazonosítása után a helyi hatóságok felvették a kapcsolatot a többi régió vezetésével, és elkezdtek visszakövetni, hogy hol jártak előtte, de nem találtak velük érintkezett beteget (Rezza et al, 2020). A két turistát február 26-ára gyógyultnak nyilvánították. Az első koronavírus fertőzöttet, akinek nem volt korábban kontaktusa senkivel, aki az azt megelőző időszakban Kínában lett volna, az olasz Egészségügyi Minisztérium (Ministero della Salute) adatai szerint február 18-án azonosították, Lombardia régióban (Ministero della Salute, 2020). Tovább súlyosbítva a helyzetet, az illetőről kiderült, hogy rengeteg helyen járt az azt megelőző napokban, potenciálisan több száz embernek átadva a vírust.

Erre pár napra, február 23-án 11 várost teljesen lezártak a hatóságok Észak-Olaszországban, majd lemondtak főbb eseményekről, mint például a Velencei Karnevál utolsó napjait. Emellett ezeken a területeken már ekkor elkezdtek bezárni az iskolákat is (Horowitz et al, 2020). Március 8-án Olaszország miniszterelnöke, Giuseppe Conte bejelentette egész Észak-Olaszország lezárását, majd egy nappal később kiterjesztették ezt a rendeletet az egész Olaszországra, ezzel első európai államként karanténba helyezve egy egész országot (Horowitz et al, 2020). Ezután a következő meghatározó esemény a március 22-én hatályba lépő rendelet volt, ami kimondta, hogy tilos elhagyni azt a közigazdasági egységet (például város vagy község), ahol az egyén jelenleg tartózkodik, kivéve, ha az igazoltan munkaügyhöz köthető, abszolút vészhelyzet, vagy egészségügyi okokból teszi az illető (Rezza et al, 2020).

3.2. A járvány terjedését könnyítő politikai tényezők

Az olasz intézkedések ekkoriban Európa-szerte még messze a legdrasztikusabb intézkedések voltak, azonban a vírus terjedésének megfékezéséhez túl későn vezették be ezeket. Az olasz intézkedések folyamatosan le voltak maradva a vírus terjedésének és halálozásának tempójához képest (Horowitz et al, 2020). Olaszországot felkészületlenül érte a vírus éppen azért, mert az első európai országgént érkezett meg oda a vírus 2020 telén. A koronavírus terjedésének kritikus időszakában, mikor még könnyebben követhető a fertőzöttek diagnosztizálás előtti tevékenységei, és az exponenciális terjedési görbe elején helyezkedik el. Az olasz kormány- és államfők eleinte nem kezelték

megfelelő súllyal a vírust. Ebben az időszakban az olasz kormánykoalíció egyik pártjának vezetője egyenesen arra buzdította az olaszokat, hogy ne hagyjanak fel korábbi szokásaikkal (Horowitz et al, 2020).

Mikor felgyorsultak az események, és pár nappal később hivatalos források az olasz kormányfők álláspontjaik teljes ellentétét javasolták, sok olasz fejében zavarodottságot és hamis biztonságérzetet keltettek a vezetők nyilatkozatai, mely a vírus terjedésének megállítását tovább nehezítette. Emellett az olasz kormány a gazdasági érdekeit szem előtt tartva kezdetben vonakodva akarta bevezetni a szükséges intézkedéseket (Horowitz et al, 2020). Mikor az olasz kormány ráébredt a veszély valódi súlyára, vírus már Olaszország legsűrűbben lakott területeire is beférkőzött, az országon és Európában belül is epicentrummá változtatva Lombardia régiót, majd Észak-Olaszország jórészét.

Emellett, Olaszország politikai berendezkedése tovább nehezítette a járvány megállításának esélyeit. A nyugati demokráciáknak versenyhátrányuk volt három fontos tényezőben: az állam végrehajtó hatalmának korlátjaiban, a szólás- és a mozgásszabadságban. Az első a másik kettő velejárója, ezek pedig az alul- és félreinformált források szabad áramlását eredményezte, melyben nehezebb volt a vírus megfékezésének létszükségleti információk hatékony terjesztése, illetve az embereknek nehezebb folyamat volt megtiltani azt, hogy kimenjenek erős indok nélkül az utcára (Keefe, 2020). Az olasz vezetés vonakodása tehát a politikai berendezkedésből adódó hátrányokkal kiegészülve feltehetően hozzájárult a COVID-19 terjedésének és halálozási arányának volumenéhez. Ezeknek a kormányzati okoknak az elemzése keresztmetszeti adatként nem értelmezhető, így később az elemzésben sem szerepelnek.

3.3. A járvány okozta halálozások mögötti általános tényezők

A politikai reakcióidő és a szigorítások volumene mellett világszerte rengetegen foglalkoznak azzal, hogy kiderítsék, hogy pontosan milyen tényezők járnak közben közvetlenül a COVID-19 által okozott halálozásoknál. A Johns Hopkins University School of Medicine jelenleg jelentős összegeket fektet a járvány feltérképezésébe és megállításába, illetve híres az egyik legelterjedtebben használt COVID-19-es adatokból álló adatbázis fenntartásáról is (Johns Hopkins University, 2020). Az egyetem kutatói potenciális magyarázó erőkként a következőket tartják számon: a teszteléseket, a demográfiai jellemzőket, az egészségügyi rendszereket, illetve egyéb, jelenleg még ismeretlen faktorokat (Johns Hopkins University, 2020).

A tesztelések és a halálozási arány között negatív kapcsolatot feltételeznek: „több teszteléssel több beteget lehet enyhe tünetekkel azonosítani, ami csökkenti a halálozási rátát”. A demográfia alatt a halálozási ráta és az idősebb társadalmak közötti pozitív kapcsolatra gondolnak. Az egészségügyi rendszerrel pedig olyan összefüggésbe hozzák a halálozást, hogy „például ahogyan a kórházak túlterhelté válnak, és kevesebb erőforrásuk lesz, a halálozási arány nőhet” (Johns Hopkins University, 2020). A tesztelésekkel és az egészségügyi rendszerrel kapcsolatos adatokról az elemzésben esik még szó, a demográfiai háttérhez köthető okokat pedig a következő tartalmi alpont tárgyalja.

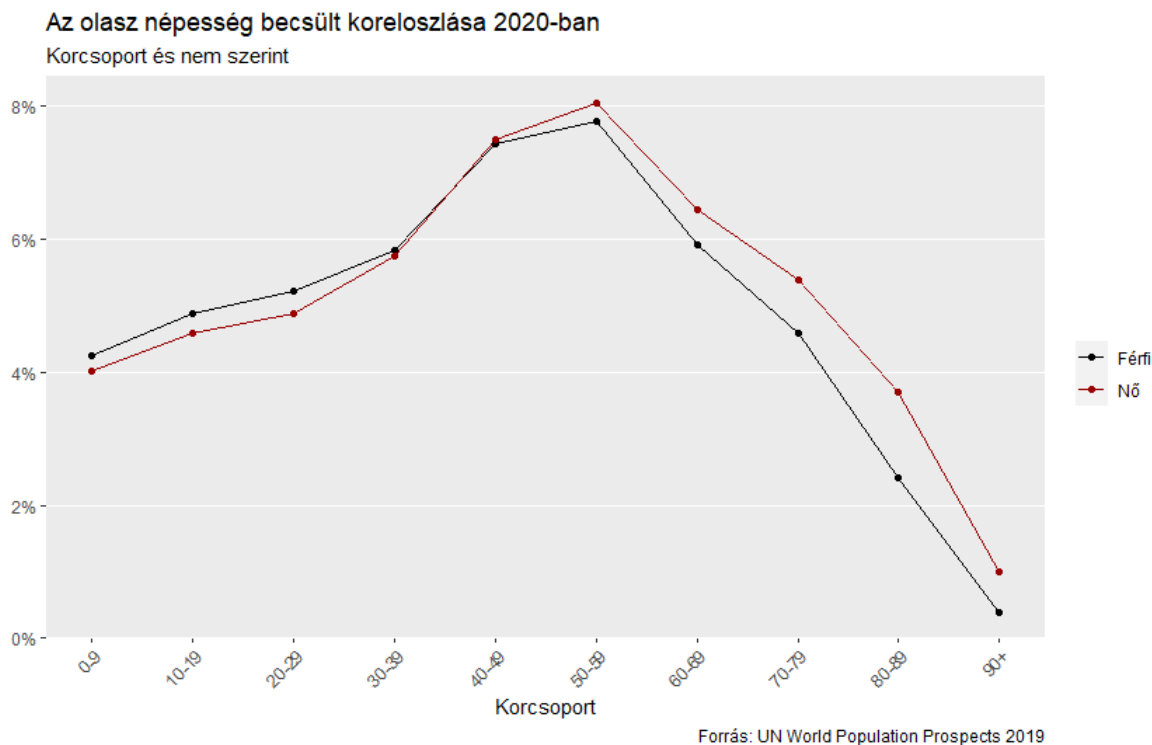
Továbbá egyre több egészségügyi szakterülettel rendelkező kutató és szervezet foglalkozik azzal, hogy milyen összefüggések vannak a korábban fennálló komoly egészségügyi problémák (Ministero della Salute, 2020), vagy éppen a dohányzás (World Health Organization, 2020e) vagy az alkoholfogyasztás (World Health Organization, 2020a) és a COVID-19 által okozott halálozások között. Az olasz Egészségügyi Minisztérium analizálta 2.351 halott korábbi egészségügyi adatait, melyből arra jutottak, hogy az elhunytak 60,3 százaléka rendelkezett 3 vagy több korábban fennálló krónikus betegséggel (Ministero della Salute, 2020). Emellett, mind a dohányzás, mind a komoly alkoholfogyasztás, mind a túlsúly segíthet a COVID-19 legsúlyosabb szövődményeinek kialakulásában, mint például az ARDS (acute respiratory distress syndrome, magyarul akut légzési distressz szindróma), mely gyakorta a halál közvetlen kiváltó oka (World Health Organization, 2020a, 2020e, Centers for Disease Control and Prevention, 2020).

3.4. A járvány okozta halálozások mögötti demográfiai tényezők

Az vírus okozta halálozások térnyerését Olaszországban fontos demográfiai tényezők is könnyítették. Az ország legfontosabb demográfiai jellemzői közé tartozik a társadalom koreloszlása. Az olasz társadalom jelenleg a legöregebb Európában, és Japán mögött a második legöregebb a világon (Boccia et al, 2020). Olaszország becsült koreloszlása korcsoport és nem szerint a 2. ábrán látható. Ez a tulajdonság, az észak-olaszországi magas népsűrűséggel, és a vírus terjedésének korai szakaszában elmulasztott határozott intézkedésekkel kombinálva fogékony táptalajává vált a COVID-19 térhódításának.

A koronavírus terjedése ezeken a területeken egy meghatározó intő jel lett a többi európai ország számára. Az egyik legnagyobb múltú és presztízsű amerikai egészségügyi folyóirat, a New England Journal of Medicine már márciusban úgy fogalmazott a SARS-

CoV-2-ről, hogy „minél medikalizáltabb és centralizáltabb a társadalom, annál szélesebb körben terjed a vírus. A gazdag Lombardiában kibontakozó katasztrófa bárhol megtörténhet” (Nacoti et al, 2020).

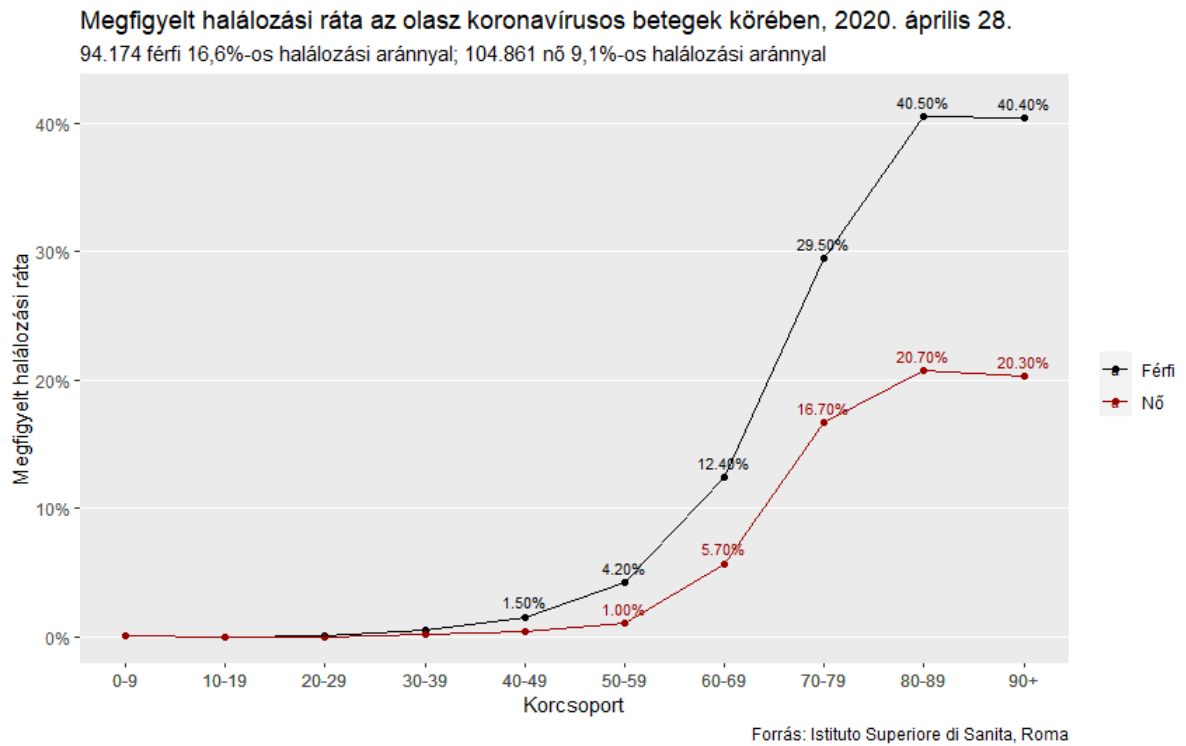


2. ábra, Az olasz népesség becsült koreloszlása korcsoport és nem szerint, 2020-ban. Saját ábra, az UN World Population Prospects 2019 adatai alapján.

Olaszországban április végig összesen körülbelül 205.000 fertőzöttet jegyeztek fel, melyből több mint 100.000 volt az aktív fertőzöttek száma, és majdnem 28.000 a halottak száma. A halálos áldozatok átlagos életkora 79 év volt, míg ugyanennek a mediánja 81 év volt (Ministero della Salute, 2020). Az átlag kisebb, mint a medián, ami arra enged következtetni, hogy korcsoportok szerint egy jobbra ferde eloszlásról beszélhetünk; ez meg is jelenik a 3. ábrán. Emellett, a 2 évnyi eltérés a két mutató között azt is jelentheti, hogy a fiatalabb korcsoportok közt találhatóak outlierek. Továbbá, az utóbbi mutató azt is magába foglalja, hogy a COVID-19-ben meghalt betegek több mint fele 80 éves vagy annál idősebb volt.

Ez, a korábban megállapított tulajdonsággal, miszerint az idősebb generációkra fokozottan veszélyesebb a vírus, illetve az előző bekezdésben tárgyalt idős olasz társadalmi szerkezettel szoros összhangban van. A 2. és a 3. ábra tökéletesen szemlélteti, hogy bár az olasz társadalom a világ legöregebbjei között van fenntartva, a 80 éves, vagy annál öregebbek a népesség 10 százalékát sem teszik ki, azonban a népességnek ezen

részen tapasztalhatóak a legnagyobb halálozási arányok. Azoknak a 80 éves, vagy annál idősebb nőknek, akik elkapják a COVID-19-et, kevesebb, mint 80 százaléka éli túl, a férfiaknak pedig kevesebb, mint 60 százaléka.



3. ábra, A megfigyelt halálozási ráta az olasz koronavírusos betegek körében, korcsoport és nem szerint. Saját ábra, az Institutio Superiore di Sanita, Roma adatai alapján (2020).

Emellett, a 2. ábra egy újabb fontos demográfiai jellemzőre is felhívja a figyelmet. A 2. ábráról jól kivehető, hogy a női népességnek nagyobb a részaránya az olasz társadalmon belül az összes 40 éves, vagy annál idősebb korosztályban, mint a férfi népességnek. A 3. ábra pedig többek között azt mutatja meg, hogy az összes olyan korosztályban, ahol az adott korosztályon belül a megfigyelt halálozási arány magasabb, mint 1 százalék, a férfiak halálozási rátái jelentősen magasabbak.

Az április 28-ai adatok alapján több mint 10.000 fővel, azaz közel 13,5 százalékkal több nő kapta el a fertőzést, mint férfi azonban az eltérő nemek szerinti halálozási arány miatt (16,6 százalék a férfiaknál, 9,1 százalék a nőknél) több, mint 5.000 fővel, azaz közel 64 százalékkal több férfi halt bele a járványba, mint nő. A tény, hogy mindez egy női többségű társadalomban történt meg felveti a kérdést, hogy milyen egyéb nemhez köthető tényezők állhatnak emögött. Jelen elemzés részben erre is keresi a választ.

4. Adatbázis és felhasznált módszertan

4.1. Adatbázisépítés

Az elemzésben használt adatbázis három fő forrásra hagyatkozott. Az első a Dipartimento della Protezione Civile (az olasz Polgári Védelmi Minisztérium) által közzétett adatbázis volt az olaszországi COVID-19 terjedésével kapcsolatban. A második az Istituto Nazionale di Statistica (az olasz Nemzeti Statisztikai Hivatal) különböző éves bontásban közzétett adatai voltak. A harmadik az Eurostat regionális adatbázisa volt, mely Olaszország térbeli adatait tartalmazza.

Az olasz Polgári Védelmi Minisztérium adatai 2020. február 24-től tartalmazzák az összes fontos COVID-19-cel kapcsolatos információt, mint például a megbetegedések száma, a halálozások száma, vagy a gyógyultak száma – összesítve, illetve napi lebontásban is. Az elérhető adatok mélységükben azonban jelentősen eltérőek az olasz régiók, és az olasz tartományok szintjei között. A régiók és tartományok szerinti felosztások közötti eltérést a 4. ábra szemlélteti.



4. ábra, Olaszország regionális (NUTS 2, baloldali) és tartományi (NUTS 3, jobboldali) felosztása. Saját ábra, az Eurostat adatai alapján (2020).

Az olasz tartományokat (107) be lehet osztani 21 régióba, amelyeket be lehet osztani 5 régiócsoportha. Ezek a felosztások az európai uniós Nomenclature of Territorial Units for Statistics rendszerének (magyarul Statisztikai Célú Területi Egységek Nomenklatúrája, röviden NUTS) felelnek meg, melynek célja a hasonló nagyságrendű gazdasági területek

statisztikáinak begyűjtése, a régiók szocioökonómiai elemzése, illetve az Európai Unió regionális politikájának kialakításának segítése (Központi Statisztikai Hivatal, 2020). Az adatbázisépítésnél is ezen területi felosztások mentén jelennek meg az elemzésben a térbeli adatok, az Eurostat adatbázisából. Ez a rendszer többek között a különböző adatok összekapcsolásában volt kifejezetten hasznos, hiszen az olasz Nemzeti Statisztikai Hivatal és Polgári Védelmi Minisztérium is ezeket a lebontásokat alkalmazva tette közzé az elemzéshez szükséges adatokat.

A COVID-19 terjedésének vizsgálatához a Polgári Védelmi Minisztérium által közzétett adatok regionális felosztás szerint részletesebb adatokat tartalmaznak mint a tartományi felosztás szerintiék. Ezek között találhatóak jelen elemzés számára irreleváns adatok is, azonban az elemzés épít olyan adatokra, amelyeket a tartományi adatbázis nem tartalmaz. Emiatt lett a regionális adatok felhasználása a kiindulópont. Ebben az adatbázisban többek között szerepel a régió középpontjának földrajzi elhelyezkedése, szélességi és hosszúsági fok szerint, a kórházban kezelt esetek száma tünetekkel, tünetmentesen, illetve az intenzív osztályon kezelték száma, valamint az összes eddig diagnosztizált koronavírusos betegek száma, a gyógyultak száma, az elhalálozottak száma, a jelenleg COVID-19 pozitív betegek száma vagy éppen az eddig tesztelt emberek száma.

Mivel az olasz Nemzeti Statisztikai Hivatalról szerzett adatok között többnyire olyanok szerepeltek, melyek a hivatalos honlapjukon tartományi szinten jelenleg nem elérhetőek, azonban regionális szinten igen, ezért innentől az elemzés kizárólag a regionális adatok felhasználásával halad tovább. A Statisztikai Hivatalból megszerzett adatbázisok általános gazdasági jellegű adatokat tartalmaznak, mint például a lakosság, vagy a régiók területe, illetve a lakosság egészségügyi hátteréhez kapcsolódnak, például a dohányzási és különböző alkoholfogyasztási szokásaikhoz, a testsúlyukhoz, vagy a krónikus betegségeik számához.

A Polgári Védelmi Minisztérium adatai a napi lebontás miatt lehetőséget teremtettek volna arra, hogy panel adatbázisként lehessen kezelni az adatokat, ezt a Nemzeti Statisztikai Hivatal adatainak tagoltsága nem tette lehetővé. Az utóbbi ugyanis éves, vagy ritkább lebontásban mér országos statisztikai adatokat, így ezekből csupán a 2017-es, 2018-as vagy 2019-es adatok voltak elérhetőek. Az elemzésben minden Nemzeti Statisztikai Hivatalról szerzett adat közül a jelenleg elérhető legfrissebb szerepel. A végleges adatbázis ezek után egészült ki a többi adat alapján létrehozott változókkal, mint például a népsűrűség vagy a regionális átlagéletkor.

Az adatbázisegyesítésnél merültek fel nehézségek. A Polgári Védelmi Minisztérium és a Nemzeti Statisztikai Hivatal sokszor ugyanazon régiókat hasonló, ám eltérő névvel látták el, nehezítve az adatbázisok egységesítését. Problémát jelentett továbbá, hogy a közelmúltban a régió kódok megváltoztak, így a térbeli adatbázishoz az egyesített koronavírussal kapcsolatos és egyéb statisztikai megfigyelések adatbázisát nem lehetett egyértelműen egymáshoz rendelni. Emellett, amíg a Polgári Védelmi Minisztérium és az Eurostat adatai között a jelenlegi NUTS rendszer szerinti 21 régió szerepel, az olasz Statisztikai Hivatal egy korábbi felosztás szerint is feltüntette az adatokat, amiket így 22 régióról közöltek. A problémát az északkelet-olaszországi Trentino-Alto Adige/Südtirol területe okozta, mely az előző évszázad egy részében egy területi egység volt, azonban ma már 2 autonóm régiót foglal magába; Bolzano-t és Trento-t.

4.2. Módszertan

Jelen elemzés egy többváltozós keresztmetszeti adatbázis ökonometriai elemzése, keresztmetszeti elemzési módszereket alkalmazva. Az elemzés elején az eredményváltozót leíró statisztikai eszközökkel elemzem, majd bemutatom a felhasznált magyarázó változókat, melyek az első, kezdetleges modellt alkotják. Az elemzésben a Legkisebb négyzetek módszerét (OLS) alkalmazom a modellépítésknél. A modelldiagnosztika keretei között Ramsey RESET tesztel megvizsgálom a modellspecifikációt, Breusch-Pagan tesztel a felmerülő heteroszkedaszticitást, F-próbán keresztül a modell relevanciáját, és VIF-mutatókon keresztül a magyarázóváltozók multikollinearitását.

Ezután a modellszűkítést először önállóan az egyetemi éveim alatt elsajátított ökonometriai szempontok szerint szűkítem, majd ellenőrzésképpen stepwise szelekció segítségével is megteszem ezt. A stepwise szelekcióhoz szükséges egy alapmodell, melyből kiindulva először a forward szelekciós eljárás szerint beemeli a modellbe a legjobb kimaradt magyarázó változót, majd a backward szelekcióhoz hasonlóan kidobja a legkevésbé relevánsat. Ezeket lépcsőről-lépésre addig végzi el a stepwise szelekció, amíg a megadott összehasonlítási kritérium alapján az új modell már nem növelné a modell jóságát. Az elemzésben a stepwise szelekció az Akaike információs kritérium alapján hasonlított össze a modellek jóságát.

Az elemzés továbbá kiegészül a témaválasztás miatt erősen indokolt térbeli autokorrelációval és térbeli regresszióval. A területek közötti térbeli autokorreláció azt

mutatja meg, hogy az egymáshoz közel fekvő térségekben megfigyelt adatok mennyire hasonlítanak egymásra (O'Sullivan és Unwin, 2014). A térbeli autokorreláció és autoregresszió vizsgálata azért kap jelen elemzésben kiemelt szerepet, mivel bármilyen járvány terjedése fizikai közelséghez kötött. A halálozás vizsgálatához hasznos információkat tudnak behozni a kezelt térbeli összefüggéseken keresztül. A térbeli autoregresszív modell úgy egészíti ki a hagyományos lineáris regressziót, hogy kontrolál a térbeli autokorrelációra, a szomszédos területek átlagos értékének maximum likelihood becslésével (Hijmans és Ghosh, 2019).

4.3. Korlátok

Az adatbázis összerakásából és a módszertanból adódóan az elemzésnek vannak limitációi. Az első ilyen abból adódik, hogy a vírus terjedésének időbeliségét a keresztmetszeti elemzési módszerek nem képesek kezelni. Habár a COVID-19-cel kapcsolatban idősorosan is rendelkezésre áll több adat Olaszországra regionális és tartományi szinten is, a többi vizsgálni kívánt adat típusa megakadályozza az idő dimenziójának bevonását a modellezésbe. Pusztán a COVID-19-ről közölt adatokat panelökonometriai eszközökkel vizsgálni funkciótlan lenne, hiszen az adatok többsége egymással szorosan összefügghet, vagy éppen egymás lineáris kombinációjaként megadható (az aktív fertőzöttek száma, a kigyógyultak száma és az elhunytak száma például egyértelműen meghatározza az összes fertőzött számát).

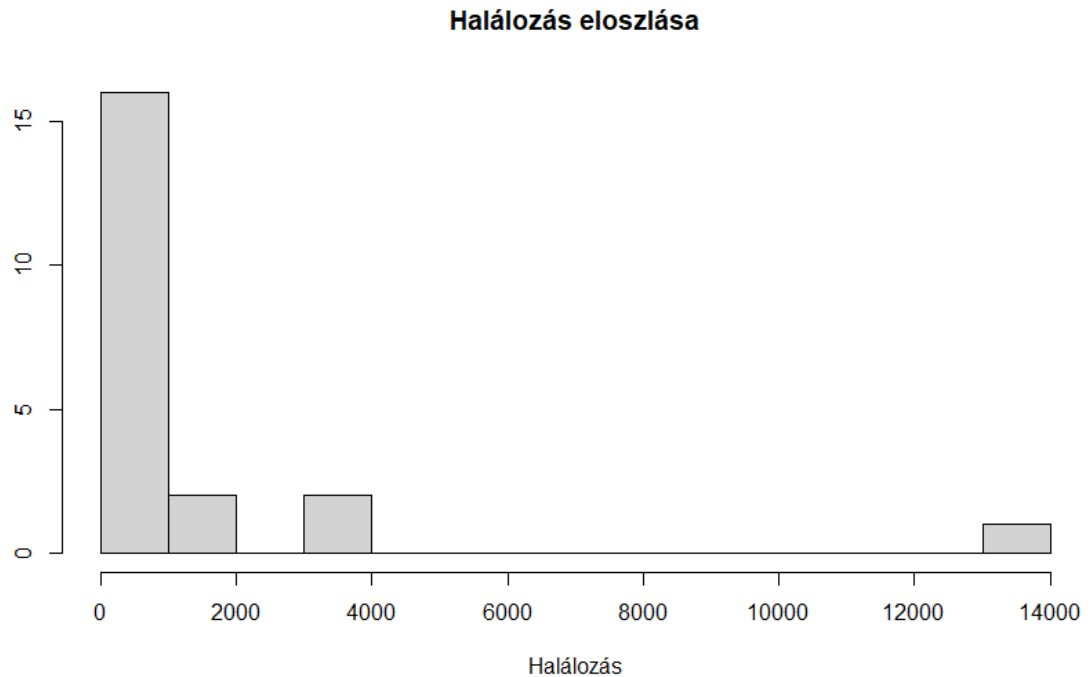
Ahogy a bevezetésben is megjelent, emiatt és az ehhez hasonló tényezők miatt az elemzés csak az április 30-ai adatokkal dolgozik tovább a későbbiekben. Az adatok emellett aktualitásukból adódóan eltérhetnek a világjárvány végén közölt adatoktól, hiszen több esetben nehézséget okozhat beazonosítani a halál kiváltó okát, a különböző régiók különböző módszereket alkalmazva becsülik meg a koronavírus fertőzésben elhunytak számát, illetve különböző sebességgel közvetítik az adatokat a központi adatbázis felé. Végül, az olasz Statisztikai Hivatal csupán a 2 vagy több krónikus betegséggel rendelkezőkről közöl adatokat az egész népességre, emiatt nem volt lehetőség arra, hogy csupán a 3 vagy annál több krónikus betegséggel rendelkezők kerüljenek be az elemzésbe, mint ahogyan a felhasznált irodalmakból következett volna.

5. Elemzés

5.1. Az eredményváltozó leíró statisztikai elemzése

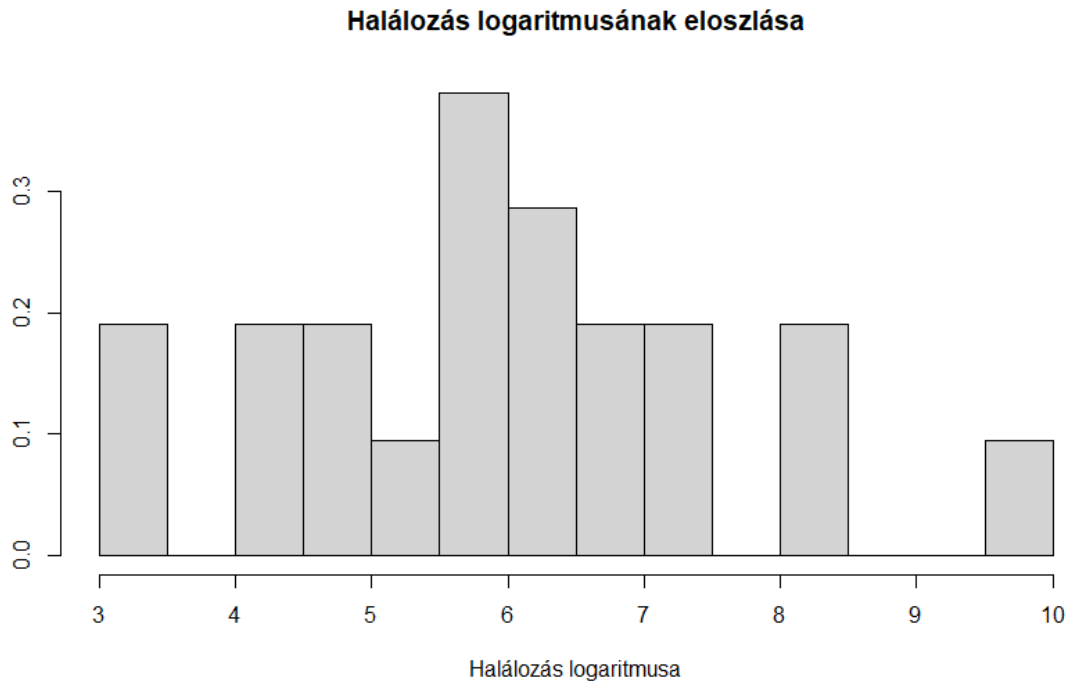
Jelen elemzés fókuszában a COVID-19 által okozott halálozások száma van. Az elemzés a 2020 április 30-ai adatokra építve arra keresi a választ, hogy milyen tényezők okozták Olaszországban a világjárvány kitörése óta megfigyelt kiemelkedő halálozási arányt. Egyes észak-olaszországi régiókban a 100.000 főre jutó halálozások száma meghaladta a 100-at is, míg több déli régióban ugyanez a mutató a 10-et sem közelítette meg. A legerősebben sújtott régióban a halálozások száma 655,8-szora a legkevésbé sújtott régióban tapasztaltnál képest. Az elemzés az Olaszországon belül megfigyelhető eltérések feltárására törekszik.

Lombardiában, a legnagyobb halálozások számával rendelkező régióban 13.772, míg Molise-ben, a legnagyobb halálozással rendelkező régióban 21 haláleset történt. Egy régióban átlagosan 1.332 haláleset volt, azonban a medián régióban csak 359 volt. Mivel az átlag közel négyszerese a mediánnak, ezért az adatbázis feltételezhetően tartalmaz egy vagy több kiugró értéket. Ez abból következik, hogy amíg a medián egy robosztus mutató, azaz nem érzékeny az outlierekre, az átlagot egy-egy kilógó érték erősen el tudja torzítani. A két középérték egymáshoz viszonyított elhelyezkedése továbbá arra enged következtetni, hogy a mutató egy erősen balra ferde eloszlást követ. Az 5. ábra az eredményváltozó eloszlását mutatja meg, mely igazolja ezt a következtetést.



5. ábra, A COVID-19 által okozott halálozások hisztogramja Olaszországban, regionális szinten. Saját ábra, az olasz Polgári Védelmi Minisztérium adatai alapján (2020).

A normalitás feltevés teljesülése érdekében, a változót eloszlásának alakjából kiindulva logaritmizálással kezeljük. Az így kapott változó eloszlását láthatjuk a 6. ábrán, mely már jobban megfelel a normalitás feltevésnek. A ferdesége (0,1974) és csúcsossága (2,9322) alapján a változó eloszlása a normálisnál egy kicsit csúcsosabb eloszlást mutat. Ezek alapján innentől az elemzésben a halálozások száma helyett ugyanennek a természetes alapú logaritmikus alakja lesz az eredményváltozó.



6. ábra, A COVID-19 által okozott halálozások logaritmusának hisztogramja Olaszországban, regionális szinten. Saját ábra, az olasz Polgári Védelmi Minisztérium adatai alapján (2020).

5.2. A magyarázó változók jellemzése

Az nullmodell felépítéséhez az elemzésben használt magyarázó változók szelektálása közben bontakozott ki az a jelenség, miszerint a mögöttes hatásokra több azonos, vagy közel azonos potenciális magyarázó változó állt rendelkezésre. Ezek kategóriákra oszthatóak, melyek többnyire átfednek a korábbiakban kiemelt potenciális magyarázó hatásokkal. A könnyen elkülöníthető kategóriák a következők: dohányzás, alkoholfogyasztás, testtömeg, krónikus betegségek, demográfiai mutatók, illetve egyéb COVID-19-es adatok. Ez a pont külön-külön bemutatja ezeket a kategóriákat.

A dohányzással kapcsolatos adatok további alkategóriákra oszthatóak. Tartalmaztak adatokat arról, hogy az adott régió népességének hány százaléka dohányzik, hogy a dohányzók napi hány cigarettát szívnak el átlagosan, illetve ötösével beosztva kategóriánként. Ezek eltérő információtartalmat hordoznak magukban, hiszen az egyik a népességben elkülönülő dohányzók csoportját jelöli ki, a másik pedig a dohányzókra korlátozva ad betekintést abba, hogy mennyire súlyos a régióban a dohányzás. Ebből kifolyólag a nullmodellbe a két változó interakciója került be. Az így kapott változó, az „Egy lakosra jutó naponta elszívott cigaretta” azt mutatja meg, hogy 1, az adott régióban elő ember átlagosan hány cigarettát szív el naponta.

Az alkoholfogyasztásra is széles körben érhetőek el adatok az olasz Statisztikai Hivatal honlapján. Ezek tartalmazzák azt, hogy mennyien fogyasztanak sört vagy bort naponta, illetve, hogy a 11 év feletti népesség hány százaléka szokott inni rendszeresen alkoholt, sört, bort, likőröket, vagy egyéb szeszesitalokat. Ezek közül az utóbbi három nagyjából ugyanannyira magyarázta a halálozások logaritmusát, az egyváltozós segédregressziók lefuttatása után. Emiatt ezek közül ezen a ponton egyik sem esett még ki, a végső nullmodell kiválasztásában ez a kategória volt a döntő tényező.

A testtömeg kategória négy változót foglalt magába: az elhízottak, a túlsúlyosak, a normál testsúlyúak és a soványok aránya az adott régió népességének. Mivel a halálozást az egészségügyi problémákkal kívánja az elemzés magyarázni, ennek a kategóriának egy részhalmazát kezelte az elemzés. A két változó, amelyet ebben a kategóriában kezeltek, a normális testsúlyúnál nagyobb, illetve kisebb testsúlyúak voltak. Az előbbi az elhízottak és a túlsúlyosak lineáris kombinációjaként lett megteremtve, az utóbbi pedig a soványokkal egyenlő. Mivel ezen a téren a szakirodalom csupán az átlagos testsúlynál nagyobb testsúlyúakat említette fokozottan kitettnek a COVID-19 által okozott komplikációknak, így a kettő közül csak az előbbi került be a nullmodellbe.

A következő kettő kategóriából a korábban felsorolt irodalmak miatt egyértelműbben lehetett a változók közül szelektálni. A krónikus betegségekre fellelhető adatok népességarányosan mutatják meg az adott régióban élők közül a különböző krónikus betegséggel rendelkezőket. Ebben a kategóriában az elérhető adatok közül a multikollinearitás kiszűrése és a legnagyobb magyarázó erő érdekében a 2 vagy több krónikus betegséggel élők aránya került be a nullmodellbe. A demográfiai adatokból létrehozott változók közül a népsűrűség, a férfi-női arány és az átlagéletkor került be. Emellett bekerült egy Észak nevű dummy változó, amely a historikus Észak-Olaszország régióit foglalja magába. Ez a következő nyolc régiót jelenti: Bolzano, Emilia-Romagna, Friuli-Venezia Giulia, Liguria, Lombardia, Piemont, Trento, Valle d'Aosta és Veneto.

Az egyéb COVID-19-es adatok többek között tartalmazták az összes megbetegedés, az intenzív osztályon kezelték és az összes koronavírus-tesztelt emberek számát is. Ebben a kategóriában az endogenitási probléma miatt az előbbi kettő változó használatát elvettem. A korábbi összes megbetegedés majdnem 100 százalékban magyarázta a halálozások számát. Az elemzésnek ellenben nem az a célja, hogy két COVID-19-ből fakadó változó közötti egyértelmű kapcsolatot tárja fel.

Továbbá, az intenzív osztályon kezelték száma részben igazolható azzal, hogy azokban a régiókban, ahol jelenleg magas az intenzíven kezelték aránya, sokkal leterheltebb az egészségügy, emiatt többen is halnak bele a járványba. Azonban, ez a változó csak a koronavírusos betegeket tartalmazza az intenzív osztályon kezelek közül, így nem meglepő, hogy ez a változó is több mint 85 százalékban magyarázta önmagában is a halottak számát. A koronavírus-tesztelt emberek számát ellenben többnyire olyan tényezők befolyásolják, mint az elérhető tesztek száma, vagy az arra vonatkozó helyi kormányzati irányelvek, így ez a változó kisebb eséllyel és kisebb mértékben függhet össze a COVID-19 nem mérhető tulajdonságaival, így ezt a változót be lehetett tenni a nullmodellbe.

5.3. Az alapmodell kiválasztása

A fentiek alapján a nullmodell négyféleképpen nézhetne ki. Mindegyik potenciális nullmodell tartalmazza a következő változókat: Egy lakosra jutó naponta elszívott cigaretta, Túlsúlyos vagy elhízott, 2 vagy több krónikus betegséggel élők, Népsűrűség, Férfi-női arány, Átlagéletkor, Észak, Összes tesztelt ember. Ezek mellett az alkoholfogyasztás mérésére a bort, likőröket, és az egyéb szeszes italokat fogyasztók arányát vizsgálja az 1. táblázat, külön-külön, illetve az összeset együtt is.

Regressziós eredmények

	A halálozások számának logaritmusa			
	nullmodell	nullmodell	nullmodell	nullmodell
	(1)	(2)	(3)	(4)
Egy lakosra jutó naponta elszívott cigaretta	-0.8533 (1.0021)	-0.8432 (0.9810)	-1.0088 (0.9815)	-0.9264 (1.0919)
Borfogyasztás	-0.0337 (0.0870)			-0.0248 (0.1189)
Likőr fogyasztás		-0.0674 (0.0847)		-0.0552 (0.1136)
Egyéb szeszesital-fogyasztás			-0.1201 (0.1263)	-0.0716 (0.1873)
Túlsúlyos vagy elhízott	-0.0583 (0.0762)	-0.0346 (0.0794)	-0.0870 (0.0803)	-0.0580 (0.1072)
2 vagy több krónikus betegséggel élők	-0.0586 (0.0855)	-0.0677 (0.0847)	-0.0819 (0.0869)	-0.0829 (0.0951)
Népsűrűség	0.0029 (0.0035)	0.0039 (0.0027)	0.0027 (0.0029)	0.0026 (0.0037)
Férfi-női arány	-34.4662 (22.6484)	-25.1362 (16.9160)	-35.7378* (17.8525)	-34.3966 (24.0005)
Átlagéletkor	-0.0215 (0.1761)	-0.0064 (0.1730)	-0.0112 (0.1705)	-0.0058 (0.1867)
Észak	0.6608 (0.7134)	0.4071 (0.5787)	0.7179 (0.6056)	0.6728 (0.7583)
Összes tesztelt ember	0.00001** (0.00001)	0.00001** (0.000005)	0.00001** (0.000005)	0.00001* (0.00001)
Constant	45.5974 (26.6883)	35.3375* (17.6748)	49.7796** (21.2503)	48.5690 (28.5733)
N	21	21	21	21
R ²	0.8331	0.8400	0.8437	0.8477
Adjusted R ²	0.6965	0.7091	0.7158	0.6615
Residual Std. Error	0.8766	0.8582	0.8483	0.9257
F Statistic	6.1005***	6.4181***	6.5964***	4.5539**

Notes:

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

1. táblázat, a potenciális nullmodellek. Saját táblázat, az Adatbázisépítésnél felsorolt adatok alapján.

A korrigált R négyzet alapján az egyéb szeszesitalok illene a legjobban a nullmodellbe, azonban ez a második legjobb korrigált R négyzetet okozó változóhoz képest jelentős mértékben növelte a modellben a multikollinearitást. A multikollinearitás mértéke a variancia infláló faktorra (VIF-mutatóval) volt tesztelve; a Szeszesital-fogyasztás 4,77-es VIF-mutatóval, míg a Likőr fogyasztás 1,92-es VIF-mutatóval rendelkezett. A variancia infláló faktor azt mutatja meg, hogy a mintavételi variancia hányszorosára nő pusztán amiatt, hogy a magyarázó változók egymást is magyarázzák (Kovács, 2008). A nullmodell kiválasztásánál a többi magyarázó változó VIF-mutatóját is megvizsgáltam. Ezek mind 5 alatti értéket vettek fel, így kijelenthető, hogy káros multikollinearitás nem található a modellben.

Ezt követően egy Ramsey RESET teszttel a modellspecifikáció is tesztelésre került. Ez a teszt a magyarázó változók magasabb polinómjai segítségével (az elemzésben a kvadratikus és a kubikus alakjaival) tesztelte, hogy a modellspecifikáció helyes volt-e, vagy szükség van-e arra, hogy más függvényformájú változókat is beemeljünk a modellbe. A teszt nullhipotézise szerint a modellspecifikáció helyes volt, és a 0,11-es p-érték mellett ezt a nullhipotézist nem tudjuk elutasítani. Új változók bevonására tehát nem volt szükség, így ezután kezdődött a modellszűkítés.

5.4. Modellszűkítés

A modellszűkítés kétféle módszer alapján történt meg. Először a modell magyarázóereje és a multikollinearitás alapján önállóan álltam neki szűkíteni a modellt, majd ezután egy stepwise szelekciós módszert is alkalmaztam ugyanerre a problémára. Ez azért volt indokolt, mert ez a módszer az Akaike-mutatók alapján állította össze a legjobb modelleket, és a térbeli regressziót az OLS regresszióval éppen az információs kritériumok alapján lehet a legegyszerűbben összehasonlítani.

Első lépésként én az Átlagéletkort távolítottam el, hiszen amellet, hogy a legmagasabb empirikus szignifikanciaszinttel rendelkezett (több mint 0,9-et vett fel), a VIF-mutatója is ennek volt a legnagyobb. Ennek az elhagyása növelte is a korrigált R négyzetet, és csökkentette is a többi változó multikollinearitását. Az átlagéletkor elhagyása, és ennek hatásai a modellre azt jelentik, hogy Olaszországon belül már nem hatott az olasz régiók koreloszlása. Míg többen kiemelték potenciális magyarázó erőként az olasz társadalom korát, ennek Olaszországon belül már nincsen szignifikáns hatása; nincsen szignifikáns kapcsolat az átlagosan öregebb régiók és a COVID-19 halálozási rátája között. A 80 év,

vagy a feletti arányát a társadalomban teszteltem az átlagéletkor változó helyett, azonban ugyanúgy arra jutottam, hogy ezt a változót el kell hagyni. Ez az előbbi állítást még jobban alátámasztja.

A következő változó, melyet elhagytam a modellből a Túlsúlyos vagy elhízott változó volt. Ez a változó regionális szinten mutatta meg a normális testtömegnél nagyobb testtömeggel rendelkezők arányát az adott régióban. Az Átlagéletkor elhagyása után ez rendelkezett a legmagasabb, a többihez képest kiemelkedően magas p-értékkel, illetve a változó kivétele ebben az esetben is javította a modell multikollinearitásának szintjét. Mivel egyes források csak az elhízottakra tartották fent a komplikációk kialakulásának fokozott veszélyét, ezért ehelyett a változó helyett az Elhízottak változóval is le lett tesztelve az új modell, azonban közel azonos eredmények mellett ezt a változót sem tartottam meg. Ez azt bizonyítja, hogy Olaszországban a COVID-19-ben elhunytak száma nincsen erős összefüggésben a testtömeggel.

Harmadik lépésként a Népsűrűség változót vettem ki. Ezt az indokolta, hogy a Túlsúlyos vagy elhízott változó után már csak kettő változó VIF-mutatóinak volt 2-nél nagyobb értéke: a Népsűrűségnek és az Összes tesztelt embernek. Mivel ebben a mutatóban és a szignifikanciájukban sem volt jelentős eltérés a kettő változó között, ezért lefuttattam két segédregressziót; az egyikből a Népsűrűséget hagytam ki, a másikkól az Összes tesztelt embert. A modell magyarázóerejében már volt észrevehető különbség, a korrigált R négyzete körülbelül 10 százalékponttal magasabb volt annak a modellnek, melyből a Népsűrűség lett kihagyva.

Ezek alapján megállapítható, hogy a népsűrűség és az összes ember, akin végeztek el koronavírustesztet szorosan összefügg. A két változó egymásra-hatása már a VIF-mutatókban is kiütközött, azonban a segédregressziók rávilágítottak arra, hogy míg a tesztelesekre kontrollálás nélkül a népsűrűségnek is minden szokásos szignifikanciaszinten szignifikáns hatása volt a halálozások számára, az elvégzett tesztek száma szorosabban függött vele össze. A két változó közötti multikollinearitást továbbá alátámasztja az is, hogy az összes VIF-mutató értéke 2 alá csökkent ezután a lépés után, azaz a modellben már zavaró multikollinearitásról sem beszélhettünk.

A korrigált R négyzet a Népsűrűség kihagyásával már csökkent, azonban további két változó kihagyásával egy közel azonos szintre vissza lehetett emelni. Ez a két változó az Egy lakosra jutó naponta elszívott cigaretta és a Likőr fogyasztás volt. Ez a jelenség azt

is igazolhatja, hogy a dohányzásnak és az alkoholfogyasztásnak kevesebb köze van a COVID-19 által okozott halálozásokhoz, mint azt a szakirodalom sejteti. Azonban a tény, hogy a 2 vagy több krónikus betegséggel élők változó VIF-mutatója egyre csökkent, illetve a változó egyre szignifikánsabbá vált a többi egészséggel kapcsolatos változó kiesésével (mint például ez a kettő, vagy a Túlsúlyos vagy elhízott) azt erősítheti meg, hogy ezek a változók közrejátszhatnak a krónikus betegségek kialakulásában, ezáltal közvetve hatva a halálózásra.

5.5. A végleges lineáris modellek

Ezután már a magyarázó változók többsége szignifikáns lett, ezek közül nem lehetett tovább szelektálni anélkül, hogy a korrigált R négyzet ne csökkent volna jelentősen. Káros, de még csak zavaró multikollinearitásról sem beszélhetünk továbbá, hiszen az összes változó VIF-mutatója 2 alatti lett. Az önálló szűkítés által létrehozott végső modellben a következő változók maradtak: 2 vagy több krónikus betegséggel élők, Férfi-női arány, Észak, Összes tesztelt ember. A modell egészének relevanciáját F-próbával tudjuk letesztelni, a próbához tartozó p-érték közel 0 lett, így a regressziós modellben van szignifikáns változó, azaz a modell releváns. Továbbá, erre a modellre is lefuttattam egy Ramsey RESET tesztet, melynek p-értéke 0,64 lett, azaz nem tudjuk elutasítani a nullhipotézist, tehát helyes a modellspecifikáció. A lefuttatott Breusch-Pagan teszt p-értéke 0,36 lett, azaz itt sem tudjuk elutasítani a nullhipotézist, tehát homoszkedaktikus a modell.

Az önálló modellszűkítés után stepwise szelekciót is futtattam az eredeti nullmodellre. A stepwise szelekció többszöri futtatás után képes eltérő eredményeket adni, a módszer működéséből adódóan, ezért az elemzés húszszori futtatás mellett foglalkozott ezekkel. Mivel a húszból több, mint 15-ször ugyanaz a modell jött ki, ezért a pótlólagos futtatásnak már nagyon alacsony hozzáadott értéke lett volna. Ez a módszer három különbözőféle modellt adott ki. Az egyik a három közül egybevágó volt az önálló szűkítés alapján létrehozottal, a másik kettő pedig nagyon hasonlított rá; a 2. modell, amit kiadott a 2 vagy több krónikus betegséggel élők, Férfi-női arány és az Összes tesztelt ember változókat tartalmazta, míg a 3. modell, amit kiadott a Férfi-női arány, az Észak és az Összes tesztelt ember változókat. A modelleket bővebben jellemzi a 2. táblázat.

Végleges regressziók

	A halálozások számának logaritmusa		
	Önálló modell Stepwise 1 Stepwise 2		
	(1)	(2)	(3)
2 vagy több krónikus betegséggel élők	-0.1131 (0.0685)	-0.1719** (0.0635)	
Férfi-női arány	-35.0939** (13.1353)	-37.3688** (13.8618)	-27.0970** (12.8149)
Észak	0.7653* (0.4335)		1.1134** (0.3976)
Összes tesztelt ember	0.00002*** (0.000003)	0.00002*** (0.000003)	0.00002*** (0.000003)
Constant	40.4367*** (13.1781)	44.0986*** (13.8001)	30.2268** (12.2141)
N	21	21	21
R ²	0.7943	0.7543	0.7593
Adjusted R ²	0.7429	0.7109	0.7168
Akaike Inf. Crit.	56.8675	58.6047	58.1719
Residual Std. Error	0.8068	0.8555	0.8468
F Statistic	15.4504***	17.3959***	17.8761***

Notes:

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

2. táblázat, a potenciális végleges modellek. Saját táblázat, az Adatbázisépítésnél felsorolt adatok alapján.

A stepwise eljárás során kapott modellek ugyanazokat a változókat tartalmazzák, melyek az önálló szűkítés eredményeként jöttek ki. A három modell összehasonlítása során jól látható, hogy az önálló szűkítésben kijött modellnek van a legmagasabb magyarázó ereje, az Akaike információs kritérium is itt a legalacsonyabb, illetve még a maradéktagok standard hibái is itt a legalacsonyabbak. A három modell F-tesztjei alapján minden szokásos szignifikanciaszinten relevánsak a modellek, a Ramsey RESET tesztek alapján a modellspecifikációik jók, azonban Stepwise 1 modell Breusch-Pagan heteroszkedaszticitás tesztjének p-értéke alapján a nullhipotézist csupán 1 százalékos szignifikanciaszinten nem tudjuk elutasítani, így magasabb szignifikanciaszintek mellett a modell heteroszkedaszticitását kell feltételeznünk.

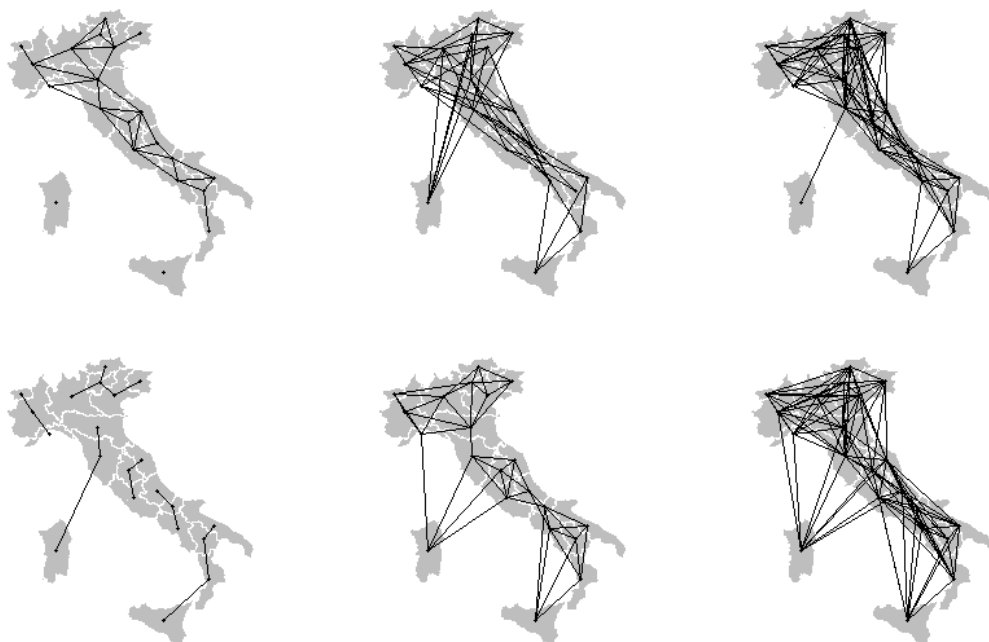
5.6. Térbeli autokorreláció

Waldo Tobler alkotta meg a Földrajz Első Törvényét, ami azt mondja ki, hogy „minden mindennel összefügg, de a közelebbi dolgok jobban összefüggenek, mint a távolabbi dolgok” (Tobler, 1970, 236. o.). Tobler törvénye természetesen csupán földrajzi kontextusban értelmezendő, azonban ez az állítás erősen meghatározta a hozzáállást a későbbi térbeli elemzésekhez. Ez adja meg az elméleti alapjait például a jelen elemzésben is szereplő térbeli autokorrelációnak és a térbeli autoregresszióknak is.

Ezekhez egyaránt elengedhetetlen a térbeli súlyok meghatározása, amely mellett az autokorrelációt vizsgáljuk. Térbeli súlynak azt nevezzük, ami kijelöli számunkra, hogy a Földrajz Első Törvényében megfogalmazott összefüggést milyen súlyok szerint szeretnénk mérni. A leggyakrabban használt térbeli súlyok a következők: a Szomszédság elve, a Közvetett szomszédság elve, a Sugár szomszédság elve, a Legközelebbi szomszédok elve és az Inverz távolság elve.

Az első kettő ezek közül a két legegyszerűbb módszer; a Szomszédság elve pusztán az adott régióval szomszédos régiókon teszteli az autokorrelációt, míg a Közvetett szomszédságnál ennek a szomszédnak az egyéb szomszédjai is a számításba kerülnek. A kettő között fontos különbség, hogy a második képes kezelni a közvetlen szomszéd nélküli területeket is.

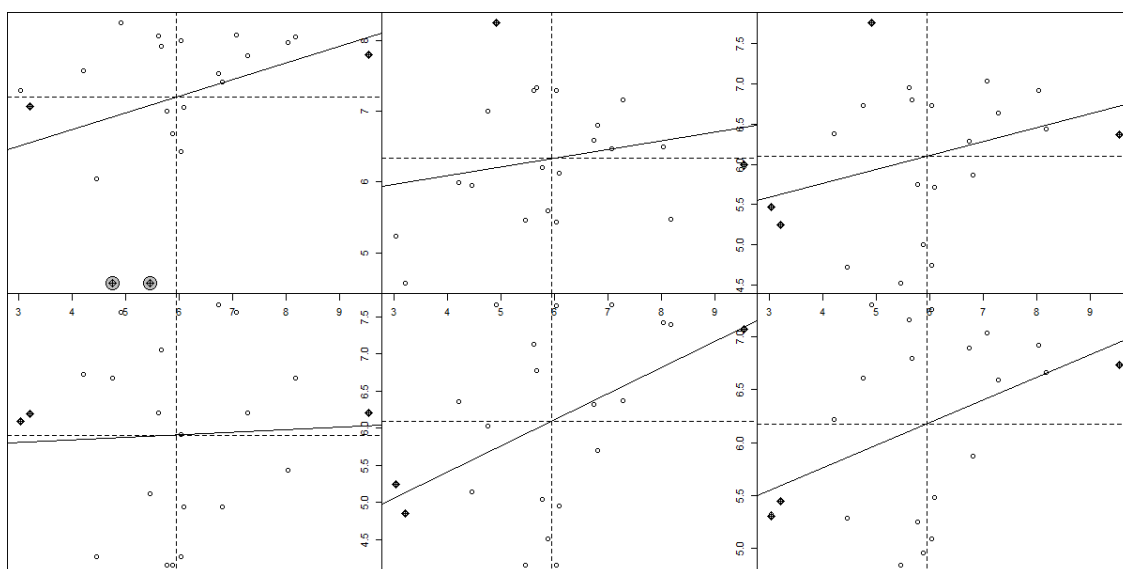
A Sugár szomszédság elve az adott terület mértani középpontjához a megadott sugár szerint írható körön belüli régiókkal számol, amennyiben ezen régiók mértani közepe beleesik az adott sugár által kijelölt körbe. A Legközelebbi szomszédok elve hasonló alapokra épül, mint a Sugár szomszédság, azonban itt nem a sugár méretét határozzuk meg, hanem a hozzá legközelebbi mértani középpontú területek számát, amikkel számolni szeretnénk. Az Inverz távolság elve a legkomplexebb ezek közül, hiszen itt a földrajzilag legtávolabb eső területek is beleszámítanak az autokorrelációba, azonban sokkal kisebb súllyal, mint a közelebb eső területek.



7. ábra, A leggyakoribb térbeli súlyok applikálása Olaszországra. A felső sor balról jobbra: a Szomszédság elve, a Közvetett szomszédság elve és a Sugár szomszédság elve, 400 km-es sugárral. Az alsó sor balról jobbra: Legközelebbi szomszédok elve, 1, 4, illetve 8 legközelebbi szomszéd mellett. Saját ábra, az Eurostat adatai alapján (2020).

A fent felsorolt leggyakoribb térbeli súlyok közül a 7. ábrán az első négy látható. A Sugár szomszédság elvénél a sugár kiválasztásánál a vizsgált sugarak közül (100-tól 1000 kilométerig, 100 kilométerenként) a térbeli autokorreláció erőssége alapján lett a 400 kilométer kiválasztva. A Legközelebbi szomszédok elvénél szintén a térbeli autokorreláció erőssége alapján történt meg a 4 legközelebbi szomszéd kiválasztása. (Az 1 és a 8 legközelebbi szomszéd a Legközelebbi szomszédok elvénél tapasztalt szélsőségek kontrasztjainak szemléltetése miatt szerepel az ábrán.) Az Inverz távolság elve definíciójából adódóan az összes régiót összekötné az összes régióval. Az elemzés ezt az elvet komplexitása miatt nem tartalmazza a továbbiakban.

A térbeli súlyok vizualizációján túl a Moran-féle I-index kiszámítása segít eldönteni, hogy mely térbeli súlyozásokat érdemes, vagy nem érdemes alkalmazni később az elemzés során. A Moran-féle I-index nullhipotézise szerint a vizsgált területek között nincsen térbeli autokorreláció. Amennyiben a teszt empirikus szignifikanciaszintje nagyon alacsony (0,01-nél kisebb), az összes szokásos szignifikancia szint mellett elvetjük ezt a nullhipotézist, ami azt jelenti, hogy a vizsgált területek között létezik területi autokorreláció, azaz fennáll jelen esetben is a Földrajz Első Törvénye. A Moran-féle I-indexek értelmezéséhez a 8. ábra nyújt segítséget.



8. ábra, A leggyakoribb térbeli súlyok melletti Moran-féle I-indexek, a halálozások számának logaritmusával számolva. A felső sor balról jobbra: a Szomszédság elve, a Közvetett szomszédság elve és a Sugár szomszédság elve melletti (400 km-es sugárral) I-indexek. Az alsó sor balról jobbra: Legközelebbi szomszédok elve, 1, 4, illetve 8 legközelebbi szomszéd melletti I-indexek. Saját ábra, az Eurostat adatai alapján (2020).

A 8. ábra az elemzés fő fókuszában lévő halálozások számán keresztül szemlélteti a térbeli autoregressziót az eltérő térbeli súlyok alkalmazása mellett. A halálozások számának itt is a természetes alapú logaritmusai jelennek meg. A Moran-féle I-indexeket ezen az ábrán az egyenesek meredekségei mutatják meg, ezek balról jobbra és felülről lefelé haladva a következők: 0,3156; 0,1226; 0,1725, 0,0508, 0,3528 és 0,2142. A Moran-féle I-index tesztelésénél a p-érték alapján 5 százalékos szignifikanciaszint mellett a hat közül kettőnél, a Közvetett szomszédság elvénél és a Legközelebbi szomszédok elvénél 1 legközelebbi szomszéd mellett nem tudjuk elvetni a nullhipotézist. Ez azt jelenti, hogy ezen súlyozások alapján az olasz régiók között nincs térbeli autokorreláció, azaz nem áll fenn a Földrajz Első Törvénye.

Továbbá, a Szomszédság elve melletti súlyokat a két sziget, Szicília és Szardínia miatt szintén el kellett vetni, hiszen bár 5 százalékos szignifikanciaszint mellett a térbeli autokorreláció fennáll Olaszország régiói között, ez az eredmény nem tartalmazza a szigeteket, mely így lekorlátozná az elemzést a kontinensen fekvő régiókra. Az összes vizsgált szomszédsági elv közül a tesztek alapján az jött ki, hogy a Legközelebbi szomszédok elve, 4 legközelebbi szomszéd mellett eredményezi a legerősebb autokorrelációt. Ennek értéke 0,3528, ami így is egy közepesen gyenge térbeli kapcsolatot mutat meg a területek között. Térbeli autoregresszív modellt alkalmazni abban az esetben

célszerű, ahol a területek közelsége a tesztelt térbeli súlyok mellett a legjobban összefügg, így a továbbiakban ezzel a szomszédsági elvvel dolgozik tovább az elemzés.

5.7. Végleges lineáris modellek és a Térbeli autoregresszív modellek

Mivel az elemezni kívánt eredményváltozó egy járvánnyal kapcsolatos mutató, ezért a terjedésének térbeli tulajdonságaival is érdemes foglalkozni a végső modell kiválasztásánál. Az olasz régiók közötti térbeli összefüggéseket a korábban megállapítottakkal összhangban a Legközelebbi szomszédok elve alapján súlyozzák a térbeli regressziók is, 4 legközelebbi szomszéd mellett. A végleges lineáris modelleket, illetve az ezekre lefuttatott térbeli autoregresszív modelleket a 3. táblázat összegzi. Mivel ez utóbbi típusú modell nem számol determinációs együtthatót, ezért a modelleket az Akaike-mutatóik alapján lehet könnyen összehasonlítani.

A 3. táblázatban felváltva szerepelnek a modellszűkítés során kiválasztott modellek, és ugyanezen modellek kiegészítve térbeli autoregresszióval. Az utóbbi modellek ugyanazokon a magyarázó változókon keresztül magyarázzák az eredményváltozót, azonban az adatok térbeliségét is számításba veszik. A térbeli autoregresszív modell vizsgálja azt, hogy a megadott szomszédsági elv szerinti szomszédokban az eredményváltozóban bekövetkező változás hogyan hat az adott régióban megfigyelhető eredményváltozó változására. A Wald-teszt és a Likelihood Ratio (LR) teszt azt vizsgálja, hogy a térbeli autoregresszió van-e szignifikáns hatása a megadott modellre.

Ezek alapján a tesztek alapján 5 százalékos szignifikanciaszinten a térbeli kapcsolatokra kontrolálni a 3 lineáris modell közül egyedül a Stepwise 1-es modellnél éri meg. Ezt indokolhatja az, hogy a másik kettő modell tartalmazza az Észak nevű dummy-t, melynek eredeti célja a historikus Észak- és Dél-Olaszország törésvonal hatásának vizsgálata miatt került be. Azonban, a két országrész közötti törésvonal, ami ma már elsősorban nem a földrajzi elhelyezkedésben manifesztálódik, egy meghatározó területi aspektus, ami egy térbeli regressziónál elronthatja a térbeli autoregresszív modell használatának jogosultságát. Jelen esetben is ez történhetett, az Észak nevű dummy a mögöttes törésvonalak reprezentálása mellett feltételezhetően lefedheti a térbeli különbségeket is, hiszen az (1)-es és a (4)-es modell több szempont szerint hasonlóknak mondható.

Végleges regressziók - térbeli autoregresszív modellekkel

	A halálozások számának logaritmusa					
	<i>OLS</i>	<i>térbeli</i>	<i>OLS</i>	<i>térbeli</i>	<i>OLS</i>	<i>térbeli</i>
		<i>autoregresszív</i>		<i>autoregresszív</i>		<i>autoregresszív</i>
	Önálló modell	Önálló modell	Stepwise 1	Stepwise 1	Stepwise 2	Stepwise 2
	(1)	(2)	(3)	(4)	(5)	(6)
2 vagy több krónikus betegséggel élők	-0.1131 (0.0685)	-0.1202** (0.0588)	-0.1719** (0.0635)	-0.1335** (0.0550)		
Férfi-női arány	-35.0939** (13.1353)	-33.4985*** (11.3516)	-37.3688** (13.8618)	-32.9330*** (11.3956)	-27.0970** (12.8149)	-25.9077** (11.5731)
Észak	0.7653* (0.4335)	0.3566 (0.5302)			1.1134** (0.3976)	0.8936 (0.5624)
Összes tesztelt ember	0.00002*** (0.000003)	0.00002*** (0.000003)	0.00002*** (0.000003)	0.00002*** (0.000003)	0.00002*** (0.000003)	0.00002*** (0.000003)
Constant	40.4367*** (13.1781)	38.1414*** (11.5400)	44.0986*** (13.8001)	37.3891*** (11.5865)	30.2268** (12.2141)	28.5617** (11.2480)
<i>N</i>	21	21	21	21	21	21
<i>R</i> ²	0.7943		0.7543		0.7593	
Adjusted <i>R</i> ²	0.7429		0.7109		0.7168	
Akaike Inf. Crit.	56.8675	58.2948	58.6047	56.6020	58.1719	60.0070
Wald Test		0.8544		4.3721**		0.2087
LR Test		0.5727		4.0028**		0.1649

Notes:

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

3. táblázat, a potenciális végleges modellek, kiegészítve a térbeli autoregresszív modellekkel.

Saját táblázat, az Adatbázisépítésnél felsorolt adatok alapján.

A végleges modell kiválasztásához az OLS regressziókat és a térbeli autoregresszív modelleket az Akaike-mutató alapján hasonlítottam össze. Ez alapján a stepwise szelekció során kiválasztott modellek közül a Stepwise 1 modell térbeli autoregresszióval kiegészülve lett a legjobb modell. Az információs kritérium csökkenése továbbá azt is jelenti, hogy a modell magyarázó erejét is megnövelte a térbeli adatokra kontrollálás alkalmazása.

5.8. A végleges modell értelmezése

A végleges regresszió azt mutatja meg, hogy Olaszországban regionálisan 2020. áprilisának végével bezárólag milyen tényezők határozták meg a legjobban a COVID-19 által okozott halálozást. A végleges modellben három magyarázó változó, illetve a halálozás logaritmusának térbeli késleltetettjeinek együtthatója maradt. Ennek értéke 0,2783 lett, ami azt jelenti, hogy ha a szomszédos területeken 1 százalékkal nő a halálozások száma, akkor minden más változatlansága mellett átlagosan várhatóan 0,2783 százalékkal nő az adott régióban a halálozások száma. Az emellett bent maradt magyarázó változók a következők: 2 vagy több krónikus betegséggel élők, Férfi-női arány, Összes tesztelt ember.

A 2 vagy több krónikus betegséggel élők nevű változó a régióban ezekkel élők arányát adja meg. Ez a változó tehát egy százalékos mutató, így a becsült regressziós együtthatót úgy kell értelmezni, hogy minden más változatlanságát feltételezve, ha 1 százalékponttal nő a krónikus betegséggel élők aránya a népességben, akkor a halálozások száma átlagosan várhatóan 0,1335 százalékkal csökken. Ez meglepő eredmény, hiszen az egészségügyi állapotok romlásával a halálozás növekedését asszociálnánk.

Hasonlóan váratlan eredményeket kapunk, amikor a férfi-női arányt vizsgáljuk. Mivel ez a változó is egy arányszám, ezért ugyanúgy kell értelmezni, mint az előzőt. Ha 1 százalékponttal nő a régióon belül a férfi-női arány, akkor átlagosan várhatóan 32,93 százalékkal csökken a halottak száma. Ez látszólag ellentmondásban van azzal, miszerint a férfiak halálozási rátája jóval magasabb a nőkéhez képest. Amennyiben nőne a férfiak részaránya, azt várnánk, hogy a várható halálozás is növekedjen.

Ezekre az anomáliákra magyarázatot adhat az elemszám, illetve a halálozási adatokban megfigyelhető outlier. Lombardia régióban jegyezték fel az olaszországi halálozások több mint felét, amely alacsony elemszám mellett torzíthatja a becsült koefficienseket, hiszen Lombardia egyes tulajdonságainak túl nagy jelentőséget tulajdoníthat a modell. Ennek

tesztelésére leszűrtem az adatbázist, mely így már nem tartalmazta a lombardiai adatokat, azonban a Stepwise 1 modellben továbbra is negatív maradt a becsült koefficiensek előjele. Továbbá az eredeti nullmodellre a szűkített adatbázisra is lefuttattam stepwise szelekciót, azonban az így kapott modell becsült koefficiensei is mind negatívak voltak, az összes tesztelt emberen kívül.

Továbbá, magyarázhatta volna a koefficiensek előjelét az is, hogy az összes tesztelt emberre közvetetten hat egy kihagyott változó, mint például a koronavírus járvány intenzitása, amely a halálozásra közvetlenül is hathat. Ennek érdekében kivettem ezt a változót a kapott modellből, azonban az előjelek továbbra is negatívak maradtak. Ezekből adódóan kijelenthető, hogy pusztán (a lombardiai adatokat is tartalmazó) végleges modell összetétele nem torzítja a koefficiensek előjeleit.

Az Összes tesztelt emberek az adott régióban az összes embert jelenti, akin elvégeztek koronavírus-tesztet. Fontos elkülöníteni az adott régióban elvégzett tesztek számával, hiszen az előbbi az embereket egyedi megfigyelésként kezeli. Mivel koronavírus-tesztek több ezer emberen végeztek el régióként (a változó minimuma 5.647, a maximuma 229.880 volt), így a változó becsült koefficiensének egy indokoltabb értelmezése a következő: Amennyiben egy adott régióban 1.000-el több embert tesztelnének, a halálozás átlagosan várhatóan 2 százalékkal nőne, minden más változatlanóságát feltételezve.

A szakirodalom szerint a minél több elvégzett teszt csökkenti a halálozási arányt, azonban az elemzésben ennek az ellenkezője jött ki. Ez tovább erősíti azt a gondolatot, hogy egy exogén változó (például a COVID-19 intenzitása) hat a halálozásra és a tesztek elvégzésére is. Ez jelentheti azt, hogy azokon a területeken, ahol már megjelent a vírus, és elkezdett egyre nagyobb arányban halálos áldozatokat is gyűjteni, ott több tesztet végeztek el később, de például azt is, hogy azokban a régiókban, ahol kevesebb tesztet végeztek el, ott kevesebb halálesetet tulajdonítottak a COVID-19-nek, pusztán az alulinformáltságuk miatt.

Ez utóbbi problémák kiküszöbölésére lehetne válasz a panel adatelemzés, azonban egyéb releváns magyarázó változók hiányában ennek korlátozott keretek között van jelenleg értelme. A járvány lecsendesedettével feltételezhetően több lehetőség nyílik majd különböző adatok begyűjtésére, mely hasznosabb összefüggésekre is fel tudná hívni a válságkezelők figyelmét. Emellett, amennyiben utólag a regionális szint mellett a

tartományi szinten is szélesebb körben lesznek elérhető adatok, úgy az elemszám méretéből adódó limitációk is feloldódhatnak.

6. Konklúzió

Jelen dolgozat arra a kérdésre kereste a választ, hogy Olaszországon belül milyen tényezők befolyásolták döntően a COVID-19 által okozott halálozások rendkívül egyenetlen eloszlását, 2020 áprilisának végéig. Az elemzés a legfrissebb szakirodalmak alapján azonosított potenciális kormányzati, egészségügyi és demográfiai okokat. A kormányzati okok közé tartoznak az olasz kormány intézkedései a járvány kirobbanásának elején, illetve ezeknek az intézkedéseknek az időzítése. Lehetséges egészségügyi okok alatt a dohányzást, a komoly alkoholfogyasztást, az elhízást és a krónikus betegségeket azonosította az elemzés. Demográfiai összefüggések alatt az olasz népesség korösszetételét, a nemek közötti különbségeket, és egyéb tényezőket ismert fel a dolgozat mint például a népsűrűség.

A dolgozat először bemutatta a jelenleg tomboló világjárványt és az azt okozó vírust. Ebben a szekcióban részletezte továbbá, hogy a COVID-19 milyen tulajdonságainak köszönhetően tudott hónapok alatt helyi vírusos megbetegedésből pandémiává válni. Ezután az eddig emberekre áterjedt koronavírusokhoz hasonló tulajdonságain keresztül szemléltette, hogy hogyan tudott nagyságrendekkel nagyobb problémákat okozni a COVID-19, mint a SARS vagy a MERS. A koronavírusok és a COVID-19 legfontosabb tulajdonságainak ismertetése után az eredetére is kitért a dolgozat. Ezután a kínai Vuhan városából kiinduló járvány globális halálozási rátáján keresztül szemléltette az elemzés középponti problémáját, illetve a vírus terjedési ütemét is.

A következő szekcióban a járvány Olaszország-specifikus bemutatása következett. Az elején a COVID-19 megjelenése és a korlátozások országos bevezetése közötti idővonal segítségével szemléltettem a járvány súlyosságát, illetve a kormányzati reakciók nehézségeit. Az emögött rejlő indokok bemutatását követően ismertettem az olasz helyzet elfajulásának egyéb, általános okait. Ezek között beazonosítottam az elvégzett koronavírus tesztek számát, az egészségügy leterheltségét, a népesség egészségügyi problémáit és demográfiai különlegességeit, mint potenciális magyarázó erők. Olaszország demográfiai jellemzői és a halálozási aránya közötti kapcsolat bemutatásával zárult az irodalomösszefoglalás.

Az elemzés elsősorban az egyetemi Statisztika és Ökonometria kurzusokon elsajátított módszertan felhasználásával elemezte az olasz régiók közötti halálozás mögötti összefüggéseket. Az elemzés továbbá kiegészült a járványok terjedésével gyakran

szorosan együtt kezelt térbeli autokorreláció és autoregresszió vizsgálatával. A felhasznált szakirodalmak alapján elkészített alapmodell szűkítését először önálló módszerekkel, majd automatizált módszerekkel (stepwise szelekcióval) is elvégeztem. A kapott modellek alapján térbeli autoregresszív modelleket is építettem, majd 6 modell közül az Akaike információs kritérium alapján választottam ki a legjobbat. Ez a modell egy térbeli autoregresszív modell lett, mely a térbeli késleltetettségre kontrollálva javított az eredeti lineáris modellen.

Az elemzés egészségügyi, demográfiai és egyéb COVID-19-hez köthető változókat vizsgált. Az első kategóriából a krónikus betegségek száma lett szignifikáns, a másodikból a régiók népességeinek férfi-női arányai, a harmadikból pedig az összes koronavírus-tesztelt emberek száma. A kormányzati intézkedésekben megbúvó oksági vagy korrelációs kapcsolatokat releváns adat hiányában nem kezelte az elemzés. Végül a modell eredményeinek interpretálását az eredmények mögötti lehetséges okok feltárására tett kísérlet követte.

Hivatkozásjegyzék

- Andersen, K.G. - Rambaut, A. - Lipkin, W.I. et al (2020): The proximal origin of SARS-CoV-2, *Nature Medicine*, 26, pp. 450–452. letöltés helye: <https://doi.org/10.1038/s41591-020-0820-9>; letöltés ideje: 2020. április 30.
- Boccia, S. - Ricciardi, W. – Ioannidis, J. P. A. (2020): What Other Countries Can Learn From Italy During the COVID-19 Pandemic, *JAMA Internal Medicine*. letöltés helye: <https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/2764369>; letöltés ideje: 2020. április 21.
- Centers for Disease Control and Prevention (2020): Groups at Higher Risk for Severe Illness. letöltés helye: <https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/groups-at-higher-risk.html>; letöltés ideje: 2020. május 16.
- Coronavirus Today (2020a): Middle East Respiratory Syndrome (MERS). letöltés helye: <https://www.coronavirustoday.com/mers>; letöltés ideje: 2020. május 9.
- Coronavirus Today (2020b): Severe Acute Respiratory Syndrome (SARS). letöltés helye: <https://www.coronavirustoday.com/sars>; letöltés ideje: 2020. április 30.
- Hijmans, R. J. - Ghosh, A. (2019): Spatial Data Analysis with R. letöltés helye: <https://rspatial.org/analysis/analysis.pdf>; letöltés ideje: 2020. április 13.
- Horowitz, J. - Bubola E. - Povoledo, E. (2020): Italy, Pandemic's New Epicenter, Has Lessons for the World, *The New York Times*. letöltés helye: <https://www.nytimes.com/2020/03/21/world/europe/italy-coronavirus-center-lessons.html>; letöltés ideje: 2020. április 11.
- Istituto Superiore di Sanità (2020): Characteristics of SARS-CoV-2 patients dying in Italy Report based on available data on May 14th, 2020. letöltés helye: https://www.epicentro.iss.it/en/coronavirus/bollettino/Report-COVID-2019_14_may_2020.pdf; letöltés ideje: 2020. május 16.
- Jakab, F. - Kemenesi, G. (2020): A koronavírusról és a kialakult járványügyi helyzetről tárgyilagosan, *Magyar Tudományos Akadémia*. letöltés helye: https://mta.hu/tudomany_hirei/a-koronavirusrol-es-a-kialakult-jarvanyugyi-helyzetrol-targyilagosan-110308; letöltés ideje: 2020. április 30.

- Johns Hopkins University and Medicine (2020): Mortality Analyses. letöltés helye: <https://coronavirus.jhu.edu/data/mortality>; letöltés ideje: 2020. április 30.
- Keefe, T. E. (2020): The Coronavirus and the American Reality. letöltés helye: https://www.researchgate.net/publication/339898953_The_Coronavirus_and_the_American_Reality; letöltés ideje: 2020. április 25.
- Kovács, P. (2008): A multikollinearitás vizsgálata lineáris regressziós modellekben, Statisztikai Szemle, 86. évfolyam, 1. szám. letöltés helye: http://www.ksh.hu/statszemle_archive/2008/2008_01/2008_01_038.pdf; letöltés ideje: 2020. április 30.
- Központi Statisztikai Hivatal (2020): NUTS-rendszer, Területi atlasz – Európai Unió. letöltés helye: https://www.ksh.hu/teruletiatlasz_eu_nuts; letöltés ideje: április 30.
- Lin, Q. - Zhao, S. - Gao, D., et al (2020): A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action, International Journal of Infectious Diseases, Vol. 93, pp. 211-216. letöltés helye: [https://www.ijidonline.com/article/S1201-9712\(20\)30117-X/fulltext](https://www.ijidonline.com/article/S1201-9712(20)30117-X/fulltext); letöltés ideje: 2020. április 4.
- Magyarország Kormánya (2020a): Mit kell tudni a koronavírusról? letöltés helye: <https://koronavirus.gov.hu/mit-kell-tudni-a-koronavirusrol>; letöltés ideje: 2020. május 9.
- Magyarország Kormánya (2020b): Mit tegyünk tünetek esetén? letöltés helye: <https://koronavirus.gov.hu/mit-tegyunk-tunetek-eseten>; letöltés ideje: 2020. május 9.
- Mesel-Lemoine M. - Millet J. - Vidalain P.-O., et al (2012): A Human Coronavirus Responsible for the Common Cold Massively Kills Dendritic Cells but Not Monocytes, Journal of Virology. letöltés helye: <https://jvi.asm.org/content/86/14/7577.full>; letöltés ideje: 2020. április 26.
- Ministero della Salute (2020): Covid-19 - Situation in Italy. letöltés helye: <http://www.salute.gov.it/portale/nuovocoronavirus/dettaglioContenutiNuovoCoronavirus.jsp?lingua=italiano&id=5351&area=nuovoCoronavirus&menu=vuoto>; letöltés ideje: 2020. április 30.

- Nacoti, M. - Ciocca, A. - Giupponi, A., et al (2020): At the Epicenter of the Covid-19 Pandemic and Humanitarian Crises in Italy: Changing Perspectives on Preparation and Mitigation, NEJM Catalyst. letöltés helye: <https://catalyst.nejm.org/doi/full/10.1056/CAT.20.0080>; letöltés ideje: 2020. április 4.
- Nowotny, N. - Kolodziejek, J. (2014): Middle East respiratory syndrome coronavirus (MERS-CoV) in dromedary camels, Oman, 2013, Eurosurveillance, Vol. 19, No. 16. letöltés helye: <https://www.eurosurveillance.org/content/10.2807/1560-7917.ES2014.19.16.20781>; letöltés ideje: 2020. május 9.
- O'Sullivan, D. - Unwin, D. (2014): Geographic Information Analysis, 2nd Edition, John Wiley & Sons, Inc, Hoboken, New Jersey. letöltés helye: https://books.google.hu/books?hl=hu&lr=&id=yVUzBAAQBAJ&oi=fnd&pg=PR7&dq=Geographic+Information+Analysis,+2nd+Edition,+John+Wiley+%26+Sons,+Inc,+Hoboken,+New+Jersey&ots=kNkXA6HtGK&sig=Ss_yZw27feuCE-ZqCpACD-tdQ3E&redir_esc=y#v=onepage&q=Geographic%20Information%20Analysis%2C%202nd%20Edition%2C%20John%20Wiley%20%26%20Sons%2C%20Inc%2C%20Hoboken%2C%20New%20Jersey&f=false; letöltés ideje: 2020. április 20.
- Rezza, G. - Bella, A. - Riccardo F., et al. (2020): SARS-CoV-2 new coronavirus infection outbreak: the situation in Italy. letöltés helye: <https://www.epicentro.iss.it/coronavirus/sars-cov-2-italia>; letöltés ideje: 2020. április 18.
- Tobler, W. R. (1970): A computer movie simulating urban growth in the Detroit region, Economic Geography, Vol. 46, pp. 234–40. letöltés helye: <https://www.tandfonline.com/doi/pdf/10.2307/143141>; letöltés ideje: 2020. április 18.
- Verity, R. - Okell, L. C. - Dorigatti, I., et al. (2020): Estimates of the severity of coronavirus disease 2019: a model-based analysis, The Lancet Infectious Diseases. letöltés helye: [https://doi.org/10.1016/S1473-3099\(20\)30243-7](https://doi.org/10.1016/S1473-3099(20)30243-7); letöltés ideje: 2020. május 14.
- World Health Organization (2003): Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003. letöltés helye: https://www.who.int/csr/sars/country/table2004_04_21/en/; letöltés ideje: 2020. március 29.

World Health Organization (2020a): Alcohol and COVID-19: what you need to know. letöltés helye: http://www.euro.who.int/__data/assets/pdf_file/0010/437608/Alcohol-and-COVID-19-what-you-need-to-know.pdf; letöltés ideje: 2020. május 16.

World Health Organization (2020b): Coronavirus disease (COVID-19) Situation Report – 51. letöltés helye: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200311-sitrep-51-covid-19.pdf?sfvrsn=1ba62e57_10; letöltés ideje: 2020. május 9.

World Health Organization (2020c): Coronavirus disease (COVID-19) Situation Report – 102. letöltés helye: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200501-covid-19-sitrep.pdf?sfvrsn=742f4a18_4; letöltés ideje: 2020. május 9.

World Health Organization (2020d): WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. letöltés helye: <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>; letöltés ideje: 2020. március 20.

World Health Organization (2020e): WHO statement: Tobacco use and COVID-19. letöltés helye: <https://www.who.int/news-room/detail/11-05-2020-who-statement-tobacco-use-and-covid-19>; letöltés ideje: 2020. május 16.

Zhou, P. - Yang, X. - Wang, X., et al. (2020): A pneumonia outbreak associated with a new coronavirus of probable bat origin, Nature 579, pp. 270–273. letöltés helye: <https://doi.org/10.1038/s41586-020-2012-7>; letöltés ideje: 2020. május 5.

Adatbázisok hivatkozása

Eurostat (2020): NUTS,saját adatbázis, letöltés helye: <https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts>; letöltés ideje: 2020. május 5.

Istat (2020a): Economic conditions of families and inequalities, saját adatbázis, letöltés helye: <http://dati.istat.it/Index.aspx?QueryId=9919>; letöltés ideje: 2020. április 30.

Istat (2020b): Health and healthcare, saját adatbázis, letöltés helye: <http://dati.istat.it/Index.aspx?QueryId=9919>; letöltés ideje: 2020. április 30.

Istat (2020c): Population and families, saját adatbázis, letöltés helye:
<http://dati.istat.it/Index.aspx?QueryId=9919>; letöltés ideje: 2020. április 30.

Istat (2020d): Territory characteristics, saját adatbázis, letöltés helye:
<http://dati.istat.it/Index.aspx?QueryId=9919>; letöltés ideje: 2020. április 30.

Protezione Civile (2020): COVID-19 Italy - Situation monitoring, saját adatbázis, letöltés helye: <https://github.com/pcm-dpc/COVID-19>; letöltés ideje: 2020. május 1.