



ISPITIVANJE PRECIZNOSTI OCENA PARAMETARA POPULACIJE PRIMENOM PROSTOG I STRATIFIKOVANOG SLUČAJNOG UZORKA

EXAMINATION OF PRECISION EVOLUTION OF POPULATION PARAMETER USING A SIMPLE AND STRATIFIED RANDOM SAMPLE

Nemanja Budimir^{3,1}

³ Doktorand Ekonomskog fakulteta Univerziteta u Beogradu, Fondacija "Andrej Budimir", Bosna i Hercegovina i Bryggan Motala kommun, Švedska

PODACI O ČLANKU

Primljen 27.02.2021.
Dostavljen na recenziju 25.08.2021.
Prihvaćen 17.09.2021.
Dostupan online od 05.11.2021.

Ključne riječi:

populacija, prost slučajni uzorak, stratifikacija, stratifikovani uzorak, SJR pokazatelj

ARTICLE INFO

Received 02/27/2021
Sent to revision 08/25/2021
Accepted 09/17/2021
Available online 11/05/2021

Keywords:

systemic risk
financial system
systemic risk indicators
regulatory framework

APSTRAKT

Da bi se istraživanja mogla sprovesti veoma važno je da se izabere odgovarajući uzorak. Naime, važno je da se izabere uzorak koji će prikazivati osnovne karakteristike populacije. Izbor jednostavnog slučajnog uzorka predstavlja postupak ili plan izbora u kojem svaki uzorak veličine n ima istu verovatnoću izbora. Ovaj oblik izbora uzorka praktično se izvodi tako što se od ukupnog broja od N jedinica populacije koje se nalaze na spisku ili listi članova populacije u svakom koraku bira po jedan član u uzorku, ali svakom članu populacije daje se jednaka šansa da bude izabran u uzorak. S druge strane, formiranje stratifikovanog uzorka uobičajeno je pri uzimanju uzoraka iz veoma heterogenih populacija, odnosno populacija koje se prema nekom obeležju mogu podeliti na stratume, slojeve ili subpopulacije. Stratifikovani uzorak predstavlja jedan od najčešće upotrebljivanih planova uzorka. U ovom radu ispitivana je preciznost ocene prostog i stratifikovanog slučajnog uzorka na tržištu naučnih časopisa iz Bosne i Hercegovine, Hrvatske i Srbije. Kao obeležje odabran je SJR pokazatelj (Scimago Journal & Country Rank). SJR pokazatelj primenjuje se za procenu i analizu časopisa iz domena nauke. Sprovedene analize pokazale su da bolje ocene daje stratifikovani uzorak, u odnosu na jednostavan slučajni uzorak. Iz ovoga sledi da stratifikovani uzorak ima manju varijansu i njegova sredina je bliža sredini osnovnog skupa. Može se zaključiti da ovaj uzorak ima manja odstupanja od parametara osnovnog skupa.

ABSTRACT

In order for research to be conducted, it is very important to choose the appropriate sample. Namely, it is important to choose a sample that will show the basic characteristics of the population. Simple random sampling is a sampling procedure or plan in which each sample of size n has the same probability of selection. This form of sampling is practically performed by selecting one member in the sample from the total number of N population units on the list or list of population members in each step, but each population member is given an equal chance to be sampled. On the other hand, stratified sampling is common when taking samples from very heterogeneous populations, i.e. populations that can be divided into strata, layers or subpopulations according to some characteristics. The stratified sample is one of the most commonly used sample plans. This paper examines the accuracy of estimating a simple and stratified random sample on the market of domestic scientific journals from Bosnia and Herzegovina, Croatia and Serbia. The SJR indicator (Scimago Journal & Country Rank) was chosen as a corresponding feature value. The SJR indicator is used for the evaluation and analysis of journals in the field of science. The conducted analyzes showed that the stratified sample provides more precise evaluation than the simple random sample. It follows that the stratified sample has less variance and its middle average is closer to the middle average of the population. It can be concluded that this sample has smaller deviations from the parameters of population.

¹ nemanja.fondacijaandrebudimir@gmail.com

Uvod

U različitim istraživanjima javlja se potreba za ispitivanjem određenih karakteristika elemenata nekog skupa. Nemogućnost pristupa svakom elementu, materijalna ili vremenska ograničenja predstavljaju osnovne razloge zbog kojih su istraživači primorani da tražene karakteristike ispituju na nekom podskupu početnog skupa. Taj početni skup naziva se populacija, a njegovi elementi jedinice i njih obeležavamo sa N (broj jedinica u populaciji). (Rakić, 2010: 68) Dakle, populacija predstavlja celokupnu kolekciju objekata u kojoj se može vršiti ispitivanje nekog obeležja, odnosno karakteristike. Kada se govori o populaciji potrebno je naglasiti da ako je ona mala onda se ona može izučavati u celini. S druge strane, kada je reč o populaciji koja sadrži veliki broj elemenata, ispitivanje cele populacije je skupo, dugotrajno, a u pojedinim slučajevima ono može da bude i destruktivno, a nekada je i nemoguće. Upravo zbog toga se uglavnom bira podskup populacije (sa n se označava broj jedinica u uzorku), koji se naziva uzorkom i na njemu se vrši ispitivanje. Na osnovu analiziranog uzorka nastoji se izvesti zaključak o celoj populaciji. Deo statistike koji se bavi proučavanjem izbora uzorka i ocenjivanjem odgovarajućih parametara populacije naziva se teorija uzorka. Iz ovoga se može zaključiti da postoji veliki broj tehnika i načina odabira uzorka, a samim time i razne vrste uzorka (Rapajić, 2010).

U radu je prikazana analiza dva plana uzorka, a to su prost slučajni uzorak bez ponavljanja i stratifikovan slučajni uzorak. Analiza je sprovedena tako što su posmatrana tržišta naučnih časopisa iz Bosne i Hercegovine, Srbije i Hrvatske, a koje se nalaze na SJR (Scimago Journal & Country Rank) za 2019. godinu. SJR pokazatelj u radu predstavlja vrednost obeležja.

Cilj rada je da se utvrdi u kom uzorku izračunata sredina je bliža sredini populacije. Primenom kog plana uzorka je izračunata varijansa sredine uzorka manja, a samim tim će se pokazati i gde se beleži manje odstupanje od parametara populacije.

Osnovna hipoteza rada glasi: Pomoću stratifikovanog slučajnog uzorka obezbeđuju se preciznije ocene, u odnosu na jednostavni (prost) slučajni uzorak bez ponavljanja.

Pojam i vrste uzorka

Populacija predstavlja skup svih elemenata na kojima se neka statistička pojava posmatra. Skup može da bude beskonačan ili konačan. U praksi ispitivanje obeležja cele populacije retko se realizuje. Ona bi mogla da bude složena, neracionalna, a ponekad i nemoguća jer populacija može da ima beskonačno mnogo elemenata ili konačno mnogo, ali da je taj broj veliki. Upravo iz tih razloga vrši se definisanje nekog podskupa uočene populacije i on se naziva uzorak. Uzorak predstavlja podskup skupa podataka, tu je reč o izdvojenom delu osnovnog skupa za potrebe statističke analize. Analiza uzorka predstavlja osnovu za donošenje zaključaka o osobinama osnovnog skupa, upravo zbog toga je od velike važnosti da uzorak bude reprezentativan. Za uzorak se može reći da je reprezentativan samo onda kada svojim osobinama prikazuje osobine osnovnog skupa iz kojeg je izdvojen (Marjanović i dr., 2016). Formiranje uzorka predstavlja složen postupak koji mora da zadovolji niz praktičnih i teorijskih kriterijuma. Dakle, svaki uzorak mora da bude: reprezentativan (svaki element populacije ima podjednaku šansu da se nađe u uzorku), dovoljno brojan (treba da sadrži optimalan broj elemenata na osnovu kojih se donosi zaključak), objektivno (potrebno je odbaciti sve subjektivne faktore, a to se postiže slučajnim izborom). Uzorak predstavlja umanjenu sliku osnovnog skupa, a njegove karakteristike, parametri kao što su aritmetička sredina, disperzija, standardno odstupanje i sl. predstavljaju procene istih parametara osnovnog skupa. S druge strane, raspodela frekvencija statističkog obeležja uzorka predstavlja aproksimaciju raspodele odgovarajuće slučajne promenljive u celoj populaciji (Kostić Kovačević, 2019, p. 148-149).

Kada se vrši izbor podataka iz osnovnog skupa primenjuju se metode izbora uzorka. Prilikom izbora podataka iz osnovnog skupa, oni mogu da budu izabrani nasumično ili metodom namernog izbora. Slučajni uzorci nazivaju se još i uzorci verovatnoće. Kod slučajnog izbora podataka postoji jednaka verovatnoća da sve jedinice budu izabrane u uzorak. Upravo na ovaj način zadovoljava se kriterijum reprezentativnosti uzorka. Dakle, od velike važnosti je obezbediti uzorak koji predstavlja "umanjenu", a nikako "iskrivljenu", niti "uveličanu" sliku jednog dela populacije (Jevremović, 2019, p. 196) Takođe, jedinice osnovnog skupa biraju se nezavisno jedna od druge. Ovakvi uzorci imaju određenu prednost, a ona se ogleda u tome što slučajni izbor podataka iz osnovne populacije smanjuje pristrasnost. Njihov nedostatak je u tome što su skupi i zahtevaju više vremena za primenu (Marjanović i dr., 2016). Osnovni skup iz kojeg se vrši izbor podataka na slučajni način može da bude konačan i beskonačan. U slučaju beskonačnih skupova broj

elemenata skupa nije moguće precizno odrediti, s druge strane konačni skupovi sastoje se od ograničenog broja elemenata.

Za razliku od slučajnih uzoraka, namerni uzorci biraju se prema ličnom nahođenju ispitivača. Istraživač navodi brojeve, odnosno kvote koji pružaju mogućnost izbora onih jedinica skupa, koje prema njegovom mišljenju na adekvatan način prikazuju skup. Neki od namernih uzoraka su: uzorci koji su formirani na osnovu subjektivnog suda, kvota uzorci, pogodni uzorci (Đorđević, 2006, p. 7).

Prema teoriji verovatnoće razlikuju se dva tipa uzoraka, a to su: mali uzorci ($n \leq 30$) i veliki uzorci ($n > 30$) (Petrović, 2018). Potrebno je naglasiti da prilikom ocenjivanja parametara osnovnog skupa na bazi malih uzoraka upotrebljava se Studentova t raspodela verovatnoće, dok se za velike uzorke t raspodela aproksimira sa normalnom (Kostić Kovačević, 2019). Dakle, stepen pouzdanosti ocenjivanja zavisi od vrste odabranog uzorka. Razlikuje se nekoliko vrsta uzoraka, a to su: prost slučajni uzorak, sistemski uzorak, sekvenčialni uzorak, stratifikovani uzorak, višefazni uzorak, višestepni uzorak, panel uzorak i procesni uzorak (Marjanović i dr., 2016, p. 21).

Podaci iz populacije

U radu će biti posmatrana populacija, odnosno skup naučnih časopisa iz Bosne i Hercegovine, Srbije i Hrvatske, a koji se nalaze rangirani na listi SJR (Scimago Journal & Country Rank). Prema podacima iz 2019. godine na SJR listi nalazi se 244 časopisa, od čega je 11 časopisa iz Bosne i Hercegovine, 163 iz Hrvatske i 70 iz Srbije. Podaci koji su upotrebljavani preuzeti su sa zvaničnog sajta SJR (Scimago Journal & Country Rank). Časopisi na sajtu rangirani su prema SJR vrednosti. SJR pokazatelj primenjuje se za procenu i analizu časopisa iz domena nauke. Ovaj pokazatelj zasnovan je na prenosu prestiža sa jednog časopisa na drugi. Prestiž se prenosi kroz citate i na taj način prestiž jednog časopisa ima uticaj na ostale i obrnuto. SJR pokazatelj je upotrebljen kao vrednost obeležja. S obzirom da postoje vrednosti obeležja za sve elemente populacije (skupa), moguće je izračunati parametre populacije, tj. moguće je izračunati sredinu i varijanse, a koje su putem uzorka ocenjivane, a zatim su dobijene ocene upoređivane.

U tabeli koja sledi prikazani su rezultati koji karakterišu populaciju.

Tabela 1: Parametri populacije

Parametar	Oznaka	Vrednost
Total	Y	53,085
Sredina	\bar{Y}	0,218
Varijansa	S^2	0,019
Maksimalna vrednost obeležja populacije		0,824
Minimalna vrednost obeležja populacije		0,100

Izvor: Proračun autora.

Na osnovu dobijenih rezultat može se zaključiti da prosečan SJR pokazatelj za 244 časopisa iznosi 0,218 ($\approx 0,22$), dok varijansa populacije ima vrednost 0,019 ($\approx 0,02$).

Prost slučajni uzorak bez ponavljanja

Prost slučajni uzorak predstavlja plan uzorka u kome se n različitih jedinica bira iz populacija, ali na takav način da svaka moguća kombinacija od n jedinica ima istu verovatnoću da bude izabrana u uzorak. Naime, uzorak od n jedinica predstavlja slučajni uzorak sa ponavljanjem, pod uslovom da se svaka jedinica nakon izbora vraća u populaciju i da može da bude ponovo izabrana. S druge strane, uzorak bez ponavljanja predstavlja izabranu jedinicu koja se može odstraniti iz populacije, ali i vratiti u istu tako da bude zanemarena prilikom kasnijeg izvlačenja. Dakle, jedinice se mogu birati jedna po jedna, ali jedinica koja je jednom izabrana više se ne vraća u osnovni skup. U svakom koraku jedinica osnovnog skupa koja još uvek nije izabrana ima istu verovatnoću da bude izabrana (Petrović, 2018). Izbor jedinica iz populacije u uzorak može da se izvede pomoću tablice slučajnih brojeva, tehnikom lutrijskog izbora ili pomoću računara. Takođe, potrebno je naglasiti da ako se dogodi da se tokom izvlačenja izabere jedinica koja je već izabrana u nekom od prethodnih izvlačenja, tada se ta jedinica eliminiše, a postupak se nastavlja.

Prost slučajni uzorak je jednostavan i njegov sastav zavisi od verovatnoće da baš pravi članovi populacije budu izabrani. Prednost ovog uzorka predstavlja to što je on jednostavan za odabir. S druge strane, njegovi nedostaci su što je potrebno znati veličinu populacije, svaki član populacije mora da je dostupan i nije efikasan kada je raspodela populacije asimetrična. Upravo zbog toga se dešava da ovakav uzorak ne reprezentuje uvek u potpunosti karakteristike populacije, pa samim tim ovaj način odabira uzorka nije uvek statistički najefikasniji (Petrović, 2018).

U ovom delu rada iz ukupne populacije od 244 elementa bira se jednostavan (prost) slučajni uzorak bez ponavljanja od 50 elementa. Uzorak je dobijen pomoću Microsoft Office 2007 paketa EXCEL, tačnije pomoću opcije RAND.

U tabeli koja sledi prikazane su ocene parametara populacije na osnovu prostog slučajnog uzorka.

Tabela 2: Ocena parametara populacije na osnovu prostog slučajnog uzorka

Parametar	Oznaka	Vrednost
Sredina	\bar{y}	0,194080
Varijansa	s^2	0,006859749
Varijansa sredine	$V(\bar{y})$	0,00051336
Ocena varijanse sredine	$\hat{V}(\bar{y})$	0,000172555

Izvor: Proračun autora.

Nakon što je izvršen odabir prostog slučajnog uzorka bez ponavljanja i ocene sredine obeležja totala, preporučljivo je ispitati i tačnost te ocene. Naime, to se može uraditi određivanjem intervala poverenja, u okviru kojih se sa dovoljnom sigurnošću nalaze vrednosti populacije. Parametar \bar{Y} nije poznat, ali je fiksiran. Za razliku od njega, krajevi intervala predstavljaju statistiku koja se menja zavisno od uzorka. Veličina $1-\alpha$ predstavlja interval poverenja. Prilikom izračunavanja intervala poverenja $\alpha=0,05$. Ovo znači da za 95% mogućih uzoraka n veličine, interval poverenja poseduje tačnu vrednost sredine obeležja populacije \bar{Y} . S obzirom da, prema Centralnoj graničnoj teoremi, uzorak $n>30$ (veliki uzorak), raspodela $\frac{\bar{y}-\bar{Y}}{\sqrt{V(\bar{y})}}$, teži približnoj raspodeli. Uzorak koji je posmatran ima veličinu 50, odnosno $n=49$. Može se zaključiti da sa veličinom uzorka od 50, može izračunati 95% interval poverenja za sredinu obeležja populacije koji se zasniva na normalnoj aproksimaciji za raspodelu sredine uzorka kod prostog slučajnog uzorka bez ponavljanja. U tabeli koja sledi prikazan je izračun intervala poverenja.

Tabela 3: Izračun intervala poverenja

Opis	Vrednost
\bar{y}	0,194080
Z	1,96
$\hat{V}(\bar{y})$	0,000172555
$\sqrt{\hat{V}(\bar{y})}$	0,013136
$\bar{y} - z\sqrt{\hat{V}(\bar{y})}$	0,16833344
$\bar{y} + z\sqrt{\hat{V}(\bar{y})}$	0,21982656

Izvor: Proračun autora.

Vrednost z ima vrednost 1,96 i dobijena je iz tablica za standardizovanu normalnu raspodelu $N(0,1)$, takva da je $P\{|Z| \leq z\} = 1-\alpha$, gde je $Z:N(0,1)$ (Petrović, 2018).

Posmatrajući tabelu može se zaključiti da je interval poverenja (0,168;0,220) za koji se tvrdi da obuhvata vrednost obeležja populacije.

Stratifikovani uzorak

Za populacije kod kojih je prisutan visok stepen heterogenosti u smislu jakog varijabiliteta obeležja koja se istražuju, pri ograničenjima u pogledu veličine uzorka, teško je na bazi jednostavnog slučajnog uzorka izvesti ocenu zadovoljavajuće preciznosti (Etikan & Bala, 2017). U skladu sa teorijom uzoraka, stratifikovani uzorak pruža preciznost ocene. Da bi se povećala preciznost uzorka kod ove vrste uzorka vrši se stratifikacija. Stratifikacija predstavlja proces u kome se vrši podela populacije na delove koji se nazivaju

stratumi. Prilikom podele populacije potrebno je voditi računa da stratumi budu relativno homogeni, ali i međusobno razgraničeni. Iz prethodno navedenog može se reći da stratifikovani uzorak predstavlja uzorak koji se zasniva na podeli osnovnog skupa na slojeve ili stratume, a iz stratuma se biraju manji jednostavniji slučajni (pod)uzorci.

Pomoću stratifikacije može se postići izuzetno velika preciznost u ocenama cele populacije. Populacije koje su heterogene mogu se podeliti na potpopulacije, a svaka od njih je homogena unutar sebe. Dakle, postupak stratifikacije dovodi do povećanja efikasnosti ako se izvrši podela populacije na stratume unutar kojih su jedinice homogene imajući u vidu varijacije proučavane varijable. Njen značaj ogleda se u grupisanju populacije u stratume zasnovane na dopunskim informacijama i izboru odvojenih uzoraka iz svakog stratuma. Korist stratifikacije ogleda se u tome što su veličine uzorka u stratumima kontrolisane od strane istraživača u odnosu na izbor slučajnog uzorka (Kanlić i dr., 2012). Naime, ako je svaki stratum homogen, razlika u merenjima između jedinica su male, pa se samim tim precizne ocene mogu dobiti na osnovu malog uzorka stratuma. Ocene koje su dobijene na ovaj način mogu biti precizne i kada se prenesu na celu populaciju.

Stratifikovani uzorak predstavlja varijantu prostog slučajnog uzorka i sistematskog uzorka i primenjuje se onda kada postoji više različitih podgrupa, a u kojima se zahteva potpuna reprezentativnost uzorka. Poces stratifikacije jedinica osnovnog skupa u uzorku nosi sa sobom fleksibilnost jer pruža mogućnost upotrebe različitih tehnika alokacije (Kanlić i dr., 2012).

Stratifikovani uzorak konstruiše se tako što se izvrši klasifikacija populacije u podgrupe, a koje su zasnovane na nekoj karakteristici populacije. Nakon toga vrši se izbor elemenata za svaki stratum posebno. Izbor elemenata može da se vrši metodom slučajnog uzorka ili metodom sistematskog uzorka. Kada se vrši proces stratifikacije potrebna su dva uslova, a to su: proporcija populacije u stratumu (ponder W_h) treba da je unapred poznata, treba biti moguće da se izvrši izbor odvojenih uzoraka iz svakog stratuma. Naime, prilikom izbora uzorka potrebno je da se iz svakog stratuma izabere najmanje jedan uzorak. U suprotnom neće biti moguće izračunati nepristrasni procenitelj za aritmetičku sredinu cele populacije (Kanlić i dr., 2012).

Kada se govori o stratifikovanom uzorku potrebno je naglasiti da se razlikuju dve vrste ovog uzorka, a to su:

- *Proporcionalni stratifikovani uzorak*- veličina uzorka u stratumima je proporcionalna veličini ciljane populacije. Odnosno, u ciljnoj populaciji obim uzorka koji je uzet iz svakog stratuma proporcionalan je relativnoj veličini tog stratuma. (Iliyasu & Etikan, 2021) Dakle, može se zaključiti da svaki stratum ima istu frakciju izbora. Pomoću ovog uzorka postiže se ista ili veća preciznost u odnosu na jednostavni slučajni uzorak koji je iste veličine. Takođe, preciznost je veća ako su vrednosti u stratumima homogene.
- *Disproporcionalni stratifikovani uzorak*- veličina uzorka u stratumima nije proporcionalna veličini ciljane populacije. Dakle, to znači da se prilikom kreiranja uzoraka iz nekih stratuma uzima više, a iz nekih manje elemenata. Ova tehnika primenjuje se onda kada troškovi prikupljanja podataka variraju od grupe do grupe ciljane populacije stratuma, u situacijama kada je za neku grupu teže proceniti stvarnu veličinu ili kada se od jedne grupe očekuje da bude manje kooperativna u istraživanju od ostalih, pa je samim tim potreban veći broj njenih članova kako bi uzorak bio reprezentativan. Ako se varijanse razlikuju kroz stratume disproporcionalna stratifikacija je preciznija u odnosu na proporcionalnu stratifikaciju.

Može se zaključiti da se teorija stratifikovanog uzorka bavi svojstvima ocena koje su dobijene pomoću stratifikovanog uzorka, najboljim izborima veličina uzorka i optimalnim rasporedom, a cilj svega toga je postizanje što veće preciznosti prilikom ocenjivanja parametara.

Postupak stratifikacije sastoji se od toga da se populacija koja sadrži N jedinica deli na L populacije, odnosno stratume koji nemaju zajedničke elemente. Neka je broj elemenata u h -tom stratumu N_h , gde je $h=1, \dots, L$. Tada je $N_1 + N_2 + \dots + N_L = L$. Nakon što se izvrši formiranje stratuma, po određenom planu se bira uzorak iz svakog stratuma, a elementi iz stratuma su među sobom nezavisni. Neka je n_h veličina uzorka iz h -tog stratuma tada je $n_1 + n_2 + \dots + n_L = n$ obim uzorka (Popović, 2015).

Vrednost i -te jedinice u h -tom stratumu označava se sa y_{hi} . S druge strane, $W_h = \frac{N_h}{N}$ predstavlja relativnu frekvenciju uzorka u stratumu (težina stratuma), $f_h = \frac{n_h}{n}$ predstavlja frakciju uzorka u stratumu, a Y_h total stratuma.

Naime, ako se iz svakog stratuma bira prost slučajni uzorak, ceo postupak naziva se stratifikovani slučajni uzorak. Ovaj oblik uzorka veoma često se kombinuje sa drugim planovima uzorka, kao što su uzorak sa nejednakim verovatnoćama, sistematski uzorak, uzorak skupine itd., u zavisnosti od potrebe koje određuju vrste i osobine populacije (Tolga, 2021).

Stratifikovani uzorak – rezultati istraživanja

U radu je izvršeno formiranje stratuma na tržištu naučnih časopisa iz Republike Srbije, Bosne i Hercegovine i Hrvatske. Kao kriterijum za formiranje stratuma upotrebljeno je rangiranje časopisa, na osnovu visine SJR pokazatelja. Kao odgovarajuća vrednost obeležja, praćen je SJR pokazatelj. SJR predstavlja prosečan broj ponderisanih citata primljenih u godini prema člancima objavljenim u časopisima u prethodne 3 godine. Dakle, SJR pokazatelj zasnovan je na prenosu prestiža sa jednog časopisa na drugi. Prestiž se prenosi kroz citate i na taj način prestiž jednog časopisa ima uticaj na ostale i povratno.

Na osnovu ukupne pozicije časopisi su podeljeni u 3 kategorije, s obzirom na visinu SJR pokazatelja. Časopisi su podeljeni u sledeće kategorije: SJR pokazatelji čija je vrednost od 0,1 do 0,153, SJR pokazatelji čija je vrednost od 0,155 do 0,26 i SJR pokazatelji čija je vrednost od 0,265 do 0,824.

U tabeli koja sledi objašnjene su ove kategorije.

Tabela 4: Kategorije časopisa sa SJR liste

Stratum	Značenje	Ukupno časopisa
1	Visoki nivo (SJR vrednost od 0,265 do 0,824)	68
2	Srednji nivo (SJR vrednost od 0,155 do 0,26)	71
3	Nizak nivo (SJR vrednost od 0,1 do 0,153)	105

Izvor: Proračun autora.

Prvi korak u ovom procesu bilo je formiranje stratuma. Granice stratuma određene su aproksimativnom metodom Dalenijusa (*eng. Dalenius*) i Hodžesa (*eng. Hodges*):

$$Z_h = \int_{y^0}^{y_h} \sqrt{f(t)} dt$$

Kada je broj stratuma veliki, $f(y)$ je približno konstanta f_h unutar datog stratuma. Kada se izvrši zamena vrednosti dobijamo da je:

$$\sqrt{12} \sum_{h=1}^L W_h S_h = \sum_{h=1}^L f_h (y_h - y_{h-1})^2 = \sum_{h=1}^L (Z_h - Z_{h-1})^2$$

S obzirom da je $Z_L - Z_0$ fiksirano, lako je proveriti da je suma koja se nalazi sa desne strane minimalna ako je $Z_h - Z_{h-1}$ konstantno. Zato se za dato $f(y)$ vrši formiranje kumulativne funkcije za funkciju $\sqrt{f(y)}$. Tačke y_h biraju se tako da formiraju konstantne intervale na kumulanti $\text{cum}\sqrt{f(t)}$. (Petrović, 2018: 64)

U našem slučaju na osnovu obračuna frekvencije SJR pokazatelja, izračunate su kumulante i dobijena je najveća vrednost kumulante 182,2299. Ovaj iznos podeljen je sa 3 (jer određujemo tri stratuma) i na taj način je dobijena prva granica koja iznosi 60,74329945 ($\approx 60,74$). Vrednost druge granice dobijena je kada je $60,74329945 \cdot 2 = 121,4865989$. U tabeli kumulanti izabrane su najbliže vrednosti koje odgovaraju dobijenim vrednostima, na ovaj način dobijene su granice intervala kumulanti, a na osnovu njih se određuju granice intervala koje se odnose na vrednost SJR pokazatelja, koji se koristi za određivanje granica stratuma.

Nakon što su određeni stratumi izvršen je izbor uzorka iz svakog stratuma. Izvršeno je biranje elemenata iz prvog, drugog i trećeg stratuma. Sam izbor elemenata iz stratuma je međusobno nezavisan. Iz svakog stratuma izabran je prost slučajni uzorak, a upravo to predstavlja stratifikovani slučajni uzorak. Stratumi će biti obeleženi indeksima h ($h=1, \dots, L$), dok će sa i biti označene jedinice u okviru stratuma. Sa N_h označen je ukupan broj jedinica stratuma, a sa Y_h total stratuma, \bar{Y}_h sredina stratuma. S druge strane, sredina uzorka označena je sa \bar{y}_h , relativna frekvencija (težina stratuma) sa W_h , a frakcija uzorka u stratumu označena je sa f_h (Petrović, 2018).

Naime, relativna frekvencija kod stratuma koje su izračunate nisu međusobno jednake. Kada imamo slučajeve da su relativne frekvencije svih stratuma $W_h = \frac{N_h}{N}$ jednake, onda se ocene sredine uzorka i sredine stratuma poklapaju, a tada kažemo da je to stratifikacija sa proporcionalnim raspodelama.

U tabeli koja sledi prikazane su osnovne informacije o stratumima.

Tabela 5: Informacije o stratumima

Opis	Stratum 1	Stratum 2	Stratum 3	Ukupno (N)
Veličina stratuma (N_h)	68	71	105	244
Sredina stratuma (\bar{Y}_h)	0,391	0,201	0,117	
Total stratuma (Y_h)	26,580	14,270	12,235	
Maksimum	0,824	0,260	0,153	
Minimum	0,265	0,155	0,100	
Rang	0,559	0,105	0,053	
Varijansa stratuma (S_h^2)	0,019701061	0,001071443	0,000293579	
Standardna devijacija stratuma (S_h)	0,140360	0,032733	0,017134	
Težina stratuma (W_h)	0,279	0,291	0,430	1,000

Izvor: Proračun autora.

Tabela 6: Obrada podataka iz stratuma

Opis	Stratum 1	Stratum 2	Stratum 3	Ukupno
N_h	68	71	105	244
S_h^2	0,019701061	0,001071443	0,000293579	
Standardna devijacija (S_h)	0,140360467	0,032733	0,01713414	
Težina stratuma (W_h)	0,278688525	0,290983607	0,430327869	1
$W_h * S_h$	0,039116851	0,009524737	0,007373298	0,056014886

Izvor: Proračun autora.

Što se tiče uzorka, u prvom stratumu uzorak iznosi 35, u drugom stratumu je 9, a u trećem je 6, što je ukupno 50 elemenata. Ista veličina upotrebljena je i kod primene prostog slučajnog uzorka. U tabeli 7. prikazano je računanje varijanse unutar stratuma i varijanse između stratuma. Važno je naglasiti da se varijansa stratifikovanog osnovnog skupa sastoji od varijanse unutar stratuma i varijanse između stratuma, a to se upotrebljava i kao provera prilikom sprovođenja analize.

Tabela 7: Varijansa unutar stratuma i varijansa između stratuma

Opis	Stratum 1	Stratum 2	Stratum 3
Veličina stratuma (N_h)	68	71	105
Sredina stratuma (\bar{Y}_h)	0,391	0,201	0,117
Total stratuma (Y_h)	26,580	14,270	12,235
Maksimum	0,824	0,260	0,153
Minimum	0,265	0,155	0,100
Rang	0,559	0,105	0,053
Varijansa stratuma (S_h^2)	0,019701061	0,001071443	0,000293579
Standardna devijacija stratuma (S_h)	0,140360	0,032733	0,017134
Težina stratuma (W_h)	0,279	0,291	0,430
$N_h - 1$	67	70	104
$(N_h - 1) * S_h^2$	1,319971059	0,075000986	0,03053219
$\Sigma(N_h - 1)S_h^2$	1,425504235		
$1/(N - 1)$	0,004115226		
Varijansa unutar stratuma (S_u^2)	0,005866273		
\bar{Y}_h	0,391	0,201	0,117
$(\bar{Y}_h - \bar{Y})$	0,173	-0,017	-0,101
$(\bar{Y}_h - \bar{Y})^2$	0,0300401	0,0002747	0,0102086

$N_h(\bar{Y}_h - \bar{Y})^2$	2,043	0,020	1,072
$\Sigma N_h(\bar{Y}_h - \bar{Y})^2$	3,134		
Varijansa između stratuma (S^2)	0,012897695		
S^2	0,018763967		

Izvor: Proračun autora.

Kao što se može videti iz prethodne tabele u našem slučaju varijansa unutar stratuma ima vrednost 0,005866273 ($\approx 0,0059$), a varijansa između stratuma iznosi 0,012897695 ($\approx 0,0129$). Na osnovu ovoga može se zaključiti da je stratifikacija dobro sprovedena, odnosno da su jedinice unutar stratuma sličnih karakteristika, dok se jedinice iz različitih stratuma značajno razlikuju. S druge strane, zbir varijanse unutar stratuma i varijanse između stratuma jednak je varijansi stratifikovanog osnovnog skupa, odnosno $0,005866273 + 0,012897695 = 0,018763967$ ($\approx 0,019$).

Sledeći korak jeste izbor elemenata iz svakog stratuma. Naime, slučajnim izborom biramo unapred definisan broj elemenata iz svakog stratuma. Za izbor slučajnih elemenata primenjena je funkcija slučajnih brojeva RANDBETWEEN. U analizi je izvršeno određivanje prosečnog SJR pokazatelja u stratifikovanom uzorku, a to je prikazano u tabeli koja sledi.

Tabela 8: Prosečna vrednost stratifikovanog uzorka

Opis	Stratum 1	Stratum 2	Stratum 3
S_h^2	0,019701061	0,001071443	0,000293579
\bar{y}_h	0,4115	0,198	0,118
Maksimum	0,824	0,250	0,147
Minimum	0,265	0,159	0,101
s_h^2 (dostupni su u izračunu)	0,025567382	0,00125494	0,00039977
Total (y_h)	14,814	1,784	0,707
$N_h * \bar{y}_h$	27,982	14,07377778	12,3725
$N_h(N_h - n_h)S_h^2/n_h$	1,263119427	0,524054508	0,508625192
$N_h(N_h - n_h)s_h^2/n_h$	1,639234457	0,613807272	0,69259575

Izvor: Proračun autora.

Izračunom dobijen je rezultat da prosečan SJR pokazatelj ima vrednost 0,223. U populaciji prosečna vrednost SJR pokazatelja iznosi 0,218, a vrednost dobijena primenom prostog slučajnog uzorka bez ponavljanja iznosi 0,194. Na osnovu ovoga može se zaključiti da je ocena sredine koja je dobijena stratifikacijom preciznija.

Nakon što je izvršeno ocenjivanje sredine obeležja populacije na osnovu stratifikovanog slučajnog uzorka, određuje se interval poverenja. Pomoću intervala poverenja procenjuje se tačnost ocena. U radu je utvrđen interval poverenja sa nivoom pouzdanosti $1 - \alpha$, gde u ovom slučaju α ima vrednost 0,05. S obzirom da je reč o dovoljno velikom uzorku, odnosno uzorku gde je $n > 30$, može se pretpostaviti da sredina ima približno normalnu raspodelu bez obzira na to koju raspodelu ima obeležje populacije. Interval poverenja koji je dobijen obračunom iznosi 0,209280153 ($\approx 0,21$) do 0,236853272 ($\approx 0,24$).

U tabeli koja sledi prikazan je interval poverenja za sredinu populacije.

Tabela 9: Interval poverenja za sredinu populacije

Opis	Vrednost
\bar{y}_{st}	0,223067
$V(\bar{y}_{st})$	0,000038562
$\hat{V}(\bar{y}_{st})$	0,000049477
Koren ocene varijanse sredine	0,007033959
$z_{0,05}$	1,96
$\bar{y}_{st} - z \sqrt{\hat{V}(\bar{y}_{st})}$	0,209280153
$\bar{y}_{st} + z \sqrt{\hat{V}(\bar{y}_{st})}$	0,236853272

Izvor: Proračun autora.

Zaključak

Prost slučajni uzorak bira se iz relativno homogenog statističkog skupa i to je onaj uzorak kod koga svaka elementarna jedinica ima podjednaku ili unapred poznatu relativnu verovatnoću da će biti izabrana u uzorak. Naime, jednaka verovatnoća izbora obezbeđuje reprezentativnost i objektivnost uzorka. U praksi najviše se upotrebljava prost slučajni uzorak bez ponavljanja, jer su po pravilu osnovni skupovi veliki i nejednakost verovatnoće izbora jedinica je zanemarljiva. Za ove uzorke se može reći da su efikasni.

Upotrebom stratifikovanog uzorka postižu se ocene koje su izuzetno precizne. Međutim, ne može se reći da su varijanse kod stratifikovanog uzorka uvek manje nego kod prostog slučajnog uzorka. U slučajevima kada je veličina uzorka unutar stratuma daleko od optimalne, one mogu da budu i znatno veće. U retkim slučajevima dešava se da i onda kada je optimalan raspored postignut, varijanse ocene budu velike, ali u opštem slučaju to nije slučaj. Takođe, potrebno je naglasiti da bi najbolje bilo kada bismo mogli da vršimo stratifikaciju na osnovu one veličine koju želimo da ocenimo. Na taj način bi se izbeglo preklapanje među stratumima, a varijansa unutar stratuma bi bila znatno manja. Međutim, u praksi to u većini slučajeva nije izvodljivo, ali moguće je pažljivo izvršiti odabir kriterijuma stratifikacije, tako da se ostvari značajan napredak u preciznosti. Kako bi se ovo postiglo, potrebno je uzeti u obzir činjenicu da se jedinice od kojih se sastoji populacija značajno razlikuju po veličini, te da su najvažnije promenljive, čije se vrednosti određuju, povezane sa veličinama jedinica. Pored ovoga, potrebno je tačno i precizno odrediti veličine jedinica i stratuma. Naime, kada se jedinice značajno razlikuju po veličini, stratifikacija sa proporcionalnim rasporedom ne daje dobre rezultate. Razlog za to je što je varijansa veća kada su jedinice veće. Pored ovoga, u obzir je potrebno uzeti i troškove sprovođenja istraživanja, jer su ponekad troškovi ispitivanja određenih elemenata uzorka suviše veliki da bi bili opravdani preciznošću koja se postigne.

Cilj rada je bio da se ispita preciznost ocene prostog slučajnog uzorka i stratifikovanog slučajnog uzorka. Kao populacija u radu poslužilo je tržište naučnih časopisa i to onih sa teritorije Republike Srbije, Bosne i Hercegovine i Republike Hrvatske. Odnosno svi časopisi koji imaju SJR pokazatelj, a dostupni su na "Scimago Journal & Country Rank" za 2019. godinu. Detaljnom analizom zaključeno je da stratifikovani uzorak ima preciznije ocene. Samim tim dokazana je i početna hipoteza rada koja glasi "Pomoću stratifikovanog slučajnog uzorka obezbeđuju se preciznije ocene, u odnosu na jednostavni slučajni uzorak bez ponavljanja." Takođe, može se zaključiti da je potreba za stratifikacijom u ovom slučaju opravdana.

Prosečni SJR pokazatelj u populaciji ima vrednost 0,218 ($\approx 0,22$). S druge strane, u stratifikovanom uzorku vrednost SJR pokazatelja iznosi 0,223 ($\approx 0,22$), a u jednostavnom slučajnom uzorku bez ponavljanja je 0,194 ($\approx 0,19$). Na osnovu ovoga, može se zaključiti da je SJR pokazatelj dobijen stratifikovanim uzorkom znatno bliži, odnosno skoro identičan kao vrednost SJR pokazatelja u populaciji. Takođe, stratifikovani uzorak ima i užu interval poverenja za sredinu obeležja populacije \bar{Y} . Preciznost stratifikacije može se posmatrati i praćenjem vrednosti varijanse i sredine uzorka. Ukoliko dve statistike predstavljaju nepristrasne ocene parametara populacije, u tom slučaju kao kriterijum kvaliteta ocene koristi se ocena varijanse. Naime, bira se ona ocena čija vrednost ima manje odstupanje od parametara populacije. Odnosno bira se ona ocena koja ima manju varijansu i za koju se može zaključiti da je efikasnija. Upotrebom plana stratifikovanog slučajnog uzorka vrednost varijanse prosečnog SJR pokazatelja iznosi 0,000038562 ($V(\bar{Y}_{st}) \approx 0,000039$). S druge strane, vrednost varijanse SJR pokazatelja koji je dobijen primenom prostog slučajnog uzorka bez ponavljanja iznosi 0,00051336 ($V(\bar{y}) \approx 0,00051$). Iz ovoga se može zaključiti da je vrednost varijanse SJR pokazatelja iz stratifikovanog uzorka $V(\bar{Y}_{st})$ značajno niža u odnosu na vrednost varijanse prosečnog SJR pokazatelja ($V(\bar{y})$).

Što se tiče budućih istraživanja bilo bi interesantno da se izvrši uključivanje i drugih planova uzorka, kao što su sistematski, višestapni ili dvofazni uzorak, kao i merenje njihove preciznosti. Istraživanje bi na ovaj način dobilo viši nivo sadržajnosti i pouzdanosti.

Literatura

1. Đorđević, V. (2006). *Statistika u ekonomiji*. Niš: Ekonomski fakultet.
2. Etikan, I., & Bala, K. (2017). Sampling and sampling methods. *Biometrics & Biostatistics International Journal, Volume 5 (6)*, 215-217.
3. Iliyasa, R., & Etikan, I. (2021). Comparison of quota sampling and stratified random sampling. *Biometrics & Biostatistics International Journal*, 24-27.

4. Jevremović, V. (2019). *Rečnik statističkih termina, drugo izdanje*. Beograd: Republički zavod za statistiku.
5. Kanlić, F., Abdić, A., & Abdić, A. (2012). Uzorkovanje i tehnike alokacije kod stratifikovanog uzorka. *Zbornik radova- Sarajevo Business and Economics Review (SBER), NO. 32.,* , 285-300.
6. Kojić, M. (2015). *Primene nekih planova teorije uzoraka u ekonomiji, (Master rad)*. Niš: Univerzitet u Nišu, Prirodno-matematički fakultet.
7. Kostić Kovačević, I. (2019). *Verovatnoća i statistika*. Beograd: Univerzitet Singidunum.
8. Marjanović, M., Mihailović, I., & Spasić, K. (2016). *Statistika u ekonomiji i poslovanju sa zbirkom rešenih zadataka*. Leskovac: Visoka poslovna škola strukovnih studija Leskovac.
9. Petrović, L. (2018). *Teorija uzoraka i planiranje eksperimenata*. Beograd: Centar za izdavačku delatnost, Ekonomski fakultet.
10. Petrović, L. (2018). *Teorijska statistika- Teorija statističkog zaključivanja, Četvrto izdanje*. Beograd: Centar za izdavačku delatnost Ekonomskog fakulteta u Beogradu.
11. Popović, B. (2015). *Matematička statistika, drugo izdanje*. Niš: Prirodno-matematički fakultet Niš.
12. Rakić, D. (2010). Određivanje veličine uzorka. *Vodič kroz metodologiju nastave primenjene statistike* (ss. 68-74). Novi Sad: Univerzitetski centar za primenjenu statistiku Univerziteta u Novom Sadu.
13. Rapajić, S. (2010). Popis ili uzorkovanje. *Vodič kroz metodologiju nastave primenjene statistike* , 12-16.
14. Tolga, Z. (2021). An efficient exponential estimator of the mean under stratified random sampling. *Mathematical Population Studies*, 28 (2) , 104-121
15. Scimago Journal & Country Rank. (april 2020). preuzeto sa: Journal Rankings: <https://www.scimagojr.com/journalrank.php?country=ID> (14. 01.2021.).