

Faktorska analiza COVID-19 epidemioloških mjera

Mario Mandušić

Sažetak

Pandemija COVID-19 bolesti izazvane novim koronavirusom SARS-CoV-2 vrlo snažno je uzdrmala cijeli svijet i izazvala velike probleme u zdravstvenom, društvenom i ekonomskom pogledu. Rezultati borbe protiv ove bolesti bitno su različiti prema državama i regijama, pa je vrlo teško detektirati pravilnosti u podacima koje bi pomogle pri zaključivanju koje su mjere učinkovite za suzbijanje ove ili slične respiratorne infekcije. Cilj ovog istraživanja je provođenje dimenzionalne redukcije 12 različitih epidemioloških mjera poduzetih u 179 različitih država u 2020. godini u svrhu otkrivanja potencijalnih latentnih varijabli u tom skupu mjera. Korištena je metoda faktorske analize (PCA, EFA, CFA). Rezultati faktorske analize pokazuju da među ispitanim epidemiološkim mjerama postoje latentne varijable, odnosno čimbenici koje su imenovane kao Factor1 = 'Lockdown', Factor2 = 'Azijska Strategija', Factor3 = 'Rani Odgovor'. U daljnjim istraživanjima potrebno je testirati hipotezu da li otkrivena struktura epidemioloških mjera ujedno znači da su iste grupirane po kronološkom redosljedu primjene i da njihov redosljed i način primjene ujedno nudi i optimalnu efikasnost u borbi protiv COVID-19 infekcije. S obzirom na grupiranje mjera u navedene latentne varijable, postavlja se pitanje jesu li neke zemlje značajno uspješnije u suzbijanju bolesti zbog toga što su odlučno i na vrijeme poduzele Factor3 ('Rani Odgovor'), zatim Factor2 ('Azijska strategija') kako bi se izbjegao Factor1 ('Lockdown'), do kojeg možda dolazi ponajviše zbog lošeg provođenja ili neprovođenja Factora3 i Factora2.

KLJUČNE RIJEČI: COVID-19 mjere, faktorska analiza, PCA, Lockdown, latentne varijable, Azijska strategija.

Uvod

Problem efikasnosti mjera za suzbijanje COVID-19 infekcije česta je tema javnih rasprava ne samo u stručnim, nego i u širim društvenim krugovima. Do dana dovršetka ovog rada u svijetu je zabilježeno više od 72,6 milijuna zaraženih osoba, a više od 1,6 milijuna ljudi je preminulo kao posljedica ove bolesti. Prema procjenama MMF-a, pad bruto društvenog proizvoda (BDP-a) za zemlje Europske unije (EU27) u 2020. godini iznosit će oko 7 %, što je najveći pad od Drugog svjetskog rata. *Invalid source specified.*

Prema autorovim saznanjima, do izrade ovog rada faktorska analiza nije se koristila za otkrivanje latentnih varijabli među nefarmaceutskim epidemiološkim mjerama. Studija kolaboracije Cochrane prikazala je kako rana primjena karantene i fizičkog distanciranja može pomoći u usporavanju širenja COVID-19 *Invalid source specified.* Međutim, određivanje kombinacije mjera koja nudi najbolje rezultate za reduciranje broja slučajeva i smrti od navedene bolesti pokazalo se jako teškim.

Nils Haug i suradnici *Invalid source specified.* proučavali su utjecaj različitih nefarmaceutskih intervencija (NPI) na širenje virusa SARS-CoV-2. Rezultati su pokazali da najveći utjecaj na smanjenje reproduktivnog čimbenika širenja virusa (R_t) imaju zabrana okupljanja većeg broja ljudi, zatvaranje edukacijskih institucija, te restrikcije na državnim granicama. Lahiri i suradnici (*Arista Lahiri, 2020*) prikazali su da je matematički racionalno implementirati mjere socijalnog distanciranja, pa i lockdown, ali studija istodobno nudi uvid u važnost primjene mjera kao što su povećanje broja testiranja i povećanje broja jedinica intenzivne njege (ICU). Ylidorim i Guler *Invalid source specified.*, koristili su faktorsku

analizu percepcije rizika s aspekta obolijevanja od COVID-19, dok Luo i suradnici *Invalid source specified*. pomoću faktorske analize obrađuju problem grupiranja simptoma bolesti. Bernardo i suradnici *Invalid source specified*. koriste eksploratornu faktorsku analizu (EFA) u svrhu istraživanja tjeskobe vezane za pandemiju koronavirusa.

Cilj ovog rada je analizirati 12 različitih epidemioloških mjera koje je poduzelo 179 država, s namjerom otkrivanja potencijalne skrivene strukture unutar tog skupa mjera. Navedeno bi bila osnova za daljnja istraživanja koja bi, korištenjem statističkih metoda, mogla utvrditi koje su mjere zaista efikasne za suzbijanje pandemije. Bez pokušaja postavljanja izravne hipoteze koje su epidemiološke mjere efikasne za suzbijanje bolesti, u ovoj studiji, u tri različita koraka provest će se faktorska analiza (PCA, EFA, CFA).

Metode

U ovom znanstvenom radu glavna metoda istraživanja je faktorska analiza *Invalid source specified*. Faktorska analiza je skup statističkih metoda koje se koriste za opisivanje varijabilnosti između varijabli koje su međusobno značajno korelirane. Cilj navedene metode je izvršiti dimenzionalnu redukciju, kojom se u skupu od više varijabli traže potencijalne neovisne i neopservirane latentne varijable. Pretpostavka za korištenje ove metode je da su opservirane varijable modelirane kao linearne kombinacije potencijalnih faktora. Faktorska analiza posebno je zahvalna za skupove podataka s relativno puno opserviranih varijabli. U tim slučajevima ponekad se događa da klasične metode statističke obrade, poput višestruke linearne regresije ili višestruke logističke regresije, za potrebe klasifikacije ne nude jasne uvide koje su varijable statistički značajne.

Primjenom faktorske analize u spomenutim slučajevima moguće je dobiti dublji uvid u potencijalnu skrivenu strukturu većeg skupa varijabli, u do tada neopservirane latentne varijable, koje onda mogu služiti za daljnju analizu i primjene drugih statističkih metoda.

U ovom radu, faktorska analiza podijeljena je u tri koraka, odnosno u komponentnu analizu (Principal Component Analysis – PCA) *Invalid source specified*., eksploratornu faktorsku analizu (Exploratory Factor Analysis - EFA) *Invalid source specified*. i konfirmatornu faktorsku analizu (Confirmatory Factor Analysis - CFA) *Invalid source specified*.

Komponentna analiza (PCA) je metoda kojom se određuje broj glavnih komponenti skupa varijabli. Navedeno znači da komponentna analiza pomaže kod odabira broja faktora i da se koristi u eksploratornoj faktorskoj analizi za prediktivno modeliranje, ili u ovom slučaju za dimenzionalnu redukciju. U osnovi, glavne komponente u komponentnoj analizi su svojstveni vektori (eigenvectors) matrice kovarijance opservacija iz zadanog skupa.

Intuitivno, komponentna analiza se može zamisliti kao optimiziranje geometrijske pozicije p-dimenzionog elipsoida prema podacima, pri čemu svaka os elipsoida predstavlja po jednu glavnu komponentu. Potrebno je napomenuti da je komponentna analiza osjetljiva na skaliranje varijabli, što znači da neće dati dovoljno precizne rezultate, ako su različite varijable iskazane bitno različitim

mjernim jedinicama. Nakon što se pomoću komponentne analize odredi broj glavnih komponenti, odnosno faktora, potrebno je pristupiti eksploratornoj faktorskoj analizi (EFA). Eksploratorna faktorska analiza služi prije svega za otkrivanje skrivene strukture relativno velikog broja varijabli u nekom skupu podataka.

Eksploratorna faktorska analiza se primjenjuje na rezultate komponentne analize s ciljem utvrđivanja značajnih koeficijenata korelacije faktora i varijable (factor loadinga) utvrđene u komponentnoj analizi. Česta tehnika korištena u eksploratornoj faktorskoj analizi je i faktorska rotacija kroz višedimenzionalni prostor kako bi se došlo do što jednostavnijeg rješenja. Najčešće korištena je ortogonalna 'varimax' rotacija. Takva rotacija zahtijeva da faktori međusobno nemaju visoku korelaciju. Interpretacija dobivenih rezultata ovisi o tome koliko snažno faktor utječe na postojeću varijablu.

Po završetku eksploratorne faktorske analize uputno je napraviti i konfirmatornu faktorsku analizu. Konfirmatorna faktorska analiza je vrsta strukturalnog modeliranja koja se koristi kada se provjerava odnos između mjerljivih varijabli (opservacija) i latentnih varijabli (faktora). Navedena metoda se koristi kod provjere pouzdanosti ljestvice i doprinosi valjanosti teorijskih konstrukata koji su u pozadini razvijenog modela. Treba napomenuti da konfirmatorna faktorska analiza ne ispituje valjanost modela, nego samo koliko dobro se razvijeni model poklapa sa podacima.

U ovoj vrsti analize provodi se više vrsta statističkih testova, a standardno se izvještava o vrijednostima Hi-kvadrat testa, RMSEA (Root mean square error of approximation) CFI (Comparative fit indeks) i SRMSR (Standardized root mean square residual).

Izvor za ovu studiju su podaci koje je objavilo Sveučilište Oxford – Oxford Covid19 Government Response Tracker (OxCRGT). Na navedenom izvoru prezentiran je skup podataka s različitim mjerama usmjerenim na suzbijanje posljedica COVID-19 infekcije za 182 različite države u svijetu. Izvor se povremeno ažurira, a u ovom radu obrađeni su podaci iz izvora na dan pristupa 22. studenog 2020. godine *Invalid source specified.*

Iz navedenog skupa varijabli odabrane su ne-farmaceutske mjere, a zanemarene ekonomske i druge mjere koje su neprimjenjive u ovakvoj analizi. Varijable su iskazane u kategoričkim jedinicama, a mjerenja su izvršena kronološki, po datumima, tako da za svaki datum postoji odgovarajuće mjerenje. Vremenska dimenzija je eliminirana iz analize na način da je za svaku zemlju uzet prosjek vrijednosti mjera za cijelo promatrano razdoblje. Isto tako su zanemarene 3 države za koje nisu bili dostupni numerički podaci o poduzetim mjerama.

Priprema skupa podataka izvršena je u MS Excel (2016), a statistička obrada, kreiranje i testiranje modela izvršena je u programskom jeziku R (4.0.0), s grafičkim okružjem RStudio (1.3.091.).

U obradu je uzeto 12 nefarmaceutskih epidemioloških mjera:

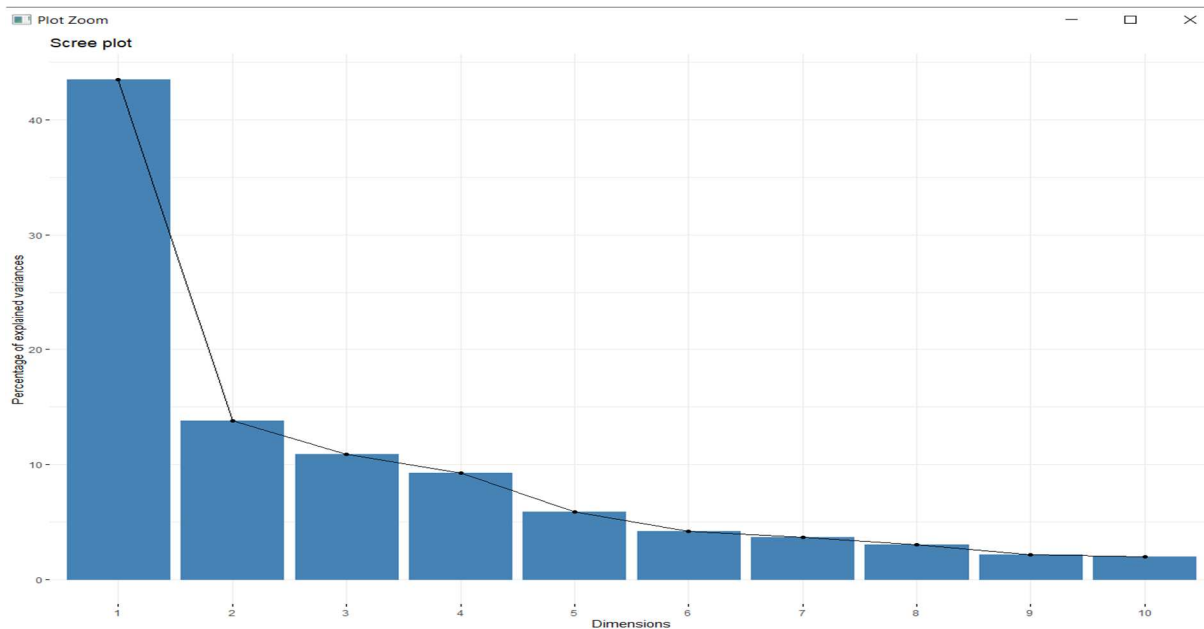
1. Zatvaranje škola
2. Zatvaranje radnih mjesta
3. Zatvaranje javnih događaja
4. Zabrana javnih okupljanja

5. Zatvaranje javnog transporta
6. Ostanak kod kuće
7. Restrikcije unutarnjeg kretanja
8. Kontrola međunarodnih putovanja
9. Testiranje
10. Informiranje javnosti
11. Praćenje kontakata
12. Nošenje maski

Rezultati

Standardni statistički testovi utvrđivanja pogodnosti nekog skupa podataka za primjenu faktorske analize su Kaiser-Meier-Olkin test (KMO test) i Bartlett test. KMO test pokazao je vrijednost 0,87, što znači da je faktorska analiza vrlo pogodna za ovaj skup podataka. P-vrijednost Bartlett testa na sfernost je $2,2 \times 10^{-16}$, što je isto tako vrlo povoljan indikator za provođenje faktorske analize.

Provedena je komponentna analiza da bi se utvrdio broj latentnih varijabli/faktora, pri čemu treba odrediti kriterije za prihvaćanje broja faktora. U ovom slučaju odabran je broj komponenti koje individualno objašnjavaju više od 10 % varijance u podacima, a kumulativno objašnjavaju 68,11 % varijance u podacima, kao što prikazuje Slika 1.



Slika 1. PCA scree plot

Nakon toga provedena je faktorska analiza na tri komponente uz varimax rotaciju. Nakon ortogonalne varimax rotacije dobivene su vrijednosti koeficijenata korelacije faktora i varijabli (factor loadings) koje su prikazane u Tablici 1.

Tablica 1. Koeficijenti korelacije faktora i varijabli (EFA factor loadings)

Epidemiološka mjera	Factor1	Factor2	Factor3
Zatvaranje škola	0.687	0.156	0.190
Zatvaranje radnih mjesta	0.787	0.117	
Zatvaranje javnih događaja	0.767	0.274	
Zabrana okupljanja	0.751	0.376	-0.109
Zatvaranje javnog prijevoza	0.735	-0.141	0.116
Ostanak kod kuće	0.818		0.145
Restrikcije unutarnjeg kretanja	0.768	-0.216	0.382
Kontrola međunarodnih putovanja	0.153	0.231	0.537
Testiranje		0.563	
Informiranje javnosti	0.214	0.666	0.293
Praćenje kontakata		0.652	
Korištenje zaštitinih maski za lice	0.382	0.224	0.115

Kao kriterij za grupiranje uzete su vrijednosti koeficijenata korelacije između faktora i varijabli > 0.5 , te su latentne varijable imenovane (Tablica 2.).

Tablica 2. Latentne varijable nakon faktorske analize

Faktor	COVID-19 mjere	Latentna varijabla
Factor1	Zatvaranje škola	"Lockdown"
	Zatvaranje radnih mjesta	
	Zatvaranje radnih mjesta	
	Zabrana okupljanja	
	Zatvaranje javnog prijevoza	
	Ostanak kod kuće	
	Restrikcije unutarnjeg kretanja	
Factor2	Testiranje	"Azijska Strategija"
	Informiranje javnosti	
	Praćenje kontakata	
Factor3	Kontrola međunarodnih putovanja	"Rani Odgovor"

Nakon komponentne analize i eksploratorne faktorske analize, primjereno je testirati koliko model odgovara podacima kroz konfirmatornu faktorsku analizu. Rezultate statističkih testova za ocjenu konfirmatorne faktorske prikazuje Tablica 3.

Tablica 3. Rezultati statističkih testova konfirmatorne faktorske analize

Statistički test	Vrijednosti iz analize	Referentne vrijednosti
Hi-kvadrat test (p-vrijednost)	0.00	<0.05
CFI	0.88	0.9
RMSEA (p-vrijednost)	0.00	<0.05
SRMR	0.09	<0.08

Prema naznačenim parametrima zaključuje se da postavljeni model dobro odgovara podacima.

Rasprava i zaključak

Početni skup podataka je dinamičan i redovno se ažurira, a za potrebe ove analize iskorišten je skup podataka kojem je pristupljeno na dan 22. studenog 2020. Moguće je da se primjenom istih metoda na ažurirani skup podataka u budućim razdobljima dobiju nešto drugačiji rezultati. Međutim, autor je mišljenja da ti rezultati neće biti bitno različiti od rezultata prezentiranih u ovom istraživanju, odnosno da će se varijable i dalje snažno grupirati na barem dva faktora.

Iz početnog skupa odabrano je 12 ne-farmaceutskih metoda (NPI), a iz analize su uklonjene varijable koje se odnose na ekonomske mjere, te indekse strogosti mjera, s obzirom da su se isti ranije pokazali kao statistički neznčajni. Skup obuhvaća opservacije 183 različite zemlje u kojima su za svaki datum od početka epidemije iskazane kategoričke vrijednosti mjera. U analizi su zanemarene 4 države za koje podaci nisu bili dostupni, pa je konačni broj opservacija 179. Za svaku državu uzet je prosjek mjera po datumima, da se iz analize eliminiira vremenska dimenzija. Skup podataka pokazao se pogodnim za faktorsku analizu, prema vrijednostima KMO i Bartlett testa.

Komponentna analiza nije ponudila dovoljno jasne kriterije za odabir broja komponenti/faktora, s obzirom da postoji samo jednu komponentu koja ima svojestvene vrijednosti (eigenvalue) veći od 1, te se ne nazire dominantna točka infleksije na scree plotu. Međutim, promatrajući p-vrijednost za prihvaćanje hipoteze o tome koliko je faktora dovoljno za obuhvaćanje statističkog značaja, dolazi se do zaključka da je potrebno odabrati tri ili četiri komponente. Odabrane su tri komponente s obzirom na p-vrijednost od 0,4, te činjenicu da iste objašnjavaju individualno više od 10 %, a kumulativno 68,11 % varijance u podacima.

Eksploratorna faktorska analiza nakon ortogonalne varimax rotacije pokazuje vrijednosti koeficijenata korelacije za različite mjere. Kao kriterij za grupiranje odabrane su vrijednosti pojedinih koeficijenata korelacije (factor loadings) koji su veći od 0,5.

Factor1 obuhvaća zatvaranje škola, zatvaranje radnih mjesta, zatvaranje javnih događaja, zabranu okupljanja, zatvaranje javnog prijevoza, ostanak kod kuće, te ograničavanje unutarnjeg kretanja. Ova

latentna varijabla imenovana je kao 'Lockdown', s obzirom da reflektira skup mjera koji se inače podrazumijevaju pod tim imenom. Iz podataka je vidljivo da je ovaj oblik borbe protiv bolesti dominantan.

Factor2 obuhvaća testiranje, informiranje javnosti i praćenje kontakata. Kako je ova strategija posebno dobro razvijena u azijskim zemljama (Tajvan, Singapur, Južna Koreja, Japan, Vijetnam i Kina) ovaj faktor je imenovan kao 'Azijska strategija'.

Napominje se kako koeficijent korelacije između faktora i varijable (factor loading) nije bio dovoljno značajan da se mjera nošenja maski uvrsti u 'Azijsku Strategiju', Razlog tome vjerojatno leži u činjenici da je politika nošenja maski općenito u svijetu bila vrlo nedosljedna, što vjerojatno stvara šum u podacima.

Factor3 pokazuje koeficijent korelacije (factor loading) veći od 0,5 samo za mjeru kontrole međunarodnih putovanja. Kako je upravo to najčešće prva mjera kod suzbijanja širenja zaraze, ova latentna varijabla nazvana je 'Rani Odgovor'. CFA pokazuje primjerene vrijednosti statističkih testova, koje pokazuju da razvijeni model dobro odgovara podacima.

Rezultati ove studije postavljaju nova pitanja i nude nove hipoteze za testiranje u daljnjim istraživanjima. Potrebno je ispitati da li latentne varijable koje su ovdje detektirane imaju značaj kod vremenske kronologije primjene i same efikasnosti mjera. Kod pojave zarazne bolesti prvo se poduzimaju kontrole međunarodnih putovanja (Factor3), kada se pojave prve zaražene osobe, vrlo je važno dobro pratiti kontakte oboljelih i testirati (Factor2), a u slučaju kada postoji snažna lokalna transmisija koja prijeti urušavanjem zdravstvenog sustava i humanitarnom katastrofom, tada se u pravilu poseže za mjerama ograničavanja i zabrane kretanja (Factor1).

Dominaciju 'Lockdowna' među navedenim latentnim varijablama, osim navedenog treba tražiti i u činjenici da se na početku pandemije nije znalo puno o prirodi bolesti, pa je najsigurnija varijanta upravo bila Lockdown, čak i u slučajevima kada nije bilo značajne lokalne transmisije. Dobro je poznato da su azijske države (Tajvan, Singapur, Južna Koreja, Japan, Vijetnam i Kina) najmanje pogođene pandemijom i u javno-zdravstvenom i u ekonomskom smislu. Postavlja se pitanje da li je značajan parametar njihovog uspjeha upravo pravodobno i kvalitetno primijenjen Factor3 (rana kontrola međunarodnih putovanja) i Factor2 (politika testiranja i praćenja kontakata, te informiranja i educiranja javnosti), s obzirom da ove države (osim Kine) u borbi protiv bolesti nisu značajno koristile skup mjera koje prepoznajemo kao Lockdown.

Dojam je da su Europa, SAD i Južna Amerika nepravovremeno i nedovoljno kvalitetno pripremile i realizirale skupove mjera koje prepoznajemo kao 'Rani Odgovor' i 'Azijsku Strategiju'. Značajne varijable kod objašnjenja različitosti rezultata u borbi protiv bolesti između Istoka i Zapada mogu biti i kultura, te iskustvo u borbi protiv zarazenih bolesti, osobito prethodnih epidemija respiratornih infekcija u zadnja dva desetljeća (SARS, MERS), pa je i to još jedna hipoteza za buduća istraživanja.

Temeljem rezultata ove studije, u novim istraživanjima potrebno je potražiti i ispitati dublje razloge postojanja latentnih varijabli ovakve strukture.

Literatura

1. Allan B. I. Bernardo, c. a.-A. (n.d.). Coronavirus Pandemic Anxiety Scale (CPAS-11): development and initial validation. Dohvaćeno iz <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7664585/>
2. Arista Lahiri, S. S. (2020). Effectiveness of preventive measures against COVID-19: A systematic review of In Silico modeling studies in indian context. Pubmed. Dohvaćeno iz <https://pubmed.ncbi.nlm.nih.gov/32496248/>
3. Brown, M. W. (Springer). Introduction to Confirmatory Factor Analysis and Structural Equation Modeling. Springer. Dohvaćeno iz https://link.springer.com/chapter/10.1007/978-94-6209-404-8_14
4. Cadima, I. T. (2016). Principal component analysis: a review and recent developments. The Royal Society Publishing. Dohvaćeno iz <https://royalsocietypublishing.org/doi/10.1098/rsta.2015.0202>
5. Cochrane. (2020). Updated Cochrane Rapid Review assesses the effectiveness of quarantine during the COVID-19 pandemic. Cochrane. Retrieved from <https://www.cochrane.org/news/updated-cochrane-rapid-review-assesses-effectiveness-quarantine-during-covid-19-pandemic>
6. IMF. (2020). Regional Economic Outlook. IMF. Dohvaćeno iz <https://www.imf.org/en/Publications/REO/EU/Issues/2020/10/19/REO-EUR-1021>
7. Nils Haug, L. G.-L. (2020). Ranking the effectiveness of worldwide COVID-19 government interventions. Nature Human Behaviour. Dohvaćeno iz <https://www.nature.com/articles/s41562-020-01009-0>
8. OxCGRT. (2020.). Oxford: Oxford. Dohvaćeno iz <https://github.com/OxCGRT/covid-policy-tracker>
9. Pearce, A. G. (n.d.). A Beginner's Guide to Factor Analysis:. Dohvaćeno iz <http://www.tqmp.org/RegularArticles/vol09-2/p079/p079.pdf>
10. Watkins, M. W. (n.d.). Exploratory Factor Analysis: A Guide to Best Practice. Sage Journals. Dohvaćeno iz <https://journals.sagepub.com/doi/abs/10.1177/0095798418771807>
11. Yildirim, G. (n.d.). Factor analysis of the COVID-19 Perceived Risk Scale: A preliminary study. Dohvaćeno iz <https://www.tandfonline.com/doi/full/10.1080/07481187.2020.1784311>
12. Yueming Luo, J. W. (n.d.). Investigation of COVID-19-related symptoms based on factor analysis. Pubmed. Dohvaćeno iz <https://pubmed.ncbi.nlm.nih.gov/32576016/>

Abstract

The pandemic of COVID-19 disease caused by the novel coronavirus SARS-CoV-2 has shaken the whole world very vigorously and caused significant problems in health, social and economic terms. The fight against the disease results are significantly different in countries and regions, so it is challenging to detect regularities in the data that would help in concluding which measures are useful in combating this or a similar respiratory infection. This paper aimed to carry out a dimensional reduction of 12 different epidemiological measures taken in 179 different countries in 2020 to identify potential underlying latent variables. Factor analysis method (PCA, EFA, CFA) was used. The factor analysis results show that there are underlying latent variables (factors) named Factor1 = 'Lockdown', Factor2 = 'Asian Strategy,' among the examined epidemiological measures Factor3 = 'Early Response.' In further research, it is necessary to test whether the detected structure of epidemiological measures also means that they are grouped in chronological order of application and that their order and method of application also offer optimal efficacy in combating COVID-19 infection. The question arises as to whether some countries are significantly more successful in controlling the disease because they have resolutely and timely taken Factor3 ('Early Response'), then Factor2 ('Asian Strategy') to avoid Factor1 ('Lockdown'), which may occur mostly due to poor implementation or non-implementation of Factor3 and Factor2.

KEYWORDS: COVID-19 measures, factor analysis, PCA; Lockdown, Latent variables, Asian Strategy