Univerzitet u Beogradu

Fakultet organizacionih nauka

Aleksandar M. Đoković

STRUKTURNA KORELACIONA ANALIZA U INTERPRETACIJI VEKTORSKIH KOEFICIJENATA KORELACIJE

Doktorska disertacija

Beograd, 2013.

University of Belgrade

Faculty of organizational sciences

Aleksandar M. Đoković

STRUCTURAL CORRELATION ANALYSIS INTERPRETED BY VECTOR CORRELATION COEFFICIENTS

Doctoral dissertation

Belgrade, 2013.

Mentor:

____

dr Zoran Radojičić,

vanredni profesor Fakulteta organizacionih nauka

Članovi komisije:

____

dr Milica Bulajić,

redovni profesor Fakulteta organizacionih nauka

____

dr Dragan Vukmirović,

redovni profesor Fakulteta organizacionih nauka

____

dr Milan Martić,

redovni profesor Fakulteta organizacionih nauka

____

dr Srđan Bogosavljević,

redovni profesor Ekonomskog fakulteta

Datum odbrane:

____

Datum promocije:

____

Strukturna korelaciona analiza u interpretaciji vektorskih koeficijenata korelacije

Rezime:

U uvodnom poglavlju se opisuju predmet i cilj istraživanja, navode se polazne hipoteze i metode istraživanja, daje sadržaj i opis disertacije uz navođenje ključnih aspekata na koje će se disertacija usmeriti.

Drugo poglavlje je posvećeno konceptu proste linearne korelacije, kanoničkoj korelacionoj analizi i vektorskom koeficijentu korelacije. Kod utvrđivanja veze između dve posmatrane varijable posebnu važnost ima koeficijent proste linearne korelacije. Na taj način način dobijamo jedan statistički pokazatelj, koji osim toga što pokazuje stepen linearne povezanosti može poslužiti i za predviđanje jedne varijable u odnosu na drugu korišćenjem linearne jednačine u slučaju da su varijable visoko korelirane. Međutim, veoma retko je u praksi da jedna veličina zavisi od neke druge, već je češći slučaj da u okviru jednog sistema i ulaz i izlaz sistema zavise od više varijabli. Jedan pristup generalizaciji ovog problema je metod multivarijacione analize koji se bavi utvrđivanjem postojanja veza i jačine povezanosti dva skupa promenljivih koji nazivamo kanonička korelaciona analiza. Ova metoda, u slučajevima kada je moguće a priori uspostaviti relaciju imeđu dva skupa promenljivih, omogućava da kvantifikujemo međusobnu povezanost i detaljno ispitamo takvu vezu (Kovačić, 1992). Originalan teorijski doprinos razvoju kanoničke korelacione analize dao je Hotelling 1936. godine. Drugi pristup generalizaciji je utvrđivanje korelacije između dva vektora. Postoje više predloženih definicija vektorskog koeficijenta korelacije (Detzius 1916, Sverdrup 1917, Charles 1959, Buell 1971, Breckling 1989, Crosby 1991), a u istraživanju kandidata, za utvrđivanje veze između m-dimenzionalne promenljive Y i n-dimenzionalne promenljive X koristićemo jednačinu

$|W| = (1 - R_v^2)\max|W| = (1 - R_v^2)|W_{yy}||W_{xx}|$ gde je vrednost označena sa R_v nazvana vektorski koeficijent korelacije (Vuković, 1977). Kod upoređivanja dve proste linearne korelacije postoje definisani statistički testovi (Fisher, 1921), dok su statistički testovi za poređenje dve korelacione, odnosno kovarijacione strukture veoma kompleksni i zahtevaju korišćenje moćnih alata. Zbog važnosti analize jedne korelacione strukture, u okviru istraživanja ćemo dati predlog test statistike za poređenje dva vektorska koeficijenta korelacije, na osnovu koje će biti baziran model strukturne korelacione analize. Do sada, pažnja nije bila usmerena na vektorski koeficijent korelacije i ono što se može dobiti njegovom interpretacijom u raznim organizacionim sistemima, pa će istraživanje u tom pravcu dati novu dimenziju u sagledavanju problematike u višedimenzionalnom rasporedu.

U trećem poglavlju pažnja se posvećuje multivarijacionoj statističkoj analizi. Multivarijaciona statistička analiza obezbeđuje mogućnost analize kompleksnih nizova podataka, tamo gde ima mnogo nezavisnih i zavisnih promenljivih koje su međusobno korelisane na različitim nivoima povezivanja. U ovom poglavlju, glavni akcenat je stavljen na dve ključne statističke tehnike: faktorsku analizu i analizu grupisanja. Faktorska analiza je statistička tehnika koja se koristi za identifikaciju relativno malog broja faktora koji se mogu koristiti za predstavljanje odnosa između grupa mnogobrojnih, međusobno povezanih, promenljivih. Na ovaj način se mogu identifikovati osnovne, ne direktno vidljive, dimenzije posmatrane pojave. Faktorska analiza i analiza glavnih komponenata imaju iste ciljeve i postupak njihovog sprovođenja je sličan, tako da metoda faktorske analize, može biti smatrana kao specijalni slučaj metode glavnih komponenata (Bulajić, 2002). Posebna pažnja je posvećena analizi grupisanja (klaster analizi), kao metodi multivarijacione statističke analize, koja se koristi za grupisanje objekata, tako da su objekti unutar grupe međusobno slični, a između grupa različiti. U okviru ovog poglavlja biće prikazan jedan a priori način grupisanja sa unapred definisanim ograničenjima, kao jedna modifikacija *K-mean* algoritma nehijerarhijskog grupisanja.

Četvrto poglavlje je bazirano na I-odstojanju kao metrici u n-dimenzionalnom prostoru, a koje je predloženo od strane prof. dr Branislava Ivanovića (Ivanović & Fanchette, 1973). Kao jedan od vodećih stručnjaka u odseku Ujedinjenih Nacija (UN) prof. Ivanović je kreirao ovu metodu sa ciljem da rangira zemlje na osnovu više kriterijuma. U ovom poglavlju dat je prikaz običnog, kvadratnog i strukturnog I-odstojanja. Glavni argument za korišćenje metode I-odstojanja je njena sposobnost da sintetizuje veliki broj varijabli u jednu numeričku vrednost. Posebna pažnja je posvećena problemu određivanja raspodele I-odstojanja. Pokazano je da kvadratno I-odstojanje ima normalnu raspodelu za slučajeve kada su varijable po kojima je vršeno rangiranje normalno raspoređene veličine. Takođe, u opštem slučaju, a za šta je korišćena Bootstrap metoda koja kao svoj sastavni deo podrazumeva primenu Monte-Karlo simulacije, pokazano je (osim u izuzetnim slučajevima) slaganje I-odstojanja sa teoretskom normalnom raspodelom.

Peto poglavlje je koncipirano na izgradnji modela strukturne korelacione analize zasnovanog na vektorskim koeficijentima korelacije. U ovom poglavlju je i najznačajniji deo disertacije koji se odnosi na definisanju test statistike za poredjenje dva vektorska koeficijenta korelacije, na osnovu koje se mogu porediti dve korelacione strukture. Na ovaj način je moguće utvrditi vezu između izlaznih i ulaznih veličina jednog organizacionog sistema, ali i videti razlike između dva različita organizaciona sistema. Rangiranje obeležja je česta pojava, ali rezultati rangiranja mogu imati vrlo ozbiljne posledice, kao što su prijemni ispiti, konkursi, participacija u UN i mnogi drugi slučajevi. Poseban osvrt u ovom istraživanju biće primena načina rangiranja metodom Ivanovićevog odstojanja, u okviru kojeg će biti inkorporiran odnos izlaznih i ulaznih veličina izražen kroz vektorski koeficijent korelacije. Naime, prillikom rangiranja entiteta metodom Ivanovićevog odstojanja eliminišu se korelacioni odnosi između ulaznih i izlaznih veličina, pa je osnovna ideja da se taj odnos izražen kroz vektorski koeficijent korelacije uzme u obzir prilikom rangiranja. Na taj način, bila bi dobijena jedna verna slika posmatranih objekata koji su radi postizanja preferenci rangirani tj. postavljeni u relacioni odnos.

U šestom poglavlju je dat zaključak sa odgovorom na pitanja u vezi sa postavljenim ciljem i hipotezama. Data je sistematizacija i pregled naučnih doprinosa koji su proistekli iz rada na doktorskoj disertaciji, skup otvorenih problema i mogućnosti za dalji rad u oblasti doktorske disertacije. Literatura sadrži skup relevantne i korišćene literature za oblast doktorske disertacije, sa pregledom literature koja se bavi navedenim oblastima. U prilogu su dati rezultati ekperimentalnog dela disertacije.

Ključne reči:

vektorski koeficijent korelacije, Ivanovićevo odstojanje, obrazovanje, rangiranje, multivarijaciona statisticka analiza

Naučna oblast:

Tehničke nauke

Uža naučna oblast:

Računarska statistika

UDK broj:

519.2

Structural correlation analysis interpreted by vector correlation coefficients

Abstract:

The introductory chapter describes the object and purpose of the research, states the initial hypotheses and research methods, gives a content and description of the dissertation by specifying the key aspects on which the thesis will be focused on.

The second chapter is devoted to the concept of simple linear correlation, canonical correlation analysis and vector correlation coefficient. In determining the relationship between two variables, the simple linear correlation coefficient plays a special role. It gives us a statistical indicator, which in addition to showing the degree of linear correlation can also be used to predict one variable according to another by using a linear equation if the variables are highly correlated. However, very rarely in practice one size depends on the other; it is more often that within a system, the input and output of the system depend on several variables. One approach to generalizing this problem is a multivariate analysis regarding the identification of links and the strength of correlation of two sets of variables called canonical correlation analysis. If it is possible to set an a priori relationship between two vital set of variables, this method allows us to quantify the interrelationship and thoroughly investigate such a link (Kovačić, 1992). Original theoretical contribution to the development of canonical correlation analysis was given by Hotelling in 1936. The second approach to a generalization is determening the correlation between the two vectors. There are several proposed definitions of vector correlation coefficient (Detzius 1916, Sverdrup 1917, Charles 1959, Buell 1971, Breckling 1989, Crosby 1991), and within researching the candidates, to determine the relationship between m-dimensional variable Y, and n-dimensional variable X, we will use an equation $|W| = (1-R_v^2)\max|W| = (1-R_v^2)|W_{yy}||W_{xx}|$ where value indicated by R_v is called the

vector correlation coefficient (Vuković, 1977). When comparing two simple linear correlations, there are defined statistical tests (Fisher, 1921), while the statistical tests for comparing two correlation and covariance structures are very complex and require the use of powerful tools. Because of the importance of the analysis of a correlation structure, within the research we will propose the statistics to compare two vector correlation coefficient, that the structural model of the correlation analysis wil be based on. So far, there were no attention focused on the vector correlation coefficient and what we can get by its interpretation in various organizational systems, so the research wil give a new dimension to understanding the problem in a multidimensional schedule.

In the third chapter, attention is given to multivariate statistical analysis. Multivariate statistical analysis provides the ability to analyze complex data sets, where there are a lot of independent and dependent variables that are correlated with each other at different levels of connectivity. In this chapter, the main focus is on two key statistical techniques: factor analysis and cluster analysis. Factor analysis is a statistical technique used to identify a relatively small number of factors that can be used to represent relationships between groups of numerous, interconnected variables. This way we can identify the basic, not directly visible, dimensions of the phenomenon. Factor analysis and principal component analysis have the same goals and the process of their implementation is similar, so that the method of factor analysis, can be considered as a special case of principal component analysis (Bulajić, 2002). Special attention is devoted to the grouping analysis (cluster analysis), as a method of multivariate statistical analysis, which is used to group objects, so that the objects within groups are similar to each other, and between groups quite different. This chapter will show an a priori grouping mehod of defined limits, as a modification of the K-mean clustering non-hierarchical algorithm.

The fourth chapter is based on the I-distance as a metric in an n-dimensional space, which is proposed by prof. Dr. Branislav Ivanović (Ivanović & Fanchette, 1973). As one of the leading experts in the department of the United Nations (UN), professor Ivanović has created this method in order to rank countries based on multiple criteria. This chapter presents the ordinary, rectangular and structural I-distance. The main argument for the use of I-distance method is its ability to synthesize large number of variables into a single numerical value. Special attention is given to the problem of determining the distribution of I-distance. It is shown that the square-distance is normaly distributed when the variables on which the ranking is based are normally distributed sizes. Also, in general, it was shown (except in exceptional cases) that the I-distance agrees with the theoretical normal distribution, using the bootstrap method, which as an integral part involves the use of Monte Carlo simulation.

The fifth chapter is based on building a structural model of correlation analysis based on vector correlation coefficients. This section is the most important part of the dissertation, refering to the definition of the test statistic for comparing two vector correlation coefficients, based on which we can compare two correlation structures. In this way it is possible to determine the relationship between the input and output size of an organizational system, and the differences between two different organizational systems. Ranking of features is a common occurrence, but the ranking results can have very serious consequences, such as entrance exams, competitions, participation in the UN and many other cases. A special emphasis in this study will be the use of Ivanović distance ranking method, within which we will incorporate ratio of output and input values expressed through the vector correlation coefficient. Through ranking entities by Ivanović distance method we eliminate correlations between input and output variables, and the basic idea is to to take this relationship based on the vector correlation coefficient into account within ranking entities. In this way, one would obtain accurate image of the the observed objects, which are ranked to achieve the preferences ie. set into relational relationship.

The sixth chapter gives a conclusion with answers to questions about the goals and hypotheses. We gave the review and systematization of scientific contributions arising from work on this doctoral dissertation, a set of open problems and opportunities for further work in the field of doctoral dissertation. Literature contains a set of relevant literature, used in the field of doctoral dissertation, with review of the literature dealing with these issues. The results of the experimental dissertational work are following.

Keywords:

vector correlation coefficient, Ivanović distance, education, ranking, multivariate statistical analysis

Scientific Area:

Technical Sciences

Specific Scientific Area:

Computational Statistics

UDK Number:

519.2

SADRŽAJ

1. UVOD **1**

1.1. Polazne hipoteze 4

1.2. Metode istraživanja 4

1.3. Doprinos doktorske disertacije 5

2. KORELACIONA ANALIZA **6**

2.1. Prosta linearna korelacija 8

2.1.1. Ocena koeficijenata proste linearne korelacije 12

2.1.2 Interpretacija koeficijenata proste linearne korelacije 16

2.1.3. Korelaciona matrica, višestruka i parcijalna korelacija 18

2.1.4. Poređenje Pirsonovog i Spirmanovog koeficijenta korelacije 21

2.1.5. Testiranje hipoteze za Pirsonov koeficijent korelacije 27

2.2. Kanonička korelaciona analiza 29

2.3. Vektorski koeficijent korelacije 30

3. MULTIVARIJACIONA STATISTIČKA ANALIZA **35**

3.1. Faktorska analiza i analiza glavnih komponenata 41

3.1.1. Model faktorske analize 43

3.1.2. Metoda glavnih komponenenata 45

3.2. Klaster analiza (analiza grupisanja) 51

3.2.1. Mere sličnosti i razlike između objekata 55

3.2.2. Mere sličnosti i razlike između grupa 58

3.2.3. Hijerarhijske metode grupisanja 61

3.2.4. Određivanje broja grupa (klastera) 62

3.3. Algoritam za rešavanje problema klasifikacije sa unapred definisanim ograničenjima 63

3.4. Analiza obavijanja podataka 69

4. IVANOVIĆEVO ODSTOJANJE **79**

4.1. Obično I-odstojanje 86

4.2. Kvadratno I-odstojanje 87

4.3. Strukturno I-odstojanje 88

4.4. Redosledna klasifikacija i I-odstojanje 90

4.5 Raspodela kvadratne forme slučajnih promenljivih koje imaju normalnu raspodelu 93

4.6 Ocena $I2$- odstojanja 98

4.7. Raspodela $I2$ – odstojanja 105
4.8. Raspodela $I2$ – odstojanja za slučajne promenljive koje nemaju normalnu raspodelu 112

5. MODEL STRUKTURNE KORELACIONE ANALIZE ZASNOVAN NA VEKTORSKIM KOEFICIJENTIMA KORELACIJE 116

5.1. Testiranje hipoteze o jednakosti dva vektorska koeficijenta korelacije 120
5.2. Primena metode I-odstojanja i vektorskog koeficijenta korelacije u rangiranju osnovnih škola u Srbiji 128
5.3. Izgradnja integralne lične karte osnovnih škola u Srbiji 139
5.3.1.Kriterijumi rangiranja škola u Velikoj Britaniji 139
5.3.2.Kriterijumi rangiranja škola u Americi 142
5.3.3.Kriterijumi rangiranja škola u Srbiji 148

6. ZAKLJUČAK 151

6.1. Doprinosi doktorske disertacije 152

7. LITERATURA 153

PRILOG 164

Prilog 1. 164
Prilog 2. 207

BIOGRAFIJA 215

IZJAVA O AUTORSTVU 218

IZJAVA O ISTOVETNOSTI ŠTAMPANE I ELEKTRONSKE VERZIJE DOKTORSKOG RADA 219

IZJAVA O KORIŠĆENJU 220

1. UVOD

U prirodi postoje događaji na kojima je moguće istovremeno definisati i proučavati dva ili više obeležja koja su veoma karakteristična za događaj. Tada do izražaja dolazi pitanje međusobne povezanosti tih obeležja, tj. da li se promena jednog od njih odražava na drugom. Ako takva povezanost postoji, onda se teži pronalaženju matematičke forme kojom se ta povezanost izražava. U proučavanju dve slučajne promenljive koje su merene na istom uzorku, veoma važnu ulogu ima koeficijent korelacije, koji meri stepen do kojeg su te dve mere u linearnoj vezi. Srodni koncept je regresioni model, u kojem je cilj da se pronađe linearna jednačina koja najbolje pokazuje vrednost jedne promenljive (ili jednog merenja), izražena preko vrednosti druge promenljive. Izračunavanje korelacije i regresionog modela zavisi od uređenog para koji se kontinuirano meri (x,y). Međutim, podaci se često predstavljaju u drugim oblicima u kojima dve promenljive nisu pogodne za uređene parove (x,y). U tom slučaju, odnos između dve promenljive može se predstaviti u tablicama kontigencije. U ovom slučaju, statističar je i dalje zainteresovan za proučavanje asocijacija između dve promenljive X i Y, i može ih izmeriti koristeći test homogenosti, testove zavisnosti ili izračunavanjem tetrahoričkog (polihoričkog) koeficijenta korelacije. Osnovno sredstvo ovih metoda je χ^2 raspodela. Istorijski gledano, i koeficijent korelacije i χ^2 raspodela nisu bili definisani i poznati u današnjem obliku, ali su koncepti koji stoje iza ovih savremenih statističkih alata prepoznatljivi u njihovim istorijskim definicijama. Ove ideje i alat uz pomoć kojih se primenjuju su razvijene tokom poslednjeg kvartala 19. veka i prvoj četvrtini 20. veka. Do sredine 19. veka, poznati matematičari, poput Paskala, Bernulija, Moavra, Simpsona, Laplasa, Gausa i Kvetela su razvili teoriju verovatnoće, meru centralne tendencije (tj. medijana), široku primenljivost greške odstupanja normalne raspodele, istorijsku centralnu graničnu teoremu.

Francis Galton, čovek odgovoran za koeficijent korelacije i rođak Čarlsa Darvina, prvi je upotrebio svoje naučne akreditive za istraživanja sprovedena u Africi u periodu od 1850. do 1852. godine. Može se zato reći da su njegov pionirski rad u području statistike, kao i njegov interes za statistiku, delovi njegovog nasleđa. Galton je komentarisao svoj odgovor na knjigu Čarlsa Darvina – „Poreklo vrsta“: „Bio sam ohrabren novim stavovima i istraživanjima koja me već duže vremena interesuju, i koje otvaraju krug centralnih tema nasleđivanja i mogućnost poboljšanja ljudske rase“. Tako je Galtonu palo na pamet da se normalni zakon raspodele može primeniti na proučavanje nasleđivanja. Kvetel je već pokazao merenjem, da grudi škotskih vojnika imaju normalni zakon raspodele, dok je Galton očekivao da kriva normalne raspodele može opisati promenljivost fizičkih i mentalnih karakteristika ljudi. U svom radu iz 1888. godine, Galton je predstavio korelaciju kraljevskom društvu u Londonu sledećom definicijom: „Za dva promenljiva organa kažemo da su u korelaciji kada je varijacija jednog praćena u manjoj ili većoj meri variranjem drugog organa, u istom pravcu... Lako se uviđa da korelacija predstavlja odnos variranja dva organa, koja su povezana uobičajenim uzrocima... Ako ne postoji povezanost usled uobičajenih uzroka, korelacija je jednaka nuli“. Ova Galtonova definicija nam otkriva svojstva koeficijenta korelacije. Međutim u prirodi, procesi i sistemi nisu jednostavni i ne mogu se svesti na samo dva parametra. Glavni problem u razmatranju takvih slučajeva je pitanje kako izmeriti koliki je uticaj jedne grupe na drugu grupu obeležja, putem nekog koeficijenta tj. izraziti postojeću vezu putem jednog broja. Rešenje za korelaciju jedne grupe obeležja i druge grupe obeležja daje nam vektorski koeficijent korelacije (Vuković, 1979).

U oblasti statističkog zaključivanja veoma važan deo predstavlja testiranje hipoteza, pa shodno tome i testiranje hipoteze o jednakosti dva koeficijenta korelacije. Obzirom da ćemo se baviti rešavanjem problema u višedimenzionom rasporedu, najveći napor biće usmeren na upoređivanju dve složene strukture kroz interpretaciju vektorskih koeficijenata korelacije. Predmet istraživanja u ovoj doktorskoj disertaciji biće određivanje funkcije (statistike) za upoređivanje dva vektorska koeficijenta korelacije, gde će nam kao osnova poslužiti statistika za nezavisnost jednog vektorskog koeficijenta korelacije.

Rangiranje obeležja je vrlo česta pojava, ali sa druge strane rezultat rangiranja može imati vrlo ozbiljne posledice, kao što su prijemni ispiti, konkursi, participacija u UN, kao i brojni drugi slučajevi. Ukoliko se radi o samo jednom obeležju, problem rangiranja je rešiv na više načina, ali zbog same vrste istraživanja potrebno je izvršiti rangiranje koje će na najrealniji način iskazati posmatrani problem. „Često se kaže da je klasifikovanje jedan od fundamentalnih procesa nauke. Činjenice i fenomeni moraju biti uređeni pre no što smo u stanju da ih shvatimo i razvijemo jedinstvene principe kojima se objašnjava njihova pojava i međusobni odnosi. Sa te tačke gledišta klasifikovanje je najviši nivo intelektualne aktivnosti neophodan za naše shvatanje prirode“ (Sokal,1977).

Pretpostavimo da se pomoću jednog kriterijuma mogu vršiti grupisanja elemenata skupa S. Svaka tako obrazovana grupa A predstavlja jedan podskup od S i naziva se deo skupa S. Neka je D dobijeni skup delova od S. Ako je unija svih delova jednaka skupu S, za D ćemo reći da predstavlja jedno pokriće skupa S. Ako su svi delovi od D neprazni, međusobom disjunktni, a unija im je jednaka skupu S, za D kažemo da predstavlja jednu podelu skupa S. Za delove jedne podele kažemo da predstavljaju klase skupa S. Ako se nad elementima skupa S meri obeležje X i ako ih uredimo prema veličini toga obeležja, svaki element imaće svoj rang u tako formiranom redosledu. Ako je vrednost od X jednog elementa i-ta po veličini u skupu S, element ćemo označiti sa e_i, a njegovu vrednost od X sa x_i. Pri tome je $\forall_i \left[i \in \{1,\ldots,n-1\} \Rightarrow x_i \geq x_{i+1} \right]$. Na ovaj način možemo formirati klasifikacionu listu (rang-listu, redoslednu listu) elemenata skupa S u odnosu na vrednosti obeležja X (Ivanović, 1977). Različiti načini formiranja klasifikacione liste tj. rang liste predstavljaju metode kako se sam proces rangiranja može izvesti. Različite metode rangiranja u statistici, proizvod su razvoja same statističke misli. Ove metode se zasnivaju na rezultatima statističkih istraživanja, pa se stoga mogu nazvati statističkim metodama rangiranja. Podrazumevaju različite pristupe rangiranju raznovrsnih pojava i objekata posmatranja, gde se sam proces rangiranja vrši korišćenjem statističkih analiza kao bi se dobio „optimalan“ rezultat rangiranja (Radojičić, 2007).

U disertaciji, pažnja će biti usmerena na jednu statističku metodu rangiranja, koja će biti zasnovana na primeni metode I-odstojanja u čiji će se obrazac uključiti obeležje dobijeno izračunavanjem vektorskog koeficijenta korelacije. Takođe, u okviru disertacije će biti prikazan jedan algoritam za rešavanje problema grupisanja. Tako definisani algoritam se može iskoristiti za rešavanje jedne određene grupe problema, u kojima su unapred definisana neka ograničenja u vidu „dozvoljenih“ i „zabranjenih“ veza određenih entiteta, gde se onda trebaju entiteti pod takvim uslovima grupisati u određen broj grupa.

1.1. Polazne hipoteze

I. Osnovna hipoteza je da testiranje hipoteze o nezavisnosti vektorskog koeficijenta korelacije, može biti osnova za definisanje statistike za testiranje hipoteze o jednakosti dva vektorska koeficijenta korelacije..

II. Na osnovu vektorskog koeficijenta korelacije uz primenu metode Ivanovićevog odstojanja može se definisati jedan novi način rangiranja.

III. Kao nova ideja apriornog grupisanja biće prikazan algoritam za rešavanje klasifikacije sa unapred definisanim ograničenjima za posebnu grupu problema.

1.2. Metode istraživanja

Osnovni metod istraživanja u disertaciji je sakupljanje i proučavanje dostupne literature, njena analiza i sistematizacija, a sve to s ciljem da se pokaže opravdanost i korisnost definisanja funkcije (statistike), kojom će se porediti jednakost dva vektorska koeficijenta korelacije. Rad će se zasnivati na primeni:

- Metoda za analizu podataka (deskriptivne mere, mere odstojanja, analiza ekstremnih vrednosti),
- Metode i tehnike eksploratorne analize podataka,
- Metoda statističke analize (korelaciona analiza, parametarski i neparametarski testovi, regresiona analiza),
- Multivarijacione statističke analize (kanonička korelaciona analiza, faktorska analiza, analiza glavnih komponenata, analiza grupisanja),
- Algoritamskih struktura

1.3. Doprinos doktorske disertacije

Glavni doprinos se ogleda u definisanju i primeni test statistike za poređenje dva vektorska koeficijenta korelacije. Na ovaj način je moguće utvrditi vezu između izlaznih i ulaznih veličina jednog organizacionog sistema, ali i utvrditi i izmeriti razlike između dva organizaciona sistema, tako da se može uvideti stepen značajnosti sličnosti ili razlike između posmatranih organizacionih sistema. U disertaciji je dat teorijski prikaz postojećih test statistika za poređenje dve korelacione strukture, kao i niz eksperimentalnih rezultata koji verifikuju usvojeni koncept. Za izračunavanje vektorskog koeficijenta korelacije, test statistike i kritične oblasti testa napisan je program u Matrix programskom jeziku za SPSS (v.21) programski paket, sa ciljem da bude pristupačan širem krugu korisnika.

Drugi doprinos se odnosi na problem određivanja raspodele I-odstojanja. Određena je raspodela kvadratne forme normalno raspoređenih vektora, pa je njenom primenom na kvadratno I-odstojanje pokazano da ima normalnu raspodelu. Takođe, u opštem slučaju, a za šta je korišćena *Bootstrap* metoda, koja kao svoj sastavni deo podrazumeva primenu Monte-Karlo simulacije, pokazano (osim u izuzetnim slučajevima) je slaganje I-odstojanja sa teoretskom normalnom raspodelom. Time je Ivanovićevo odstojanje dobilo novu dimenziju posmatranja i značajno je unapređen kvalitet za njegovu primenu.

Treći doprinos se odnosi na definisanje jednog algoritma za probleme rangiranja. Rangiranje se zasniva na I-odstojanju, ali je uzet u obzir odnos izlaznih i ulaznih veličina izražen kroz vektorski koeficijent korelacije. Ovime je pokazano da se vektorski koeficijent korelacije može koristiti kao težinski faktor u procesu rangiranja i time omogućiti bolje tj.“realnije“ proces rangiranja i same rezultate.

Četvrti doprinos je dat kroz kreiranje jednog a priori načina grupisanja sa unapred definisanim ograničenjima, kao jedna modifikacija McQeen-ovog *K-mean* algoritma nehijerarhijskog grupisanja.

2. KORELACIONA ANALIZA

Francis Galton (1822.-1911.), poznati matematičar je rekao da je cilj statističke nauke da otkrije metode pretvaranja informacija u vezi velikih grupa koje odlikuju slične činjenice u kratak i sažet izraz pogodan za diskusiju. Jedna, pre svega istorijska motivacija za područje statistike je da objasni značenje podataka u"kratkom i sažetom izrazu." Jedna je stvar ako gledamo u neku tablicu sa brojevima i tvrdimo da tu vidimo neki smisao, ali sasvim je druga stvar da pokažemo da ta tablica predstavlja u stvari dokaz za određeni zaključak. Galton je začetnik priče o korelaciji, a neki od prvih radova na ovu temu su: "Regression towards mediocrity in hereditary stature" (1885), "Family likeness in stature" (1886), and "Co-relations and their measurement, chiefly from anthropometric data." (1888). Ovim radovima je priključen dodatak od strane J.D. Hamilton Diksona, koji je ispitivao korelaciju površina u tri dimenzije.

Galtonova definicija korelacije, koju je predstavio kraljevskom društvu u Londonu, nam otkriva svojstva koeficijenta korelacije. To je mera jačine linearne veze, ako je bliža 1, onda se dve usko povezane promenljive mogu predvideti jedna na osnovu druge korišćenjem linearne jednačine. To je mera pravca: pozitivna korelacija ukazuje da se promenljive X i Y povećavaju ili smanjuju zajedno, a negativna korelacija ukazuje da dok jedna promenljiva opada, druga raste.

Primećuje se da Galton ne tvrdi da korelacija podrazumeva uzročno-posledičnu vezu (bilo bi apsurdno da je veličina jednog organa određena veličinom drugog) i sa tim u vezi on spekulira da korelacija ukazuje na prisustvo najčešćih uzroka za posmatrani odnos između organa (npr. veličina svakog organa).

Za bivarijantnu normalnu raspodelu Galton je izračunao koeficijent korelacije. Njegov metod zahtevao je da se statistički nacrtaju tačke svih podataka izmerenih u Q jedinicama, iscrta linija koja im najbolje odgovara, a zatim izračuna nagib te linije. Mogućnost greške Q, je u stvari preteča moderne standardne devijacije. Polovina posmatranih vrednosti je upala u interval (srednja vrednost - Q, srednja vrednost + Q). Dakle, za normalnu raspodelu važi, da kada je srednja vrednost jednaka medijani, Q je jedna polovina savremenog interkvartilnog domena ili je Q = 0.6745 (standardna devijacija). Galton nije koristio posebnu tehniku za crtanje ove linije, niti neku posebnu formulu računanja.

Ovaj metod, iako neprecizan po savremenim standardima, usvojen je od strane drugih naučnika koji su bili zaintresovani za nove oblasti biometrije. Profesor V.F.R Veldon je u svom radu iz 1892. godine "Certain correlated variations in Crangon vulgaris", primenio Galtonov metod na merenje fizičkih karakteristika škampi. Iako je kasnije umro mlad od upale pluća, Veldon je za života postao suosnivač biometrije sa Karl Personom u 1901. godini. Galton je 1888. godine zaključio svoj rad s komentarom o korisnosti koeficijent korelacije.

Posebnu pažnju Galton posvećuje koeficijentu korelacije r, jer bi to moglo da se koristi za predviđanje odstupanja promenljive y od x, ili x od y. Tako je od početka, koeficijent korelacije bio blisko povezan sa linijom regresije. Prvobitno r je predstavljao nagib regresione prave, ali je postojao taj problem što je nagib linije regresije bio delimično funkcija jedinica mere koju je odabrana. Galton doživljava koeficijent korelacije kao manju jedinicu regresije, i prisvaja oznaku r.

Korelacija (lat.*con* = sa, *relatio* = odnos) predstavlja odnos ili međusobnu povezanost između različitih pojava predstavljenih vrednostima dvaju promenljivih. Pri tome ova povezanost znači da je vrednost jedne promenljive moguće sa određenom verovatnoćom predvideti na osnovu saznanja o promenama druge promenljive. Klasični primeri povezanosti su npr. saznanje o uticaju količine padavina na rast žitarica, o povezanosti slane hrane i visokog krvnog pritiska i sl. Promena vrednosti jedne promenljive utiče na promenu vrednosti druge promenljive. Promenljiva koja svojom vrednošću utiče na drugu promenljivu naziva se nezavisna promenljiva. Promenljiva na koju se utiče se naziva zavisna promenljiva. Npr. unošenje vise soli u organizam utiče na porast krvnog pritiska, dok porast krvnog pritiska ne utiče na povećanje unošenja soli u organizam. U ovom primeru unošenje soli u organizam je nezavisna promenljiva, a povećanje krvnog pritiska je zavisna promenljiva. Mogući su slučajevi da dve promenljive istovremeno utiču jedna na drugu, pa su u tom slučaju obe promenljive istovremeno i zavisne i nezavisne.

Na primer, površina P kruga i poluprečnik r su u funkcionalnoj vezi ($P = r^2\pi$), a promenljive veličine koje označavaju visinu i težinu ljudi pokazuju izvesnu korelaciju, dok su brojevi tačaka koji se pojavljuju na dvema bačenim kockama nekorelativne veličine. Skup statističkih metoda kojima se proučavaju uzajamne veze statističkih obeležja i pojava (smer, jačina, oblik) naziva se teorijom korelacije, a osnovni pokazatelji korelacionih veza su jednačine regresije i koeficijent korelacije.

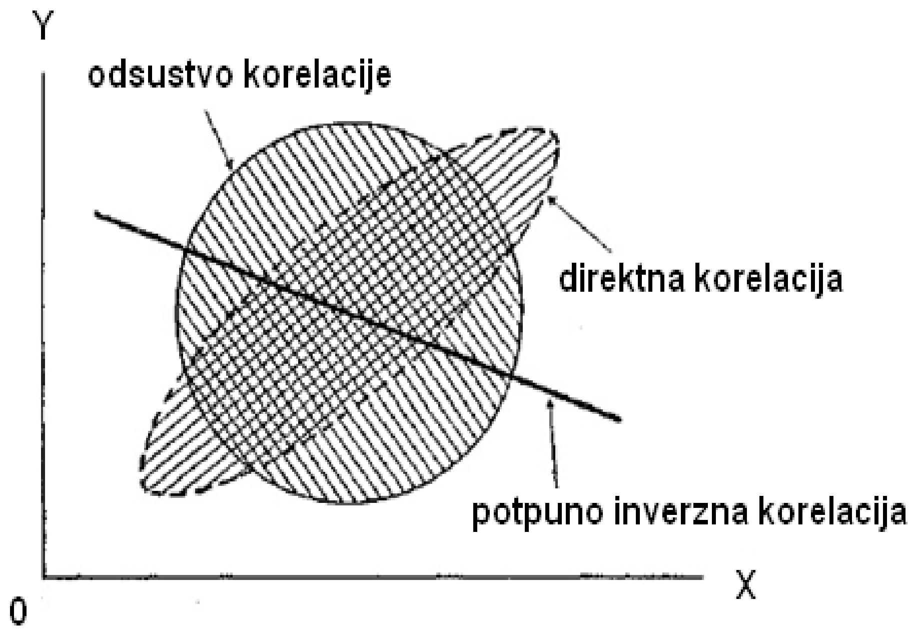
2.1. Prosta linearna korelacija

Kao što je rečeno, zadatak korelacione analize jeste da pokaže samo da li između varijabiliteta posmatranih pojava postoji kvantitativno slaganje - korelaciona veza i, ako postoji, koliki je stepen tog slaganja. Korelaciona veza dve pojave naziva se prostom korelacijom. U klasičnom modelu proste korelacije obe posmatrane pojave su slučajne promenljive, od kojih ni jednu nije nužno identifikovati kao zavisnu, odnosno nezavisnu promenljivu. Tako ćemo, na primer, posmatrajući zaposlenost stanovništva i nacionalni dohodak po stanovniku uočiti da sa porastom zaposlenosti raste i nacionalni dohodak i obrnuto, a ne možemo kategorički reći koja je od ovih pojava nezavisna promenljiva. Veća zaposlenost dovodi do povećanja nacionalnog dohotka, ali s druge strane i veći nacionalni dohodak omogućuje veću zaposlenost. Stope nataliteta i stope mortaliteta stanovništva često pokazuju istu tendenciju porasta ili opadanja, pri čemu je teško označiti jednu od ovih pojava kao nezavisnu promenljivu.

Starost supružnika pokazuje takođe kvantitativno slaganje. Mlađi muškarci stupaju po pravilu u brak s mlađim, a stariji sa starijim ženama. Korelacija između ovih pojava očito postoji, ali se ni jedna od njih ne može unapred smatrati nezavisnom, odnosno zavisnom promenljivom. Zato u ovakvim slučajevima nećemo ispitivati ponašanje jedne pojave u funkciji druge, kao kod regresijske analize, nego će nas zanimati samo mera njihovog međusobnog slaganja, stepen njihove korelacije. To ne isključuje mogućnost primene korelacione analize i na pojave identifikovane kao nezavisne, odnosno zavisne promenljive.

Postojanje kvantitativnog slaganja (korelacione veze) dve pojave otkriva, kao i kod regresione analize, dijagram raspršenosti. Tačke koje na dijagramu raspršenosti pokazuju prostu korelaciju pojava biće locirane na površini čiji se oblik priblizava elipsi. Na slici 2.1 se vidi raspored tačaka na dijagramu raspršenosti u slučaju proste korelacije, potpune korelacije (funkcionalne zavisnosti) i odsustva korelacije. U slučaju potpunog kvantitativnog slaganja varijacija - savršene korelacije, one će, kao i kod regresionog modela, biti na istoj krivoj , a sasvim raspršena kad je korelacija neznatna ili je uopšte nema.

Prema obliku rasporeda tačaka na dijagramu, prosta korelacija može biti linearna ili krivolinijska i, zavisno od smera slaganja, direktna ili inverzna. Kad obe posmatrane pojave pokazuju tendenciju istog smera (obe rastu ili obe opadaju), imaćemo direktnu korelacionu vezu, a kad se njihove promene kreću u suprotnom smeru (jedna raste dok druga opada) inverznu. Merenje jačine korelacione veze i interpretacija dobijenih mera temelji se na istim pretpostavkama na kojima se zasniva i regresiona analiza, s tim što se, kad su obe promenljive slučajne, uvodi još pretpostavka da i svakoj vrednosti Y odgovara normalan raspored vrednosti X, to jest da je njihov zajednički raspored normalan.



Slika 2.1 Dijagram raspršenosti

Kao mera jačine proste linearne korelacione veze koriste se kovarijanse i koeficijent proste linearne korelacije. Definicije ovih mera i postupak njihovog ocenjivanja na temelju podataka uzorka može se objasniti na primeru stopa nataliteta i mortaliteta. Kao što smo videli, kovarijansu linearne korelacije definišemo po sledećoj formuli:

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

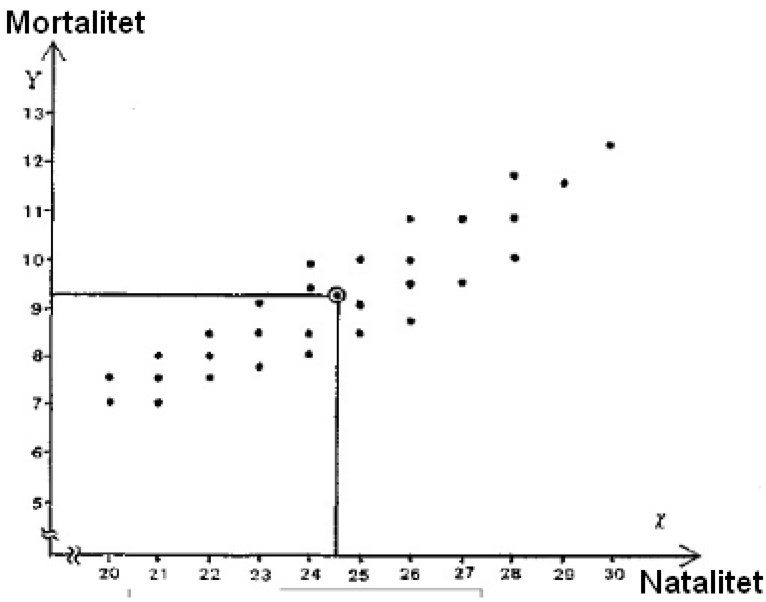
Označimo jednu od ovih pojava, stopu nataliteta na primer, sa X, a drugu stopu, koja predstavlja stopu mortaliteta označićemo sa Y. Na prvoj slici ispod imamo Tabelu 2.1 koja nam govori o broju nataliteta i mortaliteta u određenom periodu, a u poslednje tri kolone izračunate su vrednosti potrebne za izračunavanje ocene kovarijanse. Na slici 2.2 prikazan je dijagram raspršenosti za stope nataliteta i mortaliteta, a posebno je svojim koordinatama obeležena tačka koja predstavlja aritmetičke sredine posmatranih pojava x = 24.4, y = 9.2. Ta tačka se naziva centroidom podataka, a prosek proizvoda odstupanja pojedinih vrednosti X i Y od te tačke kovarijansom. Ocena kovarijanse skupa na temelju podataka uzorka dobija se po formuli:

$$S_{xy} = \sigma_{xy} = \frac{\sum\left[(X_i - \overline{X})(Y_i - \overline{Y})\right]}{n-1}$$

gde su n - veličina uzorka, a n - 1 broj stepeni slobode.

Tabela 2.1 Vrednosti za natalitet i mortalitet

Наталитет X	Морталитет Y	XY	X2	Y2
20	7,0	140	400	49
20	7,5	150	400	56,25
21	7,0	147	441	49
21	7,5	157,5	441	56,25
21	8,0	168	441	64
22	8,0	176	484	64
22	7,5	165	484	56,25
22	8,0	176	484	64
22	9,0	198	484	81
23	8,5	195,5	529	72,25
23	7,8	179,4	529	60,84
23	9,2	211,6	529	84,64
24	8,0	192	576	64
24	8,5	204	576	72,25
24	9,5	228	576	90,25
24	10,0	240	576	100
25	8,5	212,5	625	72,25
25	9,0	225	625	81
25	10,0	250	625	100
26	8,7	226,2	676	75,69
26	9,5	247	676	90,25
26	10,0	260	676	100
26	11,0	286	676	121
26	9,5	247	676	90,25
27	11,0	297	729	121
28	10,0	280	784	100
28	11,0	308	784	121
28	12,0	336	784	144
29	11,8	342,2	841	139,24
30	12,5	375	900	156,25



Slika 2.2 Dijagram raspršenosti za primer nataliteta i mortaliteta

Kovarijansa zavisi u velikoj meri od nivoa vrednosti posmatranih pojava. Tako bi, na primer, podaci o stopama nataliteta i mortaliteta pomnoženi sa 10 dali kovarijansu deset puta veću, iako bi relacije njihovih odstupanja od centroida ostale nepromenjene. Ostao bi nepromenjen i stepen njihovog međusobnog slaganja. Zato kovarijansa nije pogodna za poređenje, što ograničava njenu primenu kao mere korelacije (Brown, et al. 1977).

2.1.1. Ocena koeficijenata proste linearne korelacije

Za sagledavanje veze između dve linearno međuzavisne promenjive najčešće se upotrebljava koeficijent korelacije. U korelacionoj analizi se po pravilu koristi relativna mera korelacije, zasnovana na standardizovanim odstupanjima od centroida (odstupanjima iskazanim u standardnim devijacijama), koja se naziva Pirsonovim koeficijentom proste linearne korelacije ili samo koeficijentom proste linearne korelacije.

Pirsonov koeficijent korelacije koristi se u slučajevima kada između promenljivih posmatranog modela postoji linearna povezanost i neprekidna normalna distribucija. Vrednost Pirsonovog koeficijenta korelacije kreće se od +1 (savršena pozitivna korelacija) do -1 (savršena negativna korelacija). Predznak koeficijenta nam ukazuje na smer korelacije - da li je pozitivna ili negativna, ali ne govori o tome kolika je jačina te korelacije. Ovako definisan koeficijent korelacije bazira se na upoređivanju stvarnog uticaja posmatranih promenljivih, jedne na drugu u odnosu na maksimalni mogući uticaj dve promenljive. Za izračunavanje koeficijenta korelacije potrebne su tri različite sume kvadrata: suma kvadrata promenljive X, suma kvadrata promenljive Y i suma množilaca promenljivih X i Y. Ocenu ovog koeficijenta u skupu, ρ_{xy}, predstavlja koeficijent proste linearne korelacije uzorka r_{xy}:

$$r_{xy} = \rho_{xy} = \frac{S_{xy}}{S_x S_y},$$

gde S_{xy} predstavlja ocenu kovarijanse a,

$$S_x = \sqrt{\frac{\sum \left(X_i - \overline{X}\right)^2}{n-1}} \quad \text{i} \quad S_Y = \sqrt{\frac{\sum \left(Y_i - Y\right)^2}{n-1}}$$

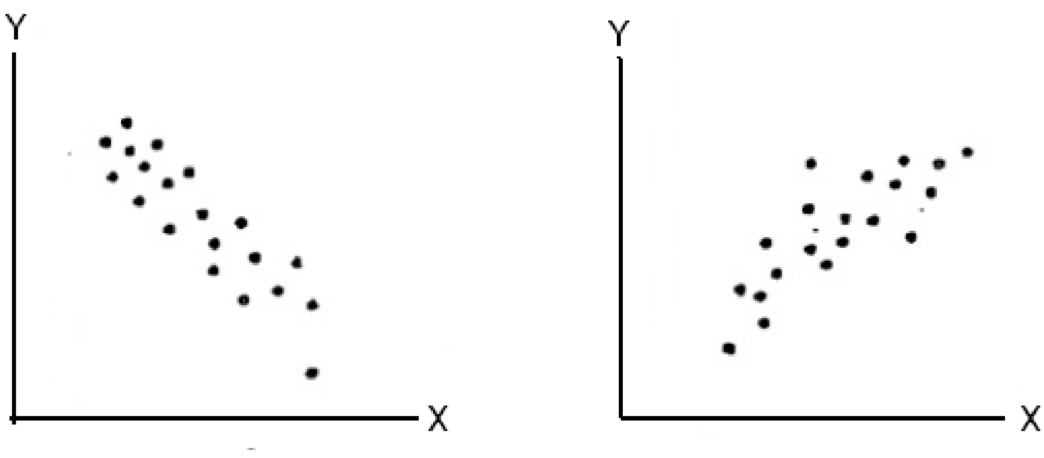
ocene standardnih devijacija promenljivih X i Y respektivno.

Jednostavnije, koeficijent proste linearne korelacije uzorka, koji predstavlja ocenu koeficijenta korelacije skupa, izračunava se (sa uprošćenom simbolikom) po formuli:

$$r_{xy} = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - \left(\sum x\right)^2}\sqrt{n\sum y^2 - \left(\sum y\right)^2}}$$

Ova formula predstavlja samo drugi oblik formule za ocenu kovarijanse. Obe su simetrične u odnosu na promenjive X i Y, pa rezultat ne zavisi od toga koju smo od posmatranih pojava obelezili sa X, a koju sa Y, odnosno koju tretiramo kao nezavisnu, a koju kao zavisnu promenljivu. Obe formule, dakle, vode računa o reverzibilnosti relacije između promenljivih X i Y. Treba primetiti da je r_{xy} pristrsana ocena, izuzev kada je r_{xy} = 0. Ta pristranost je za velike uzorke zanemarljiva, pa se za korelacionu analizu preporučuju i koriste veći uzorci.

Koeficijent proste linearne korelacije varira od -1 do 1 i označava sve čvršću korelacionu vezu - što je blize jedinici, i to od -1 do 0 negativnu ili inverznu (kad jedna pojava raste druga opada ili obrnuto), a od 0 do 1 pozitivnu ili direktnu (kad obe pojave rastu ili opadaju).



Slika 2.3 Inverzna i direktna veza

Ako između posmatranih pojava postoji potpuna (savršena) korelaciona veza, koeficijent korelacije iznosi -1 ili 1, a kada između njih uopšte nema linearne korelacije, koeficijent korelacije se izjednačuje sa nulom. Nula, dakle, ne znači odsustvo bilo kakve korelacione veze, nego odsustvo linearne korelacione veze.

Često se izračunava koeficijent determinacije koji predstavlja koeficijent korelacije dignut na kvadrat. Putem njega se sagledava koliki je udeo promenljive X u ukupnom varijabilitetu promenljive Y. Tako, na primer, ako je r = 0.9, onda je r^2 = 0.81. Znači od 100%, uzet kao ukupan varijabilitet promenljive Y, 81% se objašnjava vezom sa promenljivom X. U prethodnom primeru koji smo razmatrali izračunati koeficijent korelacije je $r_{XY} = 0.888$ i ova vrednost nam ukazuje na visok stepen pozitivne linearne korelacione veze između stope nataliteta i stope mortaliteta u posmatranom skupu.

Pored Pirsonovog koeficijenta korelacije često se koristi i Spermanov koeficijent korelacije. Spermanov koeficijent korelacije (produkt rang korelacije) koristi se za merenje povezanosti između varijabli u slučajevima kada nije moguće primeniti Pirsonov koeficijent korelacije. Bazira se na tome da se izmeri doslednost povezanosti između poređanih varijabli, a oblik povezanosti (npr. linearni oblik koji je preduslov za korišćenje Pirsonovog koeficijenta) nije bitan. Slučajevi u kojima se koristi Spermanov koficijent su na primer, kada među varijablama ne postoji linearna povezanost, a nije moguće primeniti odgovarajuću transformaciju kojom bi se povezanost prevela u linearnu (npr. veza između seizmičkog atributa i bušotinskih podataka u naftnoj geologiji). Spermanov koeficijent korelacije kao rezultat daje približnu vrednost koeficijenta korelacije koji se tretira kao njegova dovoljno dobra aproksimacija. Prilikom korišćenja Spermanovog koeficijenta, vrednosti promenljivih potrebno je rangirati i na takav način svesti na zajedničku meru. Najjednostavniji način rangiranja je da se najmanjoj vrednosti svake promenljive dodeli rang 1, sledećoj po veličini rang 2 i tako sve do poslednje kojoj se dodeljuje maksimalni rang. Izračunavanje koeficijenta radi se korišćenjem vrednosti dodeljenih rangova. Spermanov koeficijent označavaćemo sa r_S.

Formula za izračunavanje Spermanovog koeficijenta korelacije je:

$$r_s = 1 - 6\sum_{i=1}^{n} \frac{d_i^{\,2}}{n\left(n^2 - 1\right)}$$

gde je d razlika vrednosti rangova dve posmatrane varijable, a n je broj različitih serija.

2.1.2 Interpretacija koeficijenata proste linearne korelacije

Interpretacija koeficijenta proste linearne korelacije zahteva dodatnu analizu, izvesna objašnjenja i ograničenja, posebno u pogledu kauzalne veze posmatranih pojava. Potpuno odsustvo korelacione veze ne pričinjava velike teškoće, jer sasvim jasno pokazuje da između posmatranih pojava ne postoji ni uzročna veza. Treba samo imati u vidu da koeficijent proste korelacije, čija je vrednost ravna ili približna nuli označava odsustvo onog oblika korelacione veze koji je u davnom slučaju proučavan. Ako je, na primer, vrednost koeficijenta linearne korelacije nula, to znaci da između posmatranih pojava ne postoji linearna korelaciona veza, ali između tih pojava može postojati neki drugi - krivolinijski oblik korelacione veze. Postojanje korelacione veze, međutim, zahteva veću pažnju istrazivača. Koeficijenti korelacije sa vrednostima od 0 do 0.5, odnosno -0.5, iako označavaju postojanje korelacione veze, smatraju se indikatorima slabe veze. Neki autori smatraju čak da vrednosti koeficijenta proste korelacije sve do 0.7 nemaju neki veći analitički značaj, a da tek preko 0.7 označavaju dovoljno čvrstu korelacionu vezu, utoliko čvršću ukoliko su bliže jedinici. Značaj vrednosti ovih mera, pri tom, varira i zavisno od prirode posmatranih pojava. Ako dve pojave po svojoj prirodi pokazuju visok stepen korelacione veze, onda i relativno veliki iznos koeficijenta korelacije, na primer od 0.8, može u posmatranom slučaju imati mali značaj, i obrnuto. Ali korelaciona veza, bez obzira na vrednosti koeficijenta korelacije, odnosno stepen kvantitativnog slaganja posmatranih pojava, sama po sebi ne predstavlja uzročnu vezu između pojava, mada visok stepen korelacione veze indicira uzročnu vezu.

Da bi se ustanovilo da li pojave koje stoje u korelacionoj vezi pokazuju i međuzavisnost, potrebno je izvršiti detaljniju kvalitativnu i kvantitativnu analizu, ne samo posmatranih nego i drugih relevantnih pojava. Potrebno je zatim objasniti vezu koju korelacioni model predstavlja, jer slaganje varijacija dveju varijabli može imati više uzroka. Obe posmatrane pojave, pre svega, mogu pokazivati istu tendenciju kvantitativnog varijabiliteta zato što na njih utiču isti faktori pod čijim uticajem njihove promene pokazuju veći ili manji stepen slaganja. Ali one mogu pokazivati visok stepen korelacione veze i zbog toga što jedna na drugu utiču, što između njih postoji određena interakcija. Jedna od posmatranih pojava može biti faktor koji sam ili zajedno s ostalim činjenicima utiče na drugu posmatranu pojavu. Ova druga je u tom slučaju zavisna promenljiva. Takođe, dve posmatrane pojave mogu pokazivati i izvestan stepen kvantitativnog slaganja, a da između njih ne postoji nikakva kauzalna veza.

S druge strane, dve pojave mogu biti u tesnoj međusobnoj vezi iako prost koeficijent linearne korelacije ne pokazuje visok stepen slaganja. To se javlja kod vremenskih serija, čije varijacije pokazuju kvantitativno slaganje posle određenog vremenskog razmaka – sa zaostajanjem. Tako, na primer, investicijska ulaganja u jednom periodu dovode do povećane proizvodnje tek nakon određenog vremenskog intervala. Zato se kod istraživanja ovakvih pojava mora izvršiti odgovarajuće pomeranje podataka vremenskih serija.

Koeficijent proste linearne korelacije pokazuje samo stepen kvantitativnog slaganja dveju pojava. U praksi, a pogotovo u prirodnim i društvenim zbivanjima odnosi među pojavama nisu tako jednostavni (bilateralni). Na ponašanje jedne pojave utiče najčešće mnoštvo faktora, a ne samo jedan, a ono opet moze imati veće ili manje povratno dejstvo na varijacije pojedinih faktora. Ti odnosi mogu biti vrlo složeni. Metode višestruke i delimične korelacione analize, pružaju u tom pogledu šire mogućnosti, ali se i oni moraju dopunjavati i kombinovati metodama kvalitativne analize.

Korelaciona, kao i regresiona analiza, koja se zasniva na većem broju posmatranja daje jasniju i pouzdaniju sliku kvantitativnih odnosa posmatranih pojava. Kad je broj posmatranja mali, mogu se dobiti sasvim pogrešni rezultati ako se o tome ne vodi računa. Na mere korelacije, sem toga, utiču svi empirijski podaci, što znači i oni čije su vrednosti ekstremne, pa one mogu deformisati rezultat korelacione analize utoliko više ukoliko je broj posmatranja manji. Treba takođe imati u vidu da se rezultati korelacione analize odnose, pre svega, na posmatrani interval vrednosti X i Y. Njihove vrednosti izvan tog intervala mogu pokazivati veći ili manji stepen kvantitativnog slaganja, odnosno veću ili manju korelacionu vezu.

Prilikom interpretacije regresione analize ne treba gledati samo regresione koeficijente, nego treba izračunati i varijansu procene kriterijumske varijable, koja se obično označava kao $R = r^2$ i naziva se koeficijentom determinacije. Koeficijent determinacije r^2 može se predstaviti kao deo varijanse, koji se može objasniti posmatranim prediktorskim sistemom. Može se smatrati merom efikasnosti regresije, odnosno uspešnosti prognoze. Signifikanost koeficijenta determinacije r^2 testira se Fišerovim testom. Vrednost r^2 varira od 0 do 1, odnosno, ako ga izražavamo u procentima, od 0 do 100 %.

2.1.3. Korelaciona matrica, višestruka i parcijalna korelacija

Ponekad nam u istraživanju nije dovoljna informacija o korelaciji dve posmatrane varijable, već nas zanima na koji način više varijabli međusobno utiče jedna na drugu. Nakon što se posmatranjem međusobnog odnosa svih parova dveju varijabli utvrdi njihova međusobna korelacija, izrađuje se korelaciona matrica. Redovi i kolone matrice predstavljaju posmatrane varijable, a podatak na preseku određenog reda i kolone predstavlja koeficijent korelacije između varijabli u odgovarajućem redu i koloni.

Matrica na dijagonali ima vrednost 1 (pošto je svaka varijabla sama sa sobom u potpunoj korelaciji). Dobijena matrica je simetrična - podaci iznad i ispod dijagonale za isti par varijabli su identični. Zbog tih svojstava matrica je redundantna i dovoljno je posmatrati jedan njen deo, iznad dijagonale ili ispod dijagonale. Vizuelno možemo utvrditi u kojoj meri su dve pojedinačne varijable u korelaciji, koje varijable u međusobnom odnosu imaju najveći ili najmanji koeficijent korelacije, te koji se skupovi varijabli ističu sličnim koeficijentima. Vizuelno ne možemo utvrditi na koji način i u kolikoj meri više varijabli zajednički utiče na drugu pojedinačnu varijablu.

Višestruka korelacija je analitička procedura kojom se utvrđuje na koji način više nezavisnih varijabli utiče na jednu zavisnu varijablu. Koeficijent višestruke korelacije označava se velikim slovom R. Za računanje koeficijenta višestruke korelacije potrebno je prvo izračunati koeficijente korelacije između svakog para varijabli koje posmatramo. Odnos koeficijenata korelacije varijabli može se prikazati korelacionom matricom. Dobijene koeficijente potrebno je uvrstiti u formulu za izračunavanje višestruke korelacije. Podaci višestruke korelacije kod koje se posmatra međusobni uticaj tri varijable može se prikazati trodimenzionalnim dijagramom raspršenosti – scatter diagramom.

Formula za izračunavanje višestruke korelacije kada posmatramo uticaj dve nezavisne varijable na treću, zavisnu, je sledeća:

$$R = \frac{\sqrt{r_{YX_1}^2 + r_{YX_2}^2 - 2r_{YX_1} \cdot r_{YX_2}}}{\sqrt{1 - r_{X_1X_2}^2}}$$

Nezavisne varijable čije vrednosti promatramo označene su sa X_1 i X_2, a zavisna varijabla označena je sa Y. Koeficijent višestruke korelacije uzima vrednosti od 0 do +1, i u njegovoj interpretaciji primenjuju se ista pravila kao kod interpretiranja koeficijenta jednostavne korelacije (Quade, 1974). Kako bi račun višestruke korelacije bio što precizniji, potrebno je koristiti veći uzorak sa više vrednosti varijabli nego u slučaju računanja koeficijenata kod jednostavne korelacije.

Za istrazivanje korelacione veze više posmatranih pojava koriste se koeficijent višestruke korelacije i koeficijenti parcijalne ili delimične korelacije. Koeficijent višestruke linearne korelacije je relativna mera koja pokazuje stepen linearnog slaganja varijacija jedne zavisne - Y, i više nezavisnih varijabli

$X_1, X_2, \ldots, X_{k-1}$, uz pretpostavku da su sve slučajne promenljive. Pri tome, nezavisne promenljive mogu (kao i kod regresione analize) uzimati i fiksirane vrednosti kad ocenjujemo koeficijent višestruke korelacije samo u vidu jedne brojčane vrednosti (tačke). Određivanje intervala pouzdanosti ocene iziskuje ispunjenost i ove pretpostavke.

Koeficijent višestruke korelacije, označimo ga sa

$$\rho_{y, x_1, x_2, \ldots\ldots, x_{k-1}}$$

varira od 0 do 1, što znači uvek je pozitivan i za razliku od koeficijenta proste linearne korelacije, ne pokazuje smer slaganja varijacija posmatranih pojava. Što je bliže jedinici, stepen linearnog slaganja je veći, sto je bliže nuli stepen slaganja je manji. Za ocenu ovog koeficijenta koristi se koeficijent višestruke linearne korelacije uzorka, koji predstavlja kvadratni koren koeficijenta višestruke determinacije. S toga će ocena koeficijenta višestruke linearne korelacije biti:

$$\rho_{y, x_1, x_2, \ldots\ldots, x_{k-1}} = \sqrt{1 - \frac{S_e^{\,2}}{S_y^{\,2}}}$$

gdje su $S_e^{\,2}$ - rezidalna varijansa, a $S_y^{\,2}$ - ocena varijanse zavisne promenljive.

Koeficijent višestruke linearne korelacije, međutim, ne daje nikakvu informaciju o relativnom značaju nezavisnih varijabli. U mnogim slučajevima visok koeficijent višestruke korelacije javlja se pod jakim uticajem samo jedne ili dveju nezavisnih promenljivih, dok je uticaj ostalih mali ili beznačajan. Zato se nakon analize višestruke korelacije pažnja prenosi na istraživanje relativnog značaja nezavisnih varijabli.

Relativan značaj nezavisnih promenljivih meri se koeficijentima parcijalne ili delimične linearne korelacije. Svaki od njih pokazuje stepen slaganja zavisne promenljive i jedne nezavisne promenljive, pri čemu je uticaj ostalih nezavisnih isključen ili konstantan. Zbog toga se ovi koeficijenti nazivaju i ''neto'' koeficijentima. Njihove ocene predstavljaju odgovarajući koeficijenti parcijalne korelacije izračunati na osnovu podataka iz uzorka. Zadržimo se na oceni koeficijenta parcijalne korelacije sa isključenjem samo jedne nezavisne promenljive, koja se dobija po formuli:

$$r_{yx_i x_j} = \frac{r_{yx_i} - r_{yx_j} r_{yx_i x_j}}{\sqrt{\left(1 - r_{yx_j}{}^2\right)}\sqrt{\left(1 - r_{x_i x_j}{}^2\right)}},$$

gde su

$$r_{yx_i}, r_{yx_j}, r_{yx_i x_j}$$

prosti koeficijenti linearne korelacije odgovarajućih promenljivih.

Kvadrati koeficijanata parcijalne korelacije predstavljaju koeficijente parcijalne determinacije. Prilikom utvrđivanja korelacije dveju promenljivih, vrlo je važno na ispravan način izabrati promenljive koje se posmatraju. Vrednosti promenljivih bi trebale biti izabrane iz slučajnog skupa. Što je veći broj varijabli koje se posmatraju, to će rezultati biti precizniji. Povećanje broja posmatranih vrednosti varijabli može u velikoj meri promeniti rezultate izračunavanja.

2.1.4. Poređenje Pirsonovog i Spirmanovog koeficijenta korelacije

Tri najpopularnija koeficijenta korelacije su: Pirsonov koeficijent (r), Spermanov koeficijent – ρ (''ro''), i Kendalov koeficijent – τ (''tau''). Kendalov koeficijent je od strane Kendala uveden 1938. godine. Ovaj koeficijent korelacije se lako može koristiti kao alternativa za Spermanov ρ-koeficijent, za podatke prikazane u obliku redova. To je jednostavna funkcija koja odražava minimalni broj susednih razmena potrebnih za nastanak jednog reda podataka iz drugog.

Njegova svojstva Kendal analizira u svojoj knjizi u kojoj govori o metodama i stepenu korelacije, objavljenoj prvi put 1948. godine. Kao što navodi, "Koeficijent koji smo uveli daje nam neku vrstu prosečne mere sporazuma između parova nekih članova (" sporazum ", to jest, respektivno prema redu), a time je evidentno preporučljiv kao mera saglasnosti između dve lestvice''. U principu, ρ je lakši koeficijent za izračunavanje od τ. Videćemo ... da je u većini teorijskih tvrđenja poželjno da τ bude ρ... ''.

Glavna prednost korišćenja Kendalovog koeficijenta su činjenica da njegova distribucija ima nešto bolja statistička svojstva, te da postoji direktno tumačenje ovih svojstava u skladu sa mogućnostima posmatranja saglasnosti i nesaglasnosti tih parova. Ipak, koeficijent τ nije tako često upotrebljavan u prošlosti (u zadnjih šezdesetak godina), kao što je bio Spermanov koeficijent u merenju stepena korelacije, uglavnom zbog toga što je bio težak za izračunavanje. Danas izračunavanje Kendalovog τ-koeficijenta ne predstavlja nikakav problem. Kendalov τ-koeficijent je ekvivalent Spermanovom koeficijentu u smislu osnovnih pretpostavki, ali oni nisu identični u veličini, budući da su njihova osnovna logika i formule za izračunavanje sasvim drugačije. Odnos između ova dva koeficijenta sa velikim brojem parova je dao naučnik Danijels (1944):

$$-1 \leq 3\tau - 2\rho \leq 1$$

U većini slučajeva, ove vrednosti su vrlo blizu i uvek će nas dovoditi do istih zaključaka, ali kad se odstupanje dogodi, onda je najsigurnije uzeti nižu od ovuh vrednosti. Još važnije, Kendalov τ i Sperman r_s koeficijent podrazumevaju različite interpretacije. Spermanov r_s se smatra sličnim Pirsonovim r-koeficijentom korelacije, u slučaju udela varijabilnosti koja se računa, a Kendalov τ-koeficijent predstavlja verovatnoću, odnosno, razliku između verovatnoća da su posmatrani podaci u istom redosledu naspram verovatnoće da posmatrani podaci nisu u istom redosledu. Osobine i upoređivanje Kendalovog i Spearmanovog koeficijenta su analizirani od strane mnogih naučnika, a čak su i danas još uvek analiziraju.

Imajući u vidu prethodno navedeno, mi ćemo Spermanov koeficijent tretirati kao odgovarajuću prezentaciju mere za izračunavanje stepena korelacije.

Interesantno je uporediti vrednosti Pirsonovog koeficijenta korelacije koristeći podatke na kvantitativni način, u odnosu na vrednosti Spermanovog koeficijenta korelacije koristeći iste podatke na pomalo ''kvalitativan'' način. Pirsonov koeficijent korelacije je inače otkrio naučnik Bravais 1846. godine, ali ga je Karl Pirson prvi opisao 1896. godine, kao i standardne metode za izračunavanje, pokazujući da je to najbolji mogući koeficijent korelacije. Pirson je takođe dao neke komentare o proširenju ideje koju je imao Galton (koji ga je primjenjivao na antropometrijskim podacima). On je nazvao ovu metodu "proizvod-trenuci'' metoda (ili Galtonova funkcija za koeficijent korelacije r). Bitna pretpostavka u Pearsonovom radu iz 1896. godine je normalnost analiziranih promenljivih, koja može biti ostvarena samo za kvantitativne promenljive. Pirsonov koeficijent korelacije je mera jačine linearne veze između dve takve promenljive. Zato je 1904. godine Kopljanik usvojio Pirsonov koeficijent korelacije kao meru jačine veze između dve promenljive koje se ne mogu meriti kvantitativno.

Spearmanov koeficijent korelacije (tj. produkt rang korelacije) je neparametarski (distributivno besplatan) rang statistike, predstavljen kao mera jačine povezanosti dveju varijabli. To je mera monotone povezanosti koja se koristi kada se distribucija podataka uz pomoć Pirsonovog koeficijenta korelacije čini nepoželjnom ili obmanjujućom. Spermanov koeficijent nije merilo linearnog odnosa između dveju varijabli, kako se neki "statističari" izjašnjavaju. On procenjuje koliko dobro proizvoljna monotona funkcija može opisati odnos između dve varijable, bez donošenja bilo kakve pretpostavke o učestalosti distribucije varijabli. Za razliku od Pirsonovog ''produkt-trenutak'' koeficijenta korelacije, on ne zahteva pretpostavku da je odnos između varijabli linearna veza, niti zahteva da varijable budu merene na intervalnim skalama; on se može koristiti za varijable merene na nominalnoj skali.

U principu, r_s je jednostavno poseban slučaj Pirsonovog ''produkt-trenutak'' koeficijenta u kojem se podaci prikazuju u redovima pre izračunavanja koeficijenta. Spermanova statistička postignuća iz 1904. godine nisu bila cenjena od strane njegovog kolege sa univerziteta - Karl Persona, ali je takođe postojao i dugogodišnji nesporazum između njih. Istorija i naknadna praksa su pokazali da je Sperman bio u pravu, pa se danas koeficijent r_s naširoko koristi u oblasti statističke analize.

Korišćenje Pirsonovog ''proizvod-trenutak'' koeficijenta korelacije i Spearmanovog ''produkt rang'' koeficijenta korelacije u analiziranju geografskih podataka (na karti podataka koji su u prostornoj korelaciji) je prvi koristio naučnik Hajning, 1991. godine. Sada ćemo uporediti vrednosti i značaj Pirsonovog i Spermanovog koeficijenta korelacije, na istoj grupi podataka (originalni podaci za r i rangirani podaci za r_s). Podaci korišćeni u analizi su podaci Centralnog zavoda za statistiku za odabrane administrativne jedinice različitih nivoa u Poljskoj i predstavljaju regionalne indekse socijalno – ekonomskog razvoja.

X1 - Broj stanovnika prema zvaničnom mestu prebivališta

X2 - Broj telefona na 1000 stanovnika

X3 - Vodosnabdevanje: količina vode isporučene po domaćinstvima

X4 - Gustina naseljenosti po 1 kvadratnom kilometru

X5 - Oranice u okviru administrativnih granica

X6 - Površina komune u kilometrima kvadratnim

X7 – Broj zaposlenih stanovnika prema uzrastu - (18-64 za muškarce, za žene 18-59)

X8 - Stalna stopa migracija stanovništva na 1000 stanovnika

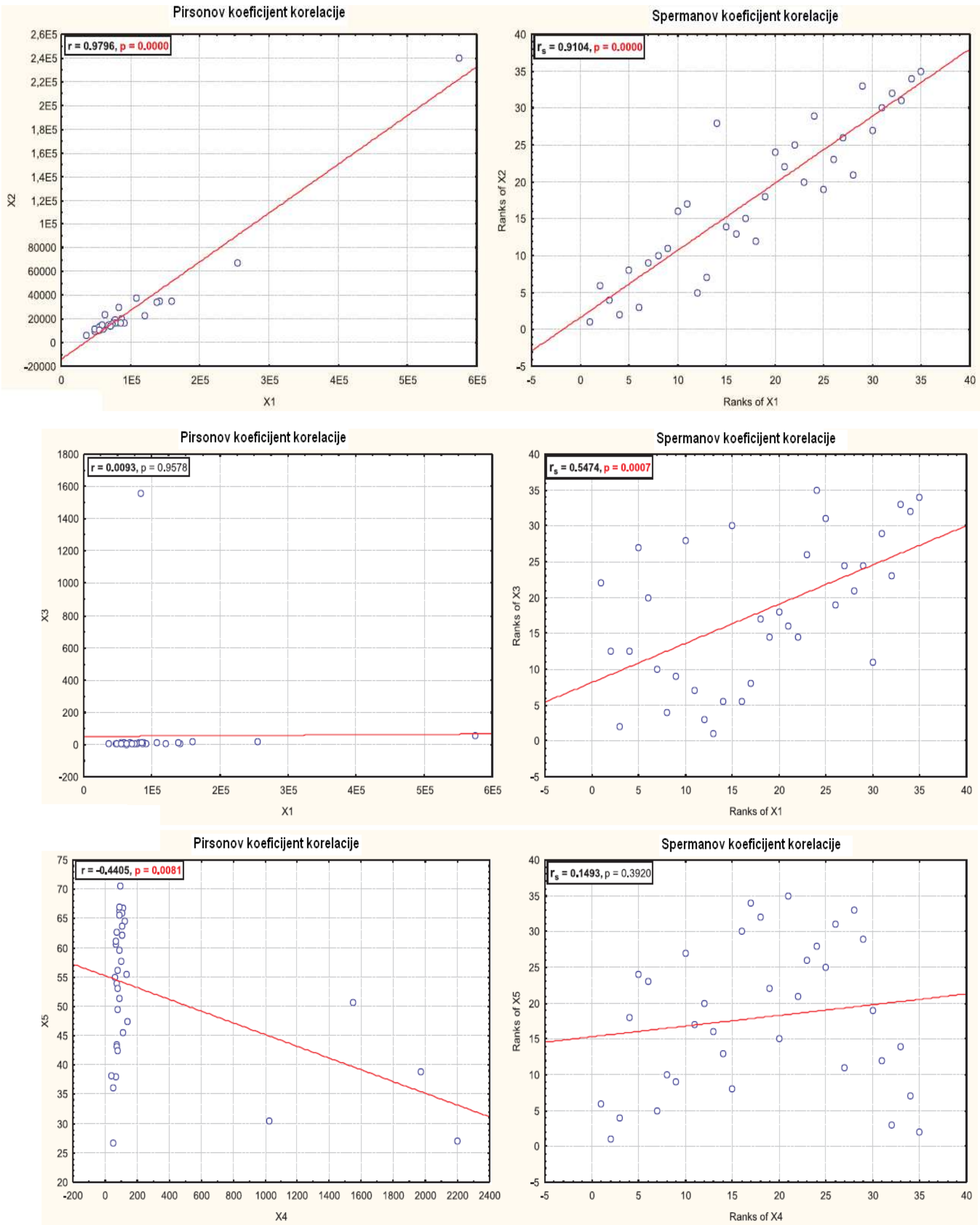
X9 - Industrijska zaposlenost na 1000 radnika

X10 – Broj rođenih na 1000 stanovnika

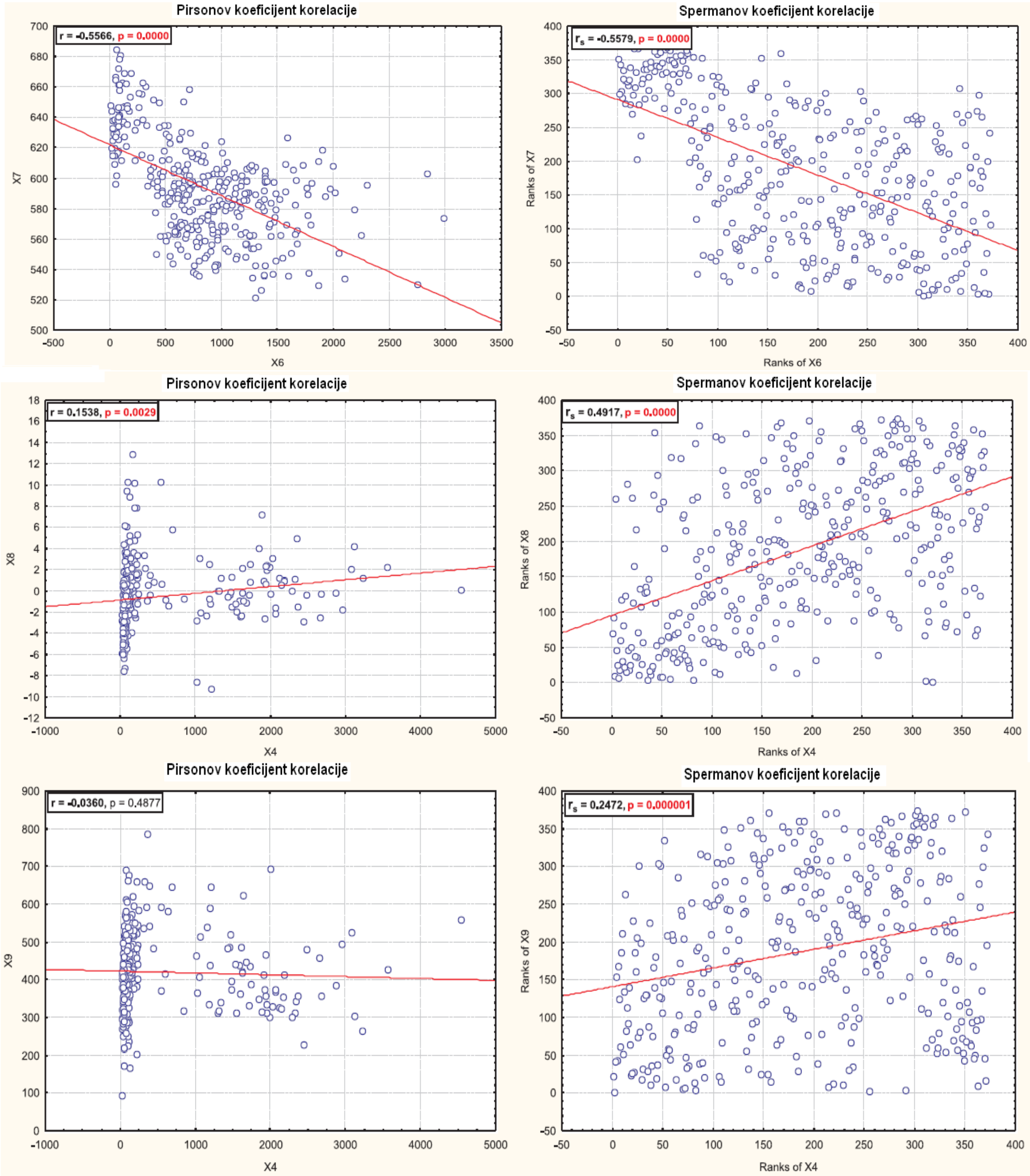
X11 - Potrošnja vode u nacionalnoj ekonomiji

X12 - Natalitet u promilima (ukupno)

Prethodni podaci su korišćeni za izračunavanje Pirsonovog i Spermanovog koeficijenta korelacije. Analiza je podeljena u tri dela, u zavisnosti od prostorne skale promenljivih. Na prvom nivou analize koristili smo n = 35 podregiona u Vijelkopolskom Vojvodstvu. U proučavanju ove oblasti, izračunali smo tri para korelacionih koeficijenata za sledeće promenljive: X1-X2, X1-X3 i X4-X5. Na Slici 2.4 se vidi da su za prvi par i Pirsonov i Spermanov koeficijent korelacije visoki i veoma značajni. U slučaju drugog para, samo je koeficijent Spermana značajan, a u trećem slučaju imamo da je samo Pirsonov koeficijent značajan. Zanimljivo je primetiti da u poslednjem slučaju imamo dva različita smera podataka, ali je samo jedna od njih značajna.



Slika 2.4 Poređenje korelacija za demografske karakteristike u Vijelkopolskom Vojvodstvu u Poljskoj



Slika 2.5 Poređenje korelacija za demografske karakteristike subregiona u Poljskoj

Druga grupa parova (drugi nivo) je dobijena za subregionalni nivo ponovo, ali u celoj Poljskoj, gde je n = 373 (Slika 2.5). Oba koeficijenta su veoma značajni za prvi par X6-X7 i iznosi oko -0.56. U slučaju drugog para pronašli smo dve značajne korelacije između X4 i X8, ali Spermanov koeficijent je bio veći od Pirsonovog. Poslednji par u ovoj seriji je X4-X9. U ovom slučaju imamo da je Pirsonov koeficijent neznačajan i negativan, ali blizu nule, dok je Spermanov koeficijent bio značajan i jednak 0.25

Kada analiziramo oba koeficijenta korelacije, i Pirsonov i Spermanov, vidimo da se logički može očekivati da će značaj jednog izražavati značaj drugog. S druge strane, obrnuta implikacija ne mora da izgleda logički tačno. Kao što smo videli iz prethodnog, značaj korelacije Spermana može dovesti do značajne ili neznačajne Pirsonove korelacije, čak i za velike skupove podataka, što je u skladu sa logičkim razumevanjem razlike između dve koeficijenata. Međutim, logično obrazloženje nije tačno u slučaju značaja Pirsonovog koeficijenta koji se prevodi na značaj koeficijenta Spermana. Tako je moguće doći do situacije da je Pirsonov koeficijent negativan, a Spermanov pozitivan. Sve nas to dovodi do sledećeg tvrđenja: ''Budite sigurni da ne treba da tumačite Spermanov rang koeficijent korelacije kao značajnu meru snage povezanosti između dve varijable.''

2.1.5. Testiranje hipoteze za Pirsonov koeficijent korelacije

Neka je osnovni dvodimenzionalni skup normalno raspoređen i ako mu pripada koeficijent korelacije $\rho = 0$, tada promenljiva:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \qquad (*)$$

ima Studentovu t raspodelu sa $k = n-2$ stepena slobode. Pretpostavićemo da se o stohastičkoj povezanosti promenljivih X i Y ništa ne zna, osim da ima karakter linearne korelacije, ako veza uopšte postoji. Na bazi koeficijenta korelacije r uzorka od n parova vrednosti, može se testirati hipoteza:

$$H_0 : \ \rho = 0$$

Prema alternativnoj hipotezi:

$$H_0 : \ \rho \neq 0$$

Iz relacije $r \neq 0$ ne sme se direktno zaključiti da među promenljivama X i Y postoji linearna korelacija sa $\rho \neq 0$, već treba sprovesti postupak testiranja hipoteze i ispitati da li se r signifikantno razlikuje od nule.

Ako se pretpostavi da je koeficijent korelacije ρ dvodimenzionalne normalne raspodele različit od nule, tada je raspodela koeficijenta korelacije r uzorka asimetrična, a varijabla t iz (*) nije više takva da ima Studentovu raspodelu. Međutim, R. A. Fischer je pokazao da za $\rho \neq 0$ raspodela varijable:

$$z = \frac{1}{2} \ln \frac{1+r}{1-r}$$

vrlo brzo teži normalnoj raspodeli porastom veličine uzorka. Pri tom su parametri te normalne raspodele dati izrazima:

$$\mu_z = \frac{1}{2} \ln \frac{1+\rho}{1-\rho} + \frac{\rho}{2(n-1)}$$

$$\sigma_z^2 = \frac{1}{n-3}$$

Činjenicu da se na raspodelu promenljive Z može aproksimativno primeniti svojstva normalne raspodele $N\left(\mu_z, \sigma_z^2\right)$, omogućava brzo i jednostavno rešavanje dva problema:

- Testiranje hipoteze $H_0: \rho = \rho_0 \neq 0$ prema alternativnoj hipotezi $H_0: \rho = \rho_1 \neq \rho_0$
- Intervalno procenjivanje koeficijenta korelacije na populaciji na osnovu izračunatog koeficijenta korelacije uzorka

2.2. Kanonička korelaciona analiza

Statistika uopšte, a multivarijaciona analiza posebno, u velikoj meri počiva na linearnim kombinacijama originalnih promenljivih. Kod kanoničke korelacije imamo dva skupa promenljivih, čije linearne kombinacije se određuju tako da korelacije između njih budu što veće. Kanoničku korelaciju je prvi predložio Hoteling 1936. godine. U svom fundamentalnom radu „Relation Between Two Sets of Variates" Hoteling kaže da „su relacije između dva skupa promenljivih kojima će se baviti samo one koje ostaju invarijantne na proizvoljnu linearnu transformaciju svakog od skupova posebno". Svaki skup promenljivih on posmatra kao jednu višedimenzionalnu promenljivu. Hoteling je koristio termine kanonička promenljiva i kanonička korelacija, koji su ubrzo postali opštepoznati i prihvaćeni u statistici. Pošto rezultujuće linearne kombinacije originalnih promenljivih iz jednog skupa predstavljaju skup promenljivih u kanoničkoj formi, i promenljive i njihove korelacije se nazivaju kanoničkim.

Momirović polazi od stanovišta da je kanonička korelacija najopštija od svih klasičnih metoda multivarijacione statistike i da se iz nje kao specijalni slučajevi mogu izvesti regresiona analiza, analiza varijanse, diskriminaciona analiza, faktorska analiza, pa čak i klaster analiza. Ove tvrdnje je on u svojim brojnim radovima i dokazao (Momirović, 1977, 1988, 1997a, 1997b; Knežević i Momirović, 1996).

Osnovna ideja kanoničke korelacije sastoji se u tome da se nastoji da se maksimiziraju korelacije između ortogonalnih linearnih kombinacija promenljivih iz dva skupa. Traži se po jedna linearna kombinacija promenljivih iz oba skupa tako da korelacija između njih bude maksimalna moguća. Zatim se traži drugi par linearnih kombinacija sa maksimalnom korelacijom, ali pod uslovom da je svaka linearna kombinacija ortogonalna na linearnu kombinaciju prethodno formiranu u istom skupu. Postupak se tako nastavlja do poslednjeg teoretski mogućeg para kanoničkih promenljivih. Konačni rezultat su dva skupa linearnih kombinacija (tj. kanoničkih promenljivih) takvih da su ispunjeni sledeći uslovi:

- Maksimizirane su korelacije između kanoničke promenljive iz jednog skupa i njoj odgovarajuće kanoničke promenljive iz drugog skupa. Ove korelacije se nazivaju kanoničke korelacije.
- Kanoničke promenljive iz svakog od skupova su međusobno ortogonalne.
- Kanonička promenljiva iz jednog skupa ortogonalna je na sve kanoničke promenljive iz drugog skupa, osim one sa kojom čini par čija je korelacija maksimizirana. Drugim rečima, krajnji rezultat kanoničke korelacione analize je jedan biortogonalni sistem.

2.3. Vektorski koeficijent korelacije

Poznato je da varijansa kao očekivana vrednost kvadrata odstupanja jednodimenzionalne slučajne promenljive X od njene aritmetičke sredine, predstavlja jednu meru disperzije te slučajne promenljive. Označimo je sa $w = E(X-m)^2$, gde je $m = E(X)$.

Definiciju varijanse jednodimenzionalne slučajne promenljive uopštićemo i na slučajne vektore. Neka je X, $p-$dimenzionalna slučajna promenljiva, tj. $X = (X_1, X_2, \dots, X_p)$. Za svaku komponentu slučajne promenljive X možemo odrediti odgovarajuću varijansu na osnovu marginalnog zakona verovatnoće tako da ćemo dobiti p varijansi koje će predstavljati mere rasturanja pojedinih komponenti promenljive X. Pored toga, za svake dve komponente promenljive X možemo odrediti kovarijansu između njih kao očekivanu vrednost proizvoda odstupanja komponenti od njihovih sredina, tj. kao

$$w_{ij} = E((X_i - m_i)(X_j - m_j)).$$

Na taj način svakoj $p-$dimenzionalnoj slučajnoj promenljivoj X odgovara jedna dispersiona matrica reda $p \times p$ koja je simetrična zbog osobine komutativnosti kovarijansi, tj. zbog

$$w_{ij} = w_{ji}, \forall i, j \in \{1, 2, \ldots, p\}$$

Na dijagonali dispersione matrice nalaze se pokazatelji rasturanja pojedinih komponenti duž odgovarajućih osa, ali nezavisnih od rasturanja drugih komponenti. Ostali elementi dispersione matrice su pokazatelji zajedničkih rasturanja parova komponenti promenljive X. Prema tome, slučajnoj promenljivoj X pridružujemo jednu matricu reda $p \times p$ kao pokazatelj rasturanja X duž $p-$dimenzionalnog prostora. Interesuje nas kako da $p-$dimenzionalnoj slučajnoj promenljivoj X dodelimo jednu vrednost koja bi predstavljala meru disperzije te promenljive.

Definicija 1. Mera rasturanja slučajne promenljive X je determinanta dispersione matrice

$$\sigma_p^2(X) = \begin{vmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{vmatrix}$$

pri čemu je $\forall i, j\left[i, j \in \{1,2,\ldots,n\} \Rightarrow w_{ij} = E((X_i - m_i)(X_j - m_j))\right]$.

Determinanta u gornjoj definiciji naziva se generalizovana varijansa $p-$ dimenzionalne slučajne promenljive. Na osnovu osobina dispersione matrice proističe da je generalizovana varijansa nenegativan broj. Vrednost nulu ima onda i samo onda kada su komponente slučajne promenljive X linearno zavisne, tj. kad se vrednosti slučajne promenljive X nalaze u hiperravni $p-$dimenzionalnog prostora.

Poznato je da determinanta pozitivno definitne simetrične matrice manja ili jednaka proizvodu njenih elemenata na dijagonali. Zato generalizovana varijansa zadovoljava nejednačinu

$$0 \le \sigma_p^2(X) \le \prod_{i=1}^{p} \sigma_i^2$$

pri čemu je $\sigma_i^2 = w_{ii}$ disperzija $i-$te komponente slučajne promenljive.

Generalizovana varijansa dostiže svoju maksimalnu vrednost $\prod_{i=1}^{p} \sigma_i^2$ ako i samo ako su komponente slučajne promenljive X međusobno linearno nezavisne.

Da bismo definisali vektorski koeficijent korelacije posmatraćemo prvo dvodimenzionalnu slučajnu promenljivu $X = \{X_1, X_2\}$. Dispersiona matrica slučajne promenljive X je

$$W = \begin{vmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{vmatrix}$$

Determinanta dispersione matrice ima maksimalnu vrednost kad su X_1 i X_2 međusobno nezavisne promenljive i ta maksimalna vrednost je jednaka $\max|W| = \sigma_1^2\sigma_2^2$. Minimalnu vrednost determinanta dispersione matrice ima onda i samo onda kad su X_1 i X_2 međusobno linerano zavisne promenljive, tj. kad se moguće vrednosti slučajne promenljive X nalaze na jednoj pravoj. Pošto determinanta dispersione matrice zadovoljava nejednačine

$$0 \leq |W| \leq \max|W|$$

onda ćemo vrednost determinante dispersione matrice odrediti tako što ćemo njenu maksimalnu vrednost pomnožiti sa jednim faktorom. Označimo taj faktor sa $\left(1-r^2\right)$. Na taj način dobićemo jednakost

$$|W| = \sigma_1^2\sigma_2^2(1-r^2).$$

Koeficijent r definisan prethodnom jednakošću, naziva se koeficijent korelacije, a r^2 je tzv. koeficijent determinacije. Koeficijent determinacije ima maksimalnu vrednost jednaku jedinici onda i samo onda kad su promenljive X_1 i X_2 linearno zavisne, a minimalnu vrednost nulu ima onda kad su X_1 i X_2 međusobno nezavisne promenljive. Ova razmatranja ćemo uopštiti na slučaj kad komponente slučajne promenljive X nisu jednodimenzionalne promenljive i na analogan način definisati koeficijent korelacije kao meru linearne zavisnosti između slučajnih vektora.

Neka je **Z** posmatrana $(m+n)$- dimenzionalna slučajna promenljiva takva da je

$$Z = (Z, X); Y^T = (Y_1, \ldots, Y_m); X^T = (X_1, \ldots, X_n)$$

Očekivana vrednost promenljive **Z** označena je sa

$$m_Z = (m_Y, m_X); m_Y = (m_{Y1}, \ldots, m_{Ym}); m_X = (m_{X1}, \ldots, m_{Xn}),$$

a dispersiona matrica sa

$$W = \begin{bmatrix} W_{yy} & W_{yx} \\ W_{xy} & W_{xx} \end{bmatrix}$$

pri čemu je

W_{yy} - dispersiona matrica komponenti vektora Y ,

W_{xx} - dispersiona matrica komponenti vektora X ,

W_{xy} - matrica kovarijansi komponenti vektora Y i komponenti vektora X .

Generalisana varijansa slučajne promenljive Z definisana je determinantom dispersione matrice i može se odrediti preko proizvoda determinanti određenih podmatricama

$$|W| = |W_{yy} - W_{yx} W_{xx}^{-1} W_{xy}| |W_{xx}|$$

Vrednost označena sa R_v , a definsana jednacinom

$$|W| = (1 - R_v^2) \max |W| = (1 - R_v^2) |W_{yy}| |W_{xx}|$$

naziva se vektorski koeficijent korelacije između m-dimenzionalne promenljive Y i n-dimenzionalne promenljive X (Vuković,1977).

Dakle, kvadrat vektorskog koeficijenta korelacije dat je izrazom

$$R_v^2 = 1 - \frac{|W|}{|W_{yy}||W_{xx}|} = 1 - \frac{|W_{yy} - W_{yx}W_{xx}^{-1}W_{xy}|}{|W_{yy}|}$$

Vektorski koeficijent korelacije ima iste osobine kao i običan koeficijent korelacije. On zadovoljava nejednačine $0 \le R_v^2 \le 1$ i ima vrednost jednaku jedinici ako i samo ako su slučajne promenljive X i Y međusobno linearno zavisne, a vrednost nula ako su komponente promenljivih X i Y međusobno nekorelirane. Pored toga, ako su slučajne promenljive X i Y međusobno nezavisne promenljive, njihov vektorski koeficijent korelacije je jednak nuli. Takođe treba napomenuti, da je vektorski koeficijent korelacije invarijantan u odnosu na linearne tranformacije (Vuković, 1977).

3. MULTIVARIJACIONA STATISTIČKA ANALIZA

Sam termin multivarijacione analize se koristi da se predstavi multivarijacioni aspekt analize podataka, u smislu da su mnogobrojne observacije izmerene na velikom broju promenljivih. Brojne se ankete koje imaju od 30 do 100 pitanja. Često se dešava da su odgovori na neka od ovih merila povezani međusobno. Poseban izazov predstavlja objašnjenje komplikovanih međuodnosa različitih varijabli nad istim observacijama. Stoga, rezultati i adekvatna analiza se ne mogu postići bez korišćenja multivarijacione analize (Agresti & Agresti, 1979).

Multivarijaciona statistika obezbeđuje mogućnost analize kompleksnih nizova podataka, tamo gde ima mnogo nezavisnih i zavisnih promenljivih koje su korelisane jedna sa drugom na različitim nivoima povezivanja. Trenutna naučna metodologija ubrzano traži kompleksne relacije između promenljivih u pokušaju da obezbedi sveobuhvatnije studije i modele (Radojičić, 2007). Da bi se došlo do niza rezultata multivarijacione analize potrebno je koristiti proces koji će nam to omogućiti, a to je iterativanog i stohastičkog karaktera. Za analizu koja zahteva multivarijacionu statistiku, odgovarajući nizovi podataka se moraju formirati od vrednosti koje odgovaraju broju promenljivih u odnosu na broj entiteta. Takođe, odgovarajući nizovi podataka mogu biti organizovani kao matrice podataka, korelacione matrice, matrice varijansi-kovarijansi, matrica sume kvadrata i matrica unakrsnih proizvoda (cross product) ili kao niz reziduala (Anderson, 1966).

U procesu naučnog objašnjenja prirode nekog fenomena polaznu osnovu analize sačinjavaju podaci koji se odnose na jedan ili više skupova objekata. Često nismo u prilici da kompleksnu prirodu objekata sagledamo u potpunosti. Međutim, na raspolaganju nam stoji mogućnost obuhvatanja različitih karakteristika jedne višedimenzione pojave. Te karakteristike, odnosno obeležja predstavljaju predmet našeg merenja. Njih ćemo jednostavno zvati promenljive. Pokušaj da se ispita priroda objekta istovremenim merenjem većeg broja promenljivih na svakoj jedinici posmatranja iz jednog ili više skupova objekata predstavlja multivarijacionu analizu (Vuković, 2000).

Mada ne postoji opšte prihvaćena definicija multivarijacione analize, možemo reći da multivarijaciona analiza predstavlja skup statističkih metoda koje simultano analiziraju višedimenziona merenja dobijena za svaku jedinicu posmatranja iz skupa koji ispitujemo (Kovačić, 1992).

Pretpostavimo da smo tokom merenja skupili podatke za i-ti objekat, pri čemu je $i=1,2,...,n$ o njihovom j-tom svojstvu, $j=1,2,..,p$. Dobijeni podaci predstavljaju osnovu multivarijacione analize i predstavljamo ih u vidu matrice podataka, tj. u tabeli u kojoj se red odnosi na objekat, a kolona na promenjivu. Ova matrica podataka nema svojstva matrice, već predstavlja uređeni skup podataka definisan od strane istraživača. Pretpostavimo da imamo n redova (objekata) i p kolona (obeležja, odnosno promenjivih), tabela podataka ili matrica podataka ima sledeći izgled:

	Prom. 1	Prom. 2	...	Prom. j	...	Prom. p
Objekat 1	X_{11}	X_{12}	...	X_{1j}	...	X_{1p}
Objekat 2	X_{21}	X_{22}	...	X_{2j}	...	X_{2p}
...	...	...		...		...
Objekat i	X_{i1}	X_{i2}	...	X_{ij}	...	X_{ip}
...	...	...		...		...
Objekat n	X_{n1}	X_{n2}	...	X_{nj}	...	X_{np}

gde X_{ij} element matrice predstavlja vrednost *j*-te promenjive merene na *i*-tom objektu. U matričnoj notaciji ovu matricu podataka označavamo sa *X*, odnosno $[X_{ij}]$, $i=1,2,...,n; j=1,2,...,p$.

Izbor odgovarajućeg metoda za analizu matrice podataka zavisi od mnogih faktora, a opredeljen je pre svega: vrstom problema, tipom podataka, karakteristikama same metode i u krajnjem slučaju ciljem istraživanja. S obzirom na dimenzije matrice podataka, zaključivanje o međuzavisnosti promenjivih je veoma teško. Upravo u te svrhe je moguće koristiti metode multivarijacione analize za redukciju velike količine podataka. Ovim metodama istovremeno postižemo pojednostavljene složene strukture posmatranog fenomena u cilju njihove lakše interpretacije. Pored ovog, pre svega deskriptivnog zadatka, metode multivarijacione analize koristimo i u procesu zaključivanja, tako što ocenjujemo, na primer stepen međuzavisnosti promenjivih i/ili testiramo njihovu statističku značajnost. Neke od metoda multivarijacione analize su istraživačkog karaktera, što će reći da se koriste ne za testiranje a priori definisanih hipoteza, nego za njihovo generisanje, odnosno konstruisanje. Klasifikacije metoda multivarijacione analize zasnovane su na različitim klasifikacionim kriterijumima (Radojičić, 2007).

Prva klasifikacija metoda pravi razliku prema tome da li su orjentisane ka ispitivanju međuzavisnosti promenjivih ili im je osnovni zadatak ispitvanje međuzavisnosti objekata. Kada istražujemo međuzavisnost promenjivih, tada posmatramo kolone matrice podataka. Osnovu ovih metoda multivarijacione analize predstavlja kovarijaciona ili korelaciona matrica. Kod drugog pristupa, u cilju poređenja dva objekta ili osobe, posmatramo odgovarajuće redove u matrici podataka, odnosno definišemo različite mere bliskosti između dva objekta ili osobe. Osnovu ovih metoda multivarijacione analize predstavlja matrica odstojanja između objekata.

Prema drugoj klasifikaciji, metode delimo u dve grupe: metode zavisnosti i metode međuzavisnosti. Ukoliko smo u istraživanju zainteresovani za ispitivanje zavisnosti između dva skupa promenjivih, gde jedan skup predstavlja zavisne promenjive, a drugi nezavisne promenjive, tada se odgovarajuća klasa metoda naziva metode zavisnosti. Sa druge strane, ako nema a priori, teorijskog osnova za podelu svih promenjivih na dva podskupa promenjivih (zavisnih i nezavisnih), tada koristimo metode međuzavisnosti. Treba uočiti da metode zavisnosti teže da objasne ili predvide jednu ili više zavisnih promenjivih na osnovu skupa nezavisnih promenivih. Metode međuzavisnosti, sa druge strane, nisu po svojoj prirodi prediktivni. Njima se pokušava učiniti prodor u kompleksnu unutrašnju strukturu podataka i to njenim pojednostavljenjem, prvenstveno kroz redukciju podataka (Kovačić, 1992).

Na osnovu podele metoda multivarijacione analize na metode zavisnosti i međuzavisnosti klasifikujemo metode (Radojičić et al., 2003) u jednu od ovih klasa.

Metode zavisnosti

1. Multivarijaciona regresija. Ovo je najpoznatija metoda multivarijacione analize. Koristimo u njenom nazivu izraz multivarijaciona da bi na taj način razlikovali dva slučaja. Prvi, u okviru koga se bavimo analizom zavisnosti jedne promenjive (zavisna promenjiva) od skupa drugih promenjivih (nezavisne promenjive). Ovaj metod analize poznatiji je pod nazivom metod višestruke regresije. Drugi slučaj je kada skup zavisnih promenjivih sadrži više od jednog člana. Za ovaj slučaj kažemo da predstavlja opštiji model multivarijacione regresije. Kod oba modela zadatak nam je ocenjivanje ili predviđanje srednje vrednosti zavisne, odnosno srednjih vrednosti zavisnih promenjivih na bazi poznatih vrednosti nesavisnih promenjivih.

2. Kanonočka korelaciona analiza. Ova analiza se može smatrati uopštenjem višestruke regresione analize. Naime, njome želimo uspostaviti linearnu zavisnost između skupa nezavisnih i skupa zavisnih promenljivih. Kod izračunavanja kanoničke korelacije formiramo dve linearne kombinacije, jednu za svaki skup nezavisnih, a drugu za skup zavisnih promenjivih. Koeficijente ovih linearnih kombinacija određujemo tako da običan koeficijent korelacije između njih bude maksimalan.
3. Diskriminaciona analiza. Bavi se problemom razdvajanja grupa i alokacijom opservacija u ranije definisane grupe. Primena diskriminacione analize omogućava identifikaciju promenjive koja je najviše doprinela razdvajanju grupa kao i predviđanje verovatnoće da će objekat pripasti jednoj od grupa, na osnovu vrednosti skupa nezavisnih promenjivih.
4. Multivarijaciona analiza varijanse (MANOVA). Multivarijaciona analiza varijanse je odgovarajuća metoda kada nam je cilj ispitivanje uticaja različitih nivoa jedne ili više "eksperimentalnih" promenjivih na dve ili više zavisnih promenjivih. U tom smislu, ona predstavlja uopštenje jednodimenzione analize varijanse (ANOVA). Od posebne je koristi u situaciji kada je moguće sprovesti kontrolisani eksperiment (manipulišući sa nekoliko tretmana). Osnovni cilj je testiranje hipoteze koja se tiče varijanse efekata grupa dve ili više zavisnih promenjivih.
5. Logit analiza. Kada je u regresionom modelu zavisna promenjiva dihotomnog tipa (na primer, promenljiva pola sa modalitetima: muško-žensko), tada takav model nazivamo regresioni model sa kvalitativnom zavisnom promenjivom. Kod njh je zavisna promenjiva , tzv. Logit funkcija, logaritam količnika verovatnoća da će dihotomna zavisna promenjiva uzeti jednu ili drugu vrednost. Modele ovog tipa nazivamo i modeli logističke regresione analize.

Metode međusobne zavisnosti

1. Analiza glavnih komponenti. Analiza glavnih komponenti je metoda za redukciju većeg broja promenjivih koje razmatramo, na manji broj novih

promenjivih (glavne komponente). Najčešće manjim brojem glavnih komponenata objašnjavamo veći deo varijanse originalnih promenjivih, što omogućava lakše razumevanje informacije sadržane u podacima. Osnovni zadatak jeste konstruisanje linearne kombinacije orginalnih promenjivih (glavnih komponenata) uz uslov da obuhvate Što je moguće veći iznos varijanse orginalnog skupa promenjivih. Sukcesivne glavne komponente izdvajaju se uz ograničenje da su međusobom nekontrolisane i da obuhvataju u maksimalnom iznosu preostali deo ukupne varijanse koji nije obuhvaćen prethodno izdvojenim komponentama.

2. Faktorska analiza. Slična je metodi glavnih komponenti, po tome što se koristi za varijaciju između promenjivih na osnovu manjeg broja promenjivih (faktora). Međutim za razliku od glavnih komponenti, pretpostavlja postojenje odgovarajućeg statističkog modela kojim orginalnu promenjivu iskazujemo kao linearnu kombinaciju faktora uz dodataka greške modela, odnosno veličina koja odražava stepen nezavisnosti posmatrane promenjive od svih ostalih. Na taj način se celokupna kovarijansa ili korelacija objašnjava zajedničkim faktorima, a neobjašnjeni deo se pridružuje grešci (specifičan faktor). Dakle, kod faktorske analize, za razliku od glavnih komponenti, gde smo zainteresovani za objašnjenje varijanse, interes faktorske analize je usmeren ka objašnjenju kovarijanse, odnosno onog dela ukupne varijanse koji promenljiva deli sa ostalim promenjivim iz posmatranog skupa promenivih.

3. Analiza grupisanja. Analiza grupisanja je metoda za redukciju podataka, no za razliku od prethodne dve metode koje su orjentisane ka kolonama (promenjivma, varijablama), ona je orjentisana ka redovima (objektima) matrice podataka. Ovom analizom kombinujemo objekte u grupe relativno homogenih objekata. Zadatak u mnogim istraživanjima upravo je identifikovanje manjeg broja grupa, tako da su elementi koji pripadaju

nekoj grupi u izvesnom smislu sličniji jedan drugom, nego što su to elementi koji pripadaju drugim grupama.

4. Višedimenziono proporcionalno prikazivanje. Pripada klasi metoda koji su orjentisani kao objektima, a koristi meru sličnosti, odnosno razlike između njih u cilju njihovog prostornog prikazivanja. Izvedena prostorna reprezentacija sadrži geometrijski raspored tačaka na mapi, gde se svaka tačka odnosi na jedan od objekata. Ukoliko se za ovo proporcionalno prikazivanje koristi mera bliskosti dobijena na osnovu merljivih (kvantitativnih) promenjivih nazivu metode dodajemo pridev kvantitativno, a ako smo za računanje mera sličnosti koristili kvalitativne promenjive, tada nazivu metode dodajemo pridev kvalitativno.
5. Loglinearni modeli. Ovi modeli omogućavaju ispitivanje međusobne zavisnosti kvalitativnih promenjivih koje formiraju višedimenzionu tabelu kontigencije. Ukoliko se jedna od promenjivih u tabeli kontigencije može smatrati zavisnom, tada na osnovu ocenjenih loglinearnih modela možemo izvesti, ranije spomenute logit modele. Međutim, kod tabela kontigencije logit funkcija se izračunava preko ćelijkih frekvencija, za razliku od modela logističke analize, gde logit funkciju iskazujemo preko skupa nezavisnih promenjivih koje mogu biti kvantitativne ili kvalitativne.

Pored ovih najčešće korišćenih metoda multivarijacione analize, u naučnim istraživanjima se pojavljuju i druge metode i modeli, koje na već definisan način pripadaju klasi multivarijacionih analiza. U daljem tekstu ćemo detaljnije obratiti pažnju na neke od najznačajnijih metoda multivarijacione analize, koje će se primenjivati u daljem radu.

3.1. Faktorska analiza i analiza glavnih komponenata

Faktorska analiza i analiza glavnih komponenata su statističke tehnike koje se koriste za identifikaciju relativno malog broja faktora koji se mogu koristiti za predstavljanje odnosa između grupa mnogobrojnih, međusobno povezanih, promenljivih. Ove metode pomažu da se identifikuju osnovne, ne direktno vidljive, dimenzije posmatrane pojave. Osnovna razlika između faktorske analize i analize glavnih komponenata je način posmatranja podataka. Kod faktorske analize u razmatranje se uzimaju vandijagonalni elementi disperzione matrice (kovarijanse), dok se analiza glavnih komponenata zasniva na dijagonalnim elementima (varijansama). Faktorske analiza i analiza glavnih komponenata imaju iste ciljeve i postupak njihovog sprovođenja je sličan, tako da metoda glavnih komponenata može biti smatrana metodom faktorske analize (Bulajić, 2002).

Prvi cilj faktorske analize, kao i analize glavnih komponenata, je da se što štedljivije predstavi odnos između promenljivih u jednoj grupi, tj. da zapažene korelacije budu objašnjene pomoću što manje faktora. Drugi važan cilj je da faktori imaju neko značenje. Dobro faktorsko rešenje je jednostavno i lako za interpretaciju. Faktorska analiza, kao i analiza glavnih komponenata, sprovodi se u četiri koraka:

- izračunavanje kovarijacione matrice
- ekstrakcija faktora
- rotacija faktora i
- izračunavanje faktorskih skorova.

U prvom koraku se izračunava kovarijaciona matrica za sve promenljive. Promenljive koje nisu međusobno povezane se mogu identifikovati iz matrice i odgovarajućih statistika. Preko korelacione matrice može biti ocenjena validnost faktorskog modela. Pošto je jedan od osnovnih ciljeva faktorske analize da pronađe one faktore koji su zajednički za više promenljivih, promenljive moraju biti u koralaciji jedna sa drugom kako bi faktorski model bio adekvatan. Ako su korelacije između promenljivih niske, vrlo je verovatno da imaju malo zajedničkih faktora (Bulajić, 2002).

Pokazatelj ja čine veza između promenljivih je parcijalni koeficijent korelacije. Ako promenljive dele zajedničke fakore i kada se eliminišu linearni efekti drugih promenljivih, vrednosti parcijalnih korelacionih koeficijenata među parovima promenljivih bi trebalo da budu male. Parcijalne korelacije su tada procene korelacija između jedinstvenih faktora i one bi trebalo da budu približne 0 kada su pretpostavke faktorske analize ispunjene. U drugom koraku se određuje broj faktora neophodnih za predstavljanje podataka, kao i metod za njihovu ekstrakciju. Razlike između faktorske analize i analize glavnih komponenata ispoljavaju se u ovom koraku. U ovom koraku se određuje i koliko kvalitetno izabrani model odražava podatke. Treći korak se fokusira na transformaciju faktora, kako bi bili lakši za interpretaciju.U četvrtom koraku se za svaku opservaciju i za svaki faktor izračunavaju skorovi. Ovi skorovi se mogu kasnije koristiti kao podaci u drugim analizama (Radojičić, 2007).

3.1.1. Model faktorske analize

Osnovna pretpostavka faktorske analize jeste da se bazni faktori mogu koristiti pri opisu kompleksnih pojava i da su zapažene korelacije između promenljivih posledica postojanja ovih faktora. Cilj faktorske analize je da identifikuje one faktore koji se ne mogu odmah uočiti na osnovu grupe posmatranih promenljivih. Matematički model faktorske analize je sličan nizu jednačina višestruke regresije. Svaka promenljiva je predstavljena kao linearna kombinacija faktora. Grupe promenljivih se izražavaju preko faktora. Faktori koji su korisni za karakterisanje grupe nisu unapred poznati, ali se mogu odrediti faktorskom analizom. Zajednički faktori su oni preko kojh se mogu izraziti sve promenljive, dok su jedinstveni oni koji služe za opisivanje uticaja pojedinih promenljivih, odnosno njegovih delova, koji nisu obuhvaćeni zajedničkim faktorima (Kovačić, 1992).

Model faktorske analize pretpostavlja da se X, vektor od p promenljivih koje se direktno posmatraju, može izraziti preko skupa od m promenljivih koje se ne posmatraju direktno i koje predstavljaju *zajedničke faktore*, u oznaci $F_1,F_2,...,F_m$ ($m<<p$), i p specifičnih, u oznaci $e_1,e_2,...,e_p$. Model se može prikazati u matričnoj notaciji na sledeći način :

$$\underset{(p\times 1)}{X} - \underset{(p\times 1)}{\mu} = \underset{(p\times m)}{B} \cdot \underset{(m\times 1)}{F} + \underset{(p\times 1)}{e}$$

Elementi matrice *B*, b_{ij} su faktorska opterećenja *i*-te promenljive na *j*-ti faktor, a sama matrica naziva je matrica faktorskih opterećenja. Uz pretpostavke modela:

$$E(F) = 0, Cov(F) = E(FF') = \Phi \text{ (Razmatramo specifični slučaj kada je } \Phi = I\text{)}$$

$$E(e) = 0, Cov(e) = E(ee') = \Psi \text{ (Dijagonalna matrica)}$$

$$Cov(e, F) = E(eF') = 0$$

Veza između odstupanja opažljivih promenljivih od njihove sredine i neopažljivih faktora, zajedno sa navedenim pretpostavkama nazivamo *model faktorske analize* . Ovaj model omogućava sledeće razlaganje kovarijacione matrice

$\Sigma = BB^T + \Psi$, gde je:

Σ - Kovarijaciona matrica,

B - Matrica faktorskih opterećenja,

Ψ - Dijagonalna matrica specifičnih faktora.

Kako je $Cov(X, F) = B$, znači da su elementi matrice faktorskih opterećenja kovarijanse između originalnih promenljivih i faktora. Na osnovu razlaganja matrice kovarijacione strukture, važi da je varijansa i-te promenljive

$$S_{ii} = \sum_{j=1}^{m} {B_{ij}}^2 + \Psi_i \ ,$$

tj. podeljena je na dva dela, prvi $h_i^2 = \sum_{j=1}^{m} B_{ij}{}^2$ koji predstavlja varijansu objašnjenu zajedničkim faktorima i naziva se zajednička varijansa ili komunalitet, a drugi deo je specifična varijansa Ψ_i (Kovačić, 1992).

3.1.2. Metoda glavnih komponenenata

Metod glavnih komponenata je metod koji se koristi za smanjivanje dimenzije skupa podataka, uz istovremeno zadržavanje maksimalno mogućeg varijabiliteta koji je prisutan u tim podacima. Na taj način se pojednostavljuje analiza jer se veliki broj pokazatelja svodi na manji broj koji, pri tome, zadržavaju skoro svu informaciju sadržanu u prvobitnom skupu podataka. Time, ne samo što je smanjen broj promenljivih u analizi, već se ostvaruje i napredak u razumevanju strukture fenomena koji se izučava, pa se može reći da metod glavnih komponenata predstavalja istraživačko sredstvo pomoću koga se generišu hipoteze o proučavanom fenomenu. Osnovni zadatak metode glavnih komponenata je određivanje linearne kombinacije originalnih promenljivih koja će imati maksimalnu varijansu. Međutim, drugi i opštiji zadatak je određivanje nekoliko linearnih kombinacija originalnih promenjivih koje će, osim toga što imaju maksimalnu varijansu, biti međusobno nekorelisane i pritom će što je moguće manje gubiti informaciju sadržanu u skupu originalnih promenjivih (Radojičić, 2001). Primenom ove metode originalne promenjive se transformišu u nove promenjive koje se nazivaju glavne komponente (Rao, 1965). Prva glavna komponenta se izdvaja tako da obuhvata najveći deo varijanse originalnog skupa podataka, a naredne komponente se formiraju tako da obuhvate onaj deo varijanse originalnog skupa podataka koji nije obuhvaćen prethodno izdvojenim glavnim komponentama (Kovačić, 1992).

Realizacijom ovih zadataka se postižu dva cilja:

- Redukuje se originalni skup podataka
- Olakšava se njegova interpretacija.

Redukcija se sastoji u tome da se višedimenzioni skup podataka sumira sa manjim brojem linearnih kombinacija (novih promenjivih). Ovom redukcijom se postiže drugi cilj – olakšava se interpretacija kovarijacione strukture originalnog skupa promenjivih i to na osnovu manjeg broja međusobno nekorelisanih glavnih komponenata (Bartholomew, 1980).

Neka X predstavlja p-dimenzioni slučajan vektor sa kovarijacionom matricom Σ. Neka je $Y_1 = \alpha_{11}X_1 + \alpha_{12}X_2 + \ldots + \alpha_{1p}X_p = \alpha'_1 X$ linearna kombinacija elemenata vektora X, gde su $\alpha_{11}, \alpha_{12}, \ldots, \alpha_{1p}$ koeficijenti linearne kombinacije. Poznato je da je $Var(Y_1) = Var(\alpha'_1 X) = \alpha'_1 \sum \alpha_1$. Treba odrediti vektor koeficijenata α_1, tako da se maksimizira varijansa od Y_1. Pošto se $Var(Y_1)$ može uvećavati množenjem vektora $\alpha_1$proizvoljnim skalarom, uvodi se ograničenje da je vektor koeficijenata jedinične dužine,tj. $\alpha_1\alpha_1 = 1$. Ovaj problem maksimiziranja $\alpha_1 \sum \alpha_1$ pri ograničenju $\alpha_1\alpha_1 = 1$ se rešava korišćenjem Lagranžovih množitelja tako što se maksimizira Lagranžova funkcija $\alpha'_1 \sum \alpha_1 - \lambda(\alpha'_1 \alpha_1 - 1)$, gde je λ Lagranžov množitelj.

Diferenciranjem funkcije po koeficijentima α_1 i izjednačavanjem sa nulom se dobija $\sum \alpha_1 - \lambda\alpha_1 = 0$ ili $(\sum - \lambda I)\alpha_1 = 0$, gde je I jedinična matrica dimenzija (pxp). Determinanta $|\Sigma - \lambda I|$ mora biti jednaka 0 da bi se dobilo netrivijalno rešenje za α_1, što znači da λ mora biti jedan od karakterističnih korena kovarijacione matrice Σ. Množenjem sa leve strane gornjeg izraza sa α'_1 dobija se $\alpha'_1\Sigma\alpha_1 - \lambda\, \alpha'_1\alpha_1 = 0$.

Pošto je $\alpha'_1\alpha_1 - 1 = 0$, sledi da je $\lambda = \alpha'_1\Sigma\alpha_1 = Var(Y_1)$. S obzirom da želimo da maksimiziramo varijansu, za λ ćemo uzeti najveći karakteristični koren, na primer λ_1. Iz uslova$(\Sigma - \lambda I)\alpha_1 = 0$sledi da je α_1 odgovarajući karakteristični vektor pridružen karakterističnom korenu λ_1. Njegovim normiranjem$(\alpha'_1\alpha_1 = 1)$ dobija se traženi vektor α_1, a $Y_1 = \alpha_{11}X_1 + \alpha_{12}X_2 + \ldots + \alpha_{1p}X_p = \alpha'_1 X$ predstavlja *prvu glavnu komponentu* .

Ukoliko treba odrediti više od jedne linearne kombinacije, postupak se ponavlja, ali uz dodatni uslov da kovarijansa prve i druge glavne komponente bude jednaka nuli. Neka je $Y_2=\alpha_{21}X_1+\alpha_{22}X_2+....+\alpha_{2p}X_p= \alpha'_2X$ linearna kombinacija, čije koeficijente $\alpha_{21},\alpha_{22},....,\alpha_{2p}$ treba odrediti uz uslov normiranja $\alpha'_2\alpha_2=1$. Dodatni uslov nekorelacije prve i druge komponente se svodi na uslov $\alpha'_2\alpha_1=0$. Pošto je$\Sigma\alpha_1 = \alpha_1\lambda_1$,sledi da je $Cov(Y_2,Y_1) = Cov(\alpha'_2X, \alpha'_1X) = \alpha'_2\Sigma\alpha_1 = \alpha'_1\Sigma\alpha_2 = \alpha'_2\alpha_1\lambda_1 = \alpha'_1\alpha_2\lambda_1$.Formira se Lagranžova funkcija $\alpha'_2\Sigma\alpha_2 - \lambda(\alpha'_2\alpha_2 - 1) - \Phi\ \alpha'_2\alpha_1$, gde su λ i Φ Lagranžovi množitelji. Diferenciranjem po α_2 i izjednačavanjem sa 0 se dobija $\Sigma\alpha_2 - \lambda\alpha_2 - \Phi\alpha_1 = 0$.

Kada ovaj izraz pomnožimo sa leve strane sa α'_1, dobijamo $\alpha'_1\Sigma\alpha_2 - \lambda\ \alpha'_1\alpha_2 - \Phi\ \alpha'_1\alpha_2 = 0$.S obzirom na to da su prva dva člana u ovom izrazu jednaki nuli, a $\alpha'_1\alpha_1=1$, sledi da je $\Phi=0$. Zbog toga je $\Sigma\alpha_2-\lambda\alpha_2=0$, tj. $(\Sigma-\lambda I)\alpha_2=0$, što znači da je λ karakteristični koren kovarijacione matrice Σ, a α_2 je odgovarajući karakteristični vektor. Za λ opet biramo što je moguće veću vrednost, jer je $\lambda= \alpha'_2\Sigma\alpha_2$. Drugi po veličini karakteristični koren označićemo sa λ_2, njemu pridruženi karakteristični vektor je α_2, a linearna kombinacija $Y_2= \alpha'_2X$ je *druga glavna komponenta*.

Na ovaj način se mogu odrediti sve glavne komponente. Njih ima onoliko koliko ima i različitih karakterističnih korena kovarijacione matrice. Ako su svi karakteristični koreni matrice Σ međusobno različiti $(\lambda_1>\lambda_2>....>\lambda_p>0)$, tada postoji p glavnih komponenata$Y_1,Y_2,...,Y_p$ $(Y_j= \alpha'_jX, j=1,2,...,p)$. Vektori koeficijenata $\alpha_1,\alpha_2,....,\alpha_p$ su karakteristični vektori matrice Σ koji su pridruženi karakterističnim korenima λ_j.

Iz definicije glavnih komponenata proizlaze sledeće osobine:

- Očekivana vrednost glavnih komponenata je $E(Y_j)=0$;
- Varijansa $Var(Y_j)= \lambda_j$;
- Kovarijansa svakog para glavnih komponenti je jednaka nuli: $Cov(Y_i,Y_j)=0,\ i \neq j$;
- $Var(Y_1) \geq Var(Y_2) \geq ....\geq Var(Y_p) \geq 0$ (Kovačić, 1992).

Kovarijaciona matrica pruža informaciju o varijansi i kovarijansi promenjljivih, ali na osnovu $p(p+1)/2$ elemenata. U cilju iskazivanja stepena varijabiliteta pomoću jednog broja, u višedimenzionalnom slučaju, definiše se sintetički pokazatelj, *generalizovana varijansa*. Postoje dve alternativne definicije generalizovane varijanse (Kovačić, 1992). Prema prvoj, češće korišćenoj definiciji, generalizovana varijansa je determinanta kovarijacione matrice, a prema drugoj, trag kovarijacione matrice.

Važna osobina glavnih komponenata je da su generalizovane varijanse glavnih komponenata jednake generalizovnim varijansama originalnog skupa promenjivih. Ovo tvrđenje se može dokazati za slučaj obe definicije generalizovane varijanse:

Neka je Y vektor glavnih komponenata takav da je $Y'=[Y_1,Y_2,...,Y_p]$.Transformacija originalnog skupa promenjivih sadržanog u vektoru X se može pisati na ovaj način:

$$Y=AX$$

gde je A (pxp) matrica čiji su redovi karakteristični vektori kovarijacione matrice Σ, tj. $\alpha_1,\alpha_2,\dots, \alpha_p$, pridruženi odgovarajućim karakterističnim korenima $\lambda_1,\lambda_2,\dots,\lambda_p$. Ova matrica je ortoginalna i ima sledeće osobine:

$$A'=A^{-1}, \quad |A| = \pm 1$$

Y = AX je *ortogonalna transformacija*ili *rotacija*, jer se njome vrši rotacija koordinatnih osa za određeni ugao pri čemu ose ostaju međusobno normalne, a ugao između bilo koja dva vektora ostaje isti nakon transformacije.

Primenom matrice A se može izvršiti ortogonalna dekompozicija kvadratne simetrične matrice Σ čiji su koreni različiti. Važi da je $\Sigma =A'\Lambda A$, gde je Λ dijagonalna matrica čiji su elementi karakteristični koreni matrice Σ. Pošto je vektor glavnih komponenata $Y=AX$, njegova kovarijaciona matrica je $Var(Y)=A\Sigma A'$. Zamenom Σ, dobija se$Var(Y)=A(A'\Lambda A)A'=\Lambda$, zato što je A ortogonalna matrica i $A'A=I$.

Sada možemo odrediti generalizovanu varijansu vektora Y. Na osnovu prve definicije, generalizovana varijansa je jednaka determinanti kovarijacione matrice. Kovarijaciona matrica glavnih komponenata je Λ, a njena determinanta $|\Lambda|$ je jednaka proizvodu karakterističnih korena λ_j. Na osnovu izraza ortogonalne dekompozicije matrice Σ se dobija da je $\Lambda=A\Sigma A'$.

Pošto je determinanta proizvoda dve matrice jednaka proizvodu njihovih determinanti, sledi da je$|\Lambda| = |A\Sigma A'| = |A||\Sigma||A'| = |\Sigma|$

Dakle, generalizovane varijanse originalnog i transformisanog skupa podataka su međusobno jednake. Po drugoj definiciji, generalizovana varijansa jednaka je tragu kovarijacione matrice. Trag kovarijacione matrice glavnih komponenata jednak je zbiru karakterističnih korena λ_j. Na osnovu izraza ortogonalne dekompozicije matrice $\Sigma = A'\Lambda A$, dobija se da je$\Lambda = A\Sigma A'$.

Koristeći osobinu traga matrice$(tr(BC)) = (tr(CB))$važi da je $tr(\Lambda) = tr(A\Sigma A') = tr(A'A\Sigma) = tr(\Sigma)$, jer je $A'A=I$, što znači da su i prema drugoj definiciji generalizovane varijanse originalnog i transformisanog skupa podataka međusobno jednake (Kovačić, 1992).

Konstatacija da metod glavnih komponenata predstavlja statistički postupak transformacije originalnog skupa podataka u skup međusobno nekorelisanih promenjivih sa opadajućom vrednošću varijanse, često je pravdanje neuspešnih pokušaja interpretacije glavnih komponenata. Problem koji se javlja u interpretaciji glavnih komponenata nastaje zbog njihove osetljivosti na različite merne skale originalnih promenjivih. U slučaju da u analizi neka od promenjivih ima znatno veću varijansu nego ostale, ona će dominirati prvom glavnom komponentom, bez obzira na to takva je korelaciona struktura podataka. Tada postoje dve mogućnosti:

Prva je da ne koristimo direktno koeficijente linearne korelacije u cilju interpretacije glavnih komponenata, već da zasnujemo analizu na koeficijentima korelacije originalnih promenjivih i glavnih komponenata (Bulajić, 2002).

Druga mogućnost je da kompletnu analizu baziramo na korelacionoj, a ne kovarijacionoj matrici originalnih podataka. Kako se korelaciona matrica može smatrati kovarijacionom za standardizovane podatke, ukupan varijabilitet meren generalizovanom varijansom jednak je p, gde je p broj promenjljivih, dimenzija korelacione matrice i njen trag. Koeficijent korelacije između *k*-te originalne promenjljive i *j*-te glavne komponente je $\alpha_{jk}\sqrt{\lambda_j}$. Rezultati analize glavnih komponenata na osnovu korelacione i kovarijacione matrice mogu se značajno razlikovati, a biće isti kada su originalne promenljive istovrsne, tj. merene na istoj mernoj skali (Kovačić, 1992).

Treba skrenuti pažnju i na iznos varijanse originalnih promenjivih koji se objašnjava zadržanim skupom glavnih komponenata. On pokazuje u kom stepenu zadržane glavne komponente dobro aproksimiraju varijansu svake originalne promenjive posebno. Na osnovu izraza ortogonalne dekompozicije kovarijacione matrice (Σ=A'ΛA) sledi da je varijansa k-te promenjive

$$\sigma_{kk}^{2} = \sum_{j=1}^{p} \lambda_i \alpha_{jk}^{2}, k = 1, 2, \ldots, p$$

Dakle, doprinos svake glavne komponente varijansi k-te promenjive jednak je kvadratu koeficijenata korelacije glavne komponente i te originalne promenjive. Doprinos svih glavnih komponenata izračunavamo kao sumu kvadrata elemenata u k-tom redu korelacione matrice. Količnik dobijene sume i odgovarajuće varijanse originalne promenjive predstavlja proporciju varijanse te promenjive koja je objašnjena zadržanim glavnim komponentama. Ova proporcija se zove komunalitet promenjive. Ako umesto kovarijacione koristimo korelacionu matricu, odmah ćemo dobiti proporciju varijanse originalne promenjive objašnjene zadržanim glavnim komponentama, jer je, standardizacijom promenjivih, vrednost varijanse svedena na jedinicu.

Moguće je izračunati onoliko glavnih komponenata koliko ima promenljivih. Ako su sve glanvne komponente zadržane u analizi, svaka promenljiva će biti tačno predstavljena njima, ali neće doći do smanjenja obima skupa podataka jer postoji onoliko faktora (glavnih komponenata) koliko i promenljivih. U tom slučaju su komunaliteti za svaku promenjljivu jednaki jedinici, jer je glavnim komponentama obuhvaćen ukupni varijabilitet polaznog skupa podataka. Sve izdvojene glavne komponente mogu biti zadržane u analizi onda kad je potrebno da promenjljive, tj. njihove linearne kombinacije, budu međusobno nekorelisane.

3.2. Klaster analiza (analiza grupisanja)

Klaster analiza, ili ***analiza grupisanja,*** je metoda multivarijacione statističke analize, koja se koristi za grupisanje, objekata u grupe, tako da su objekti unutar grupe međusobno slični, a između grupa znatno različiti. Objekti se grupišu u grupe na osnovu mera bliskosti koje se definišu na osnovu njihovih karakteristika.

Ciljevi analize grupisanja su:

- Istraživanje podataka- Ako ne znamo kako je skup objekata strukturiran, analizom grupisanja otkrivamo nepoznatu strukturu;
- Redukcija podataka**;**
- Generisanje hipoteza- Za skup podataka nepoznate strukture, analizom grupisanja formiraju se grupe čiji broj i sastav pomažu u definisanju hipoteza o strukturi podataka. Tako, na primer, broj grupa sugerisan prvobitnom analizom može biti hipoteza koja bi se testirala novim skupom podataka;
- Predviđanje (Anderberg, 1973)**.**

Zadatak analize grupisanja je vrlo sličan problemu koji rešava diskriminaciona analiza, kada se ova koristi kao sredstvo za klasifikaciju objekata. Razlika je u tome što su kod diskriminacione analize grupe već poznate, dok to kod analize grupisanja nije slučaj.

Svi postupci grupisanja objekata (Bogosavljević, 1988) mogu se podeliti u dve grupe:

Hijerarhijske metode grupisanja:

- Aglomerativne
- Dividivne
- Preklapajuće
- Fazi

Nehijerarhijske metode grupisanja:

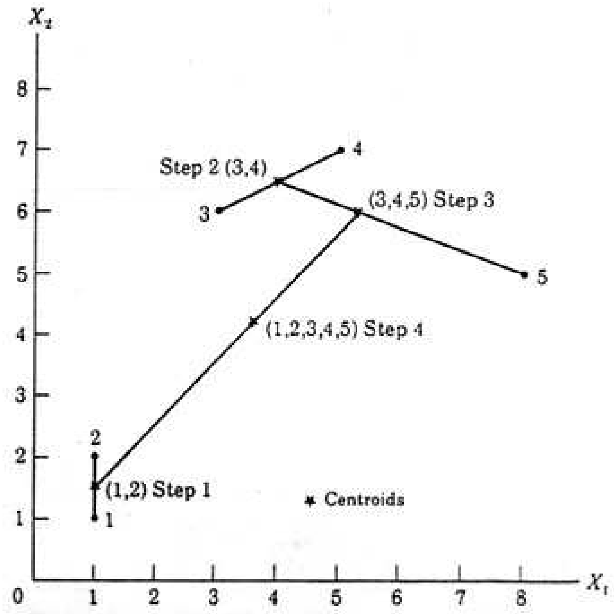
- K-mean algoritam
- Frogy algoritam itd.

Hijerarhijske metode se, u osnovi, sastoje iz iterativnog procesa u kome se spajaju objekti u grupe, a u narednoj iteraciji se spajaju objekti i prethodno formirane grupe, tako da se jednom formirane grupe, u stvari, samo proširuju novim objektima, bez mogućnosti prelaska objekata iz jedne grupe u drugu. Nehijerarhijske metode, međutim, tu mogućnost dozvoljavaju.

Na slici 3.3. predstavljen je primer hijerarhijskog grupisnja objekata, po koracima. Jednom formirane grupe objekata, u narednim koracima se spajaju sa bliskim objektima ili grupama objekata.

Grupisanje objekata u grupe je zasnovano na karakteristikama koje merimo kod svakog objekta. Uzmimo, na primer, dve karakteristike koje merimo kod svakog objekta. U tom slučaju za grafički prikaz podataka u cilju određivanja grupa možemo uzeti dijagram rasturanja. Na osnovu dijagrama rasturanja možemo definisati prirodne grupe kao oblasti u dvodimenzionalnom prostoru sa velikom gustinom tačaka koje su razdvojene od drugih oblasti, oblastima sa malom gustinom tačaka. Međutim, ako definišemo prirodne grupe na osnovu kriterijuma bliskosti, možemo smatrati da objekti unutar grupe treba da budu bliži jedni drugima, nego objektima u drugim grupama.

Osim grafičkih metoda, kod kojih se subjektivnom procenom formiraju grupe, postoje i analitički postupci pomoću kojih se prema skupu formalnih pravila vrši grupisanje objekata u grupe. U osnovi svih ovih metoda se nalazi matrica podataka, tj. matrica sa n redova (objekata) i p kolona (promenljivih). Elementi u jednom redu odnose se na različite karakteristike jednog objekta i formiraju njegov profil.

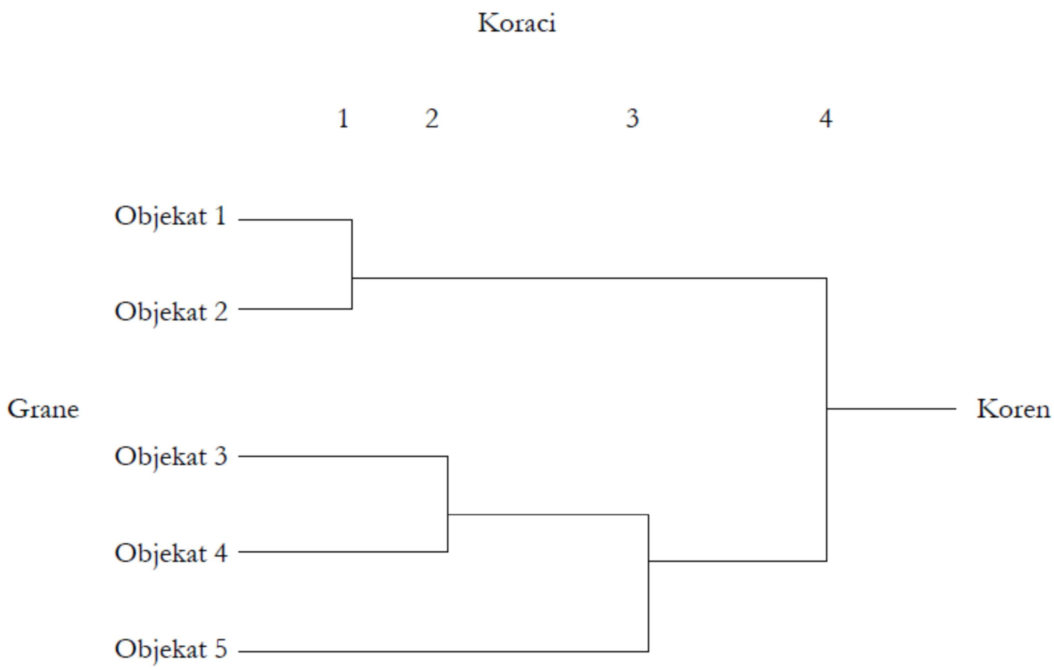


Slika 3.3. Hijerarhijsko grupisanje

Na osnovu (n x p) matrice podataka formiramo (n x n) matricu bliskosti (P) čiji elementi mere stepen sličnosti ili razlike između svih parova profila iz matrice podataka. Na primer, element p_{rs} (r,s = 1,2,...,n) je mera bliskosti između r-tog i s-tog objekta (Kaufman, 1990).

Sledeći korak u analizi grupisanja, nakon što smo formirali matricu bliskosti, je izbor metode grupisanja. Metoda grupisanja je skup pravila pridruživanja objekata u grupe na osnovu mere bliskosti između objekata. Postoji veliki broj metoda grupisanja od kojih treba izabrati onu koja najviše odgovara posmatranom problemu. Najčešće se koriste hijerarhijske metode grupisanja kod kojih se u svakoj iteraciji objekti pridružuju već formiranim grupama, ili sa drugim objektom formiraju novu grupu. Na kraju se dobija hijerarhijska struktura datog skupa objekata koja se zove *hijerarhijsko drvo* ili *dendogram* (Radojičić, 1998).

Na slici 3.4. je prikazan dendogram koji odgovara grupisanju objekata sa slike 3.3.



Slika 3.4. Dendrogram

Postoje dva načina za formiranje hijerarhijske strukture. Prvi način je *udruživanjem*, koje se vrši tako što se grupe formiraju od grana ka korenu drveta, a drugi je *deobom*, gde se krećemo u obrnutom smeru (prvo se formira jedna grupa koja sadrži sve objekte i onda se ona deli dok ne dođemo do grana). Ako nam nije potrebna cela hijerarhijska struktura, jednostavno ćemo "preseći" hijerarhijsko drvo, dobijajući na taj način jedno rešenje analize grupisanja (Vukmirović et al., 1994).

Analizom grupisanja se, kao i faktorskom analizom i analizom glavnih komponenata, može vršiti redukcija podataka. Međutim, analiza grupisanja se bavi redukcijom podataka u odnosu na broj objekata, dok druge dve vrše redukciju u odnosu na broj promenjivih.

Kao u svakoj statističkoj proceduri, određeni broj odluka mora biti donet pre samog početka sprovođenja analize grupisanja:

- Koje promenljive će poslužiti kao osnova za klaster formaciju?
- Kako će se meriti odstojanje između slučajeva?
- Koji će se kriterijum koristiti za spajanje slučajeva u klastere?

Uvek je najvažnije izabrati promenljive koje će se uključiti u analizu. Ako se isključe važne promenljive, analiza može dati slabe rezultate. U klaster analizi, prvobitan izbor promenljivih određuje karakteristike koje će se koristiti u identifikaciji podgrupa.

Koncepti odstojanja i bliskosti su osnove u mnogim statističkim tehnikama. Odstojanje je mera koja meri koliko su daleko dva objekta, a bliskost koliko su blizu (Radojičić et al., 2001). Mere odstojanja su niske, a mere bliskosti visoke za slične entitete.

3.2.1. Mere sličnosti i razlike između objekata

Kada nam je cilj grupisanje objekata, mera bliskosti iskazuje međusobne razlike i sličnosti između dva objekta. Tada mera bliskosti meri stepen međusobnog rastojanja, tj. predstavlja *meru odstojanja* među objekata.

Mera bliskosti p_{rs} predstavlja *meru razlike objekata r i s* ako su ispunjeni sledeći uslovi:

- *Uslov ne-negativnosti:* $p_{rs} > 0$ ako se objekti r i s razlikuju, a $p_{rs} = 0$ ako i samo ako su objekti r i s identični.
- *Uslov simetričnosti:* $p_{rs} = p_{sr}$
- *Uslov triangularnosti:* $p_{rs} \leq p_{rq} + p_{qs}$, za sve r, s i q.

Mera bliskosti p_{rs} predstavlja *meru sličnosti objekata r i s* ako su ispunjeni sledeći uslovi:

- *Uslov normiranosti:* $0 \leq p_{rs} \leq 1$, za sve r i s.
- p_{rs}**= 1,** samo ako su objekti identični
- *Uslov simetričnosti:* $p_{rs} = p_{sr}$.

Najpoznatija mera razlike (odstojanja) je tzv. Euklidska mera odstojanja na bazi kvantitativnih promenljivih. Na primer, ako su x_r i x_s r-ti i s-ti red matrice podataka tada je kvadrat Euklidskog odstojanja:

$$d_{rs}^2 = \sum_{j=1}^{p} (x_{rj} - x_{sj})^2$$

Euklidsko odstojanje je specijalan slučaj tzv. odstojanja Minkowskog koje glasi

$$M = \left[\sum_{j=1}^{p} \left| x_{rj} - x_{sj} \right|^{\lambda} \right]^{1/\lambda}$$

Odstojanje Minkowskog se, kada je λ = 2, svodi na Euklidsko odstojanje (Kovačić, 1992). Na osnovu odstojanja Minkowskog se takođe može definisati i " odstojanje tipa gradskog bloka " tj. tzv. Menhetn odstojanje koje se dobija za λ = 1. U opštem slučaju, što je λ veće, to je mera odstojanja manje osetljiva na prisustvo nestandardnih opservacija.

Mahalanobisovo odstojanje je odstojanje koje vodi računa i o kovarijacionoj strukturi podataka. Naziva se još i multivarijaciona mera odstojanja. Mahalanobisovo odstojanje eliminiše efekat korelisanosti promenljivih, tako da ga ne treba koristiti kada je u analizi upravo taj efekat bitan za razlikovanje objekata.

Merenje bliskosti objekata može se bazirati i na merama sličnosti. Ako posmatramo dva objekta r i s u p-dimenzionalnom prostoru, možemo uzeti veličinu ugla između dva (px1) vektora x_r i x_s da bismo izmerili stepen sličnosti između tih objekata. Što je taj ugao manji, objekti r i s su sličniji međusobom, tako da kao meru sličnosti koristimo kosinus tog ugla:

$$c_{rs} = \frac{\sum_{j=1}^{p} x_{rj} x_{sj}}{\sqrt{\sum_{j=1}^{p} x_{rj}^2 \sum_{j=1}^{p} x_{sj}^2}}$$

Pošto je u gornjem izrazu kvadrat dužine vektora $\Sigma\, x_{rj}^2$ i $\Sigma\, x_{sj}^2$, to znači da mera sličnosti c_{rs} ne zavisi od dužine dva vektora. Mera sličnosti c_{rs} se zove *konusni koeficijent* ili *koeficijent podudarnosti.*

Meru sličnosti takođe možemo konstruisati na osnovu mere odstojanja. Ako je d_{rs} Euklidsko odstojanje između dva objekta, mera sličnosti bi bila

$$p_{rs} = \frac{1}{1+d_{rs}}$$

Pošto je $d_{rs} \geq 0$ očigledno važi $0 \leq p_{rs} \leq 1$.Ako je matrica sličnosti nenegativno definitna, tada možemo i meru odstojanja konstruisati na osnovu mere sličnosti (Bulajić, 2002). Na primer, ako uspostavimo relaciju između kvadrata Euklidskog odstojanja $d_{rs}^2 2$ i kosinusnog koeficijenta c_{rs}, važi $d_{rs}^2 = d_r^2 + d_s^2 - 2d_r d_s c_{rs.}$ Ako usvojimo da je $d_r^2 = d_s^2 = 1$, tada je $d_{rs}^2 = 2(1\text{-}c_{rs})$.

3.2.2. Mere sličnosti i razlike između grupa

Način merenja sličnosti ili razlike između grupa je karakteristika po kojoj se metode analize grupisanja razlikuju. Zbog toga se i naziv metoda grupisanja poklapa sa nazivom mera bliskosti između grupa. Postoji mnogo mera sličnosti i razlike, ali najpoznatije su sledećih pet:

- Jednostruko povezivanje;
- Potpuno povezivanje;
- Prosečno povezivanje;
- Metod centroida i
- Wardov metod (metod minimalne sume kvadrata).

Jednostruko povezivanje definiše odstojanje između dve grupe kao najmanje odstojanje parova objekata iz posmatrane dve grupe. Potpuno povezivanje definiše odstojanje između dve grupe kao najveće odstojanje između parova objekata iz te dve grupe, dok se prema prosečnom povezivanju odstojanje između dve grupe određuje na osnovu prosečnog odstojanja svih parova objekata iz dve posmatrane grupe.

Ako uzmemo dve grupe objekata (r i s) koje sadrže n_r i n_s objekata, i ako označimo opservacije p promenljivih za n objekata u r-toj grupi sa x_{rjm} (j=1,2,...,p ; m=1,2,...n_r), i za n_s objekata u s-toj grupi sa x_{sjm}, i ako centroide r-te grupe označimo sa $x'_r = [x_{r1*}, x_{r2*}, ... x_{rp*}]$ i centroide s-te grupe sa $x'_s = [x_{s1*}, x_{s2*}, \ldots x_{sp*}]$, tada prvu meru odstojanja između ove dve grupe možemo definisati kao

$$d_{rs}^2 = \sum_{j=1}^{p} (x_{rj*} - x_{sj*})^2$$

Pošto postoji ukupno ($n_r n_s$) odstojanja između dve grupe, druga mera odstojanja definiše meru ukupnog odstojanja između dve grupe kao $n_r n_s d_{rs}^2$, a prosečno rastojanje je $n_r n_s d_{rs}^2 / (n_r + n_s)$. Može se pokazati da je ova mera odstojanja između grupa ekvivalentna promeni u sumi kvadrata unutar grupa do koje je došlo zbog udruživanja r-te i s-te grupe (Bulajić, 2002).

Suma kvadrata odstupanja opservacija od svoje sredine tj. suma kvadrata unutar grupe, se za r-tu grupu definiše kao

$$SKW_r = \sum_{m=1}^{n_r} \sum_{j=1}^{p} (x_{rjm} - \overline{x}_{rj*})^2$$

dok je za s-tu grupu

$$SKW_s = \sum_{m=1}^{n_s} \sum_{j=1}^{p} (x_{sjm} - \overline{x}_{sj*})^2$$

Kada udružimo ove dve grupe, dobijamo kombinovanu grupu (na primer t). Ako posmatramo odstupanja opservacija grupe t od novog centroida x'_t = $[x_{t1*}, x_{t2*}, \ldots x_{tp*}]$ dobijamo novu sumu kvadrata unutar t-te grupe

$$SKW_t = \sum_{m=1}^{n_r+n_s} \sum_{j=1}^{p} (x_{tjm} - \overline{x}_{tj*})^2$$

Usled udruživanja r-te i s-te grupe dolazi do povećanja ukupne sume kvadrata unutar grupe koje je dato izrazom: SKW_t – (SKW_r + SKW_s) i ekvivalentno je prosečnom odstojanju između grupa ($n_r n_s d_{rs}^2/(n_r+n_s)$). Do ove relacije dolazimo ako uspostavimo vezu između analize varijanse i određivanja odstojanja između grupa (Radojičić, 2007). U analizi varijanse možemo ukupnu sumu kvadrata unutar kombinovane grupe t (SKW_t) posmatrati kao ukupnu sumu kvadrata u analizi varijanse. Ukupna suma kvadrata u analizi varijanse se razlaže na dva dela: sumu kvadrata unutar grupa (u našem slučaju SKW_r+SKW_s) i sumu kvadrata između grupa (SKB_t) do koje dolazimo na osnovu razlike ukupne sume kvadrata i sume kvadrata unutar grupa, ili direktno

$$SKB_t = \sum_{j=1}^{p} \left[n_r (\bar{x}_{rj*} - \bar{x}_{tj*})^2 + n_s (\bar{x}_{sj*} - \bar{x}_{tj*})^2 \right]$$

$$SKB_t = \frac{n_r n_s}{(n_r + n_s)} \sum_{j=1}^{p} (\bar{x}_{rj*} - \bar{x}_{sj*})^2$$

$$SKB_t = \frac{n_r n_s}{(n_r + n_s)} d_{rs}^2$$

Zaključujemo da je druga mera odstojanja između grupa ekvivalentna sumi kvadrata između grupa, tj. priraštaju u sumi kvadrata unutar grupa do koga je došlo udruživanjem r-te i s-te grupe. Osnovu Wardove metode hijerarhijskog udruživanja predstavlja upravo ova druga mera odstojanja (Radojičić, 1994).

Nakon formiranja nove grupe potrebno je izračunati odstojanja novoformirane grupe i ostalih grupa:

$$d_{tu}^2 = \alpha_r d_{ru}^2 + \alpha_s d_{su}^2 + \beta d_{rs}^2 + \gamma \left| d_{ru}^2 - d_{su}^2 \right|$$

gde je t novoformirana grupa, u jedna od ostalih grupa (različita od r i s), a α_r, α_s, β i γ su koeficijenti koji zavise od toga koji se metod udruživanja koristi. U gornjem izrazu koristili smo kvadrat Euklidskog odstojanja, što je obavezno samo ako koristimo metod centroida ili Wardov metod (Radojičić, 2001). Za ostale metode možemo koristiti neku drugu meru odstojanja između grupa.

Vrednosti parametara se menjaju u zavisnosti od korišćene mere odstojanja između grupa (Bulajić, 2002):

- Jednostruko povezivanje: $\alpha_r = \alpha_s = \frac{1}{2}, \beta = 0, \gamma = -\frac{1}{2}$

- Potpuno povezivanje: $\alpha_r = \alpha_s = \frac{1}{2}, \beta = 0, \gamma = \frac{1}{2}$

- Prosečno povezivanje: $\alpha_r = \frac{n_r}{n_r + n_s}, \alpha_s = \frac{n_s}{n_r + n_s}, \beta = \gamma = 0$

- Metod centroida: $\alpha_r = \frac{n_r}{n_r + n_s}, \alpha_s = \frac{n_s}{n_r + n_s}, \beta = -\frac{n_r n_s}{(n_r + n_s)^2}, \gamma = 0$

- Wardov metod: $\alpha_r = \frac{n_r + n_u}{n_t + n_u}, \alpha_s = \frac{n_s + n_u}{n_t + n_u}, \beta = -\frac{n_u}{n_t + n_u}, \gamma = 0$

3.2.3. Hijerarhijske metode grupisanja

Hijerarhijske metode grupisanja se mogu svrstati u dve kategorije prema tome da li su zasnovane na iterativnom spajanju (aglomerativne metode) ili deljenju grupa i objekata (dividivne metode).

Prva grupa polazi od pojedinačnih objekata koje udružuje u grupe, a zatim u sledećim iteracijama spaja prethodno formirane grupe i pojedinačne objekte, s tim da jednom formirane grupe ostaju zajedno, tj. nema mogućnosti prelaska objekta iz jedne u drugu grupu. Metode koje spadaju u ovu grupu se zajednički nazivaju *hijerarhijske metode udruživanja*. Na početku postupka hijerarhijskog udruživanja imamo n grupa sa po jednim objektom, a nadalje se postupak odvija po sledećim koracima:

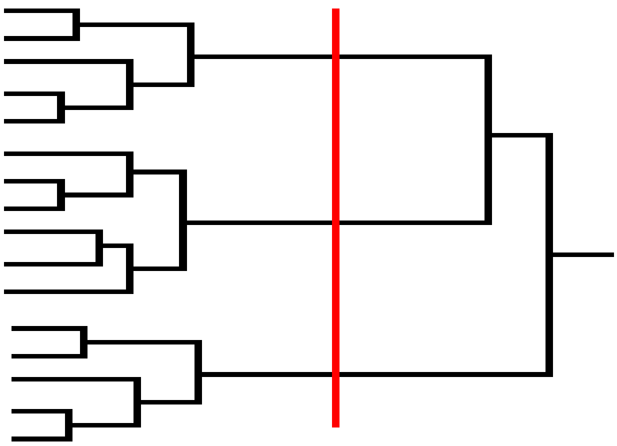
- Na osnovu matrice odstojanja biramo dve najbliže grupe i udružujemo ih u novu grupu (neka su r-ta i s-ta grupa udružene u novu grupu t)
- Određujemo odstojanje ostalih grupa i novoformirane grupe, i ponovo izračunavamo matricu odstojanja
- Prethodna dva koraka se ponavljaju (n-1) put sve dok se ne formira jedna grupa.

Druga grupa metoda rade isto to, ali u suprotnom smeru. One polaze od jedne grupe u kojoj se nalaze svi objekti, i iz nje izdvajaju po jedan objekat ili grupu sve dok se ne formira onoliko grupa koliko ima pojedinačnih objekata. Ove metode se zajednički nazivaju *hijerarhijske metode deobe*.

Najpopularnije metode grupisanja pripadaju hijerarhijskim metodama udruživanja, a među njima se posebno izdvajaju metode udruživanja. Metode hijerarhijskog udruživanja se razlikuju po tome kako u drugoj fazi gornjeg iterativnog postupka određuju međusobnu bliskost grupa (Vuković, 1987).

3.2.4. Određivanje broja grupa (klastera)

Na osnovu dendrograma možemo formirati *izvedenu matricu odstojanja*. Do elemenata ove matrice dolazimo tako što svim parovima objekata iz dve različite grupe koje se udružuju u jednu, pripisujemo istu vrednost odstojanja, onu pri kojoj smo ih udružili u dve grupe. Međusobnim poređenjem odgovarajućih elemenata originalne i izvedene matrice odstojanja može se utvrditi u kom stepenu formirane grupe predstavljaju dobro rešenje problema grupisanja.



Slika 3.5. "Seča" dendrograma, podela na odgovarajući broj grupa

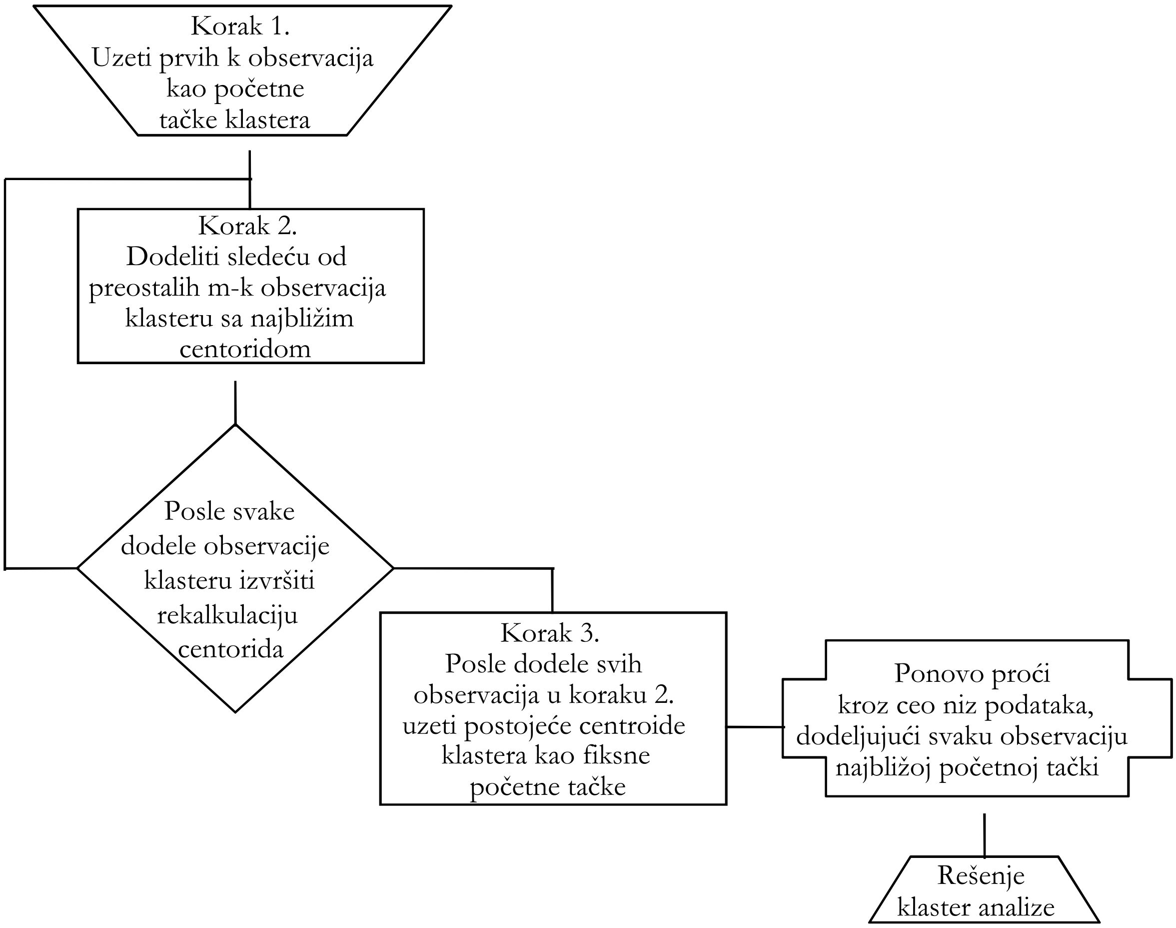
U cilju određivanja broja grupa, grafički prikaz hijerarhijskog grupisanja, odnosno dendrogram, možemo "preseći" na određenoj visini izborom željenog broja grupa. Time smo dobili jedno od mogućih rešenja problema grupisanja. Problem izbora broja grupa se može rešiti praćenjem vrednosti mere odstojanja pri kojoj se dve grupe udružuju u jednu. Krećući se od prvog ka n-1 koraku, vrednost mere odstojanja će rasti, ali u početku sporije, a kasnije brže tj. eksponencijalno. Ako se u okolini očekivanog broja grupa u određenom koraku zabeleži velika promena vrednosti mere odstojanja između grupa, tada taj broj grupa koji je prethodio tom koraku proglašavamo optimalnim.

3.3. Algoritam za rešavanje problema klasifikacije sa unapred definisanim ograničenjima

Za razliku od hijerarhijskih, nehijerarhijski metodi klasifikovanja dozvoljavaju mogućnost premeštanja objekta iz ranije formiranih grupa. U primeni ovih metoda pretpostavlja se da je broj klasa unapred poznat.

Postupak nehijerarhijskog klasifikovanja započinje inicijalnom podelom objekata u izabrani broj grupa ili prema inicijalno određenim centroidima za svaku grupu. Potom se odredi odstojanje između svakog objekta i svake grupe (inicijalnog centroida). Objekti se pridružuju najbližoj grupi. Nakon pridruživanja objekta nekoj grupi, izračunava se centroid grupe iz koje je objekat "otišao" i grupe u kojoj se objekat "pridružio". Ponovo se za svaki objekat izračunava njegovo odstojanje od centroida grupa i vršimo preraspodelu objekata između grupa sve dotle dok izabrana funkcija to sugeriše. [Kovačić, 1992]

Jedan od najpopularnijih metoda za nehijerarhijsko klasifikovanje je metod k-means algoritam. MacQueen (1967) koristi termin "*k-means*" čime objašnjava proces dodeljivanja svake observacije u klaster (od *k* klastera) sa najbližim centroidom (srednja vrednost). Ovaj proces se zasniva na izračunavanju centroida klastera, na osnovu trenutnih veza između klastera. MacQueen's algoritam za klasifikovanje *m* observacija u *k* klastera sadrži sledeće korake [Radojičić, 1994]:



Slika 3.6. MacQueen k-means algoritam

Ovim algoritmom se na bazi početnih centroida koji se može izabrati na sledeće načine :

1. prvih q
2. slučajnih q
3. datih q elemenata e
4. datih q centroida
5. dati nukleusi klasa
6. step-wise selekcija (traži skup q najudaljenijih)

vrši podela na klase, zatim se ponovo proračunavaju centroidi itd.

Jedna od varijanti ovog metoda je konvergentna metoda klasifikovanja, koja koristi *k-means* proces. Implementacija ove varijante je kroz sledeće korake:

1. Početi sa inicijalnom raspodelom observacija u klastere (raspodela observacija, prema izboru ili korišćenjem neke od pomenutih metoda).

2. Za posmatranu observaciju treba izračunati distance do svih centroida klastera. U slučaju da najbliži centorid nije centroid posmatrane observacije, tj. observacija je bliža nekom drugom klasteru (centoridu) nego pripadajućem klasteru (centroidu), potrebno je realocirati posmatranu observaciju ka najbližem centroidu i izvršiti ponovno izračunavanje centrodia klastera (za onaj klaster koji gubi i za onaj klaster koji dobija observaciju).

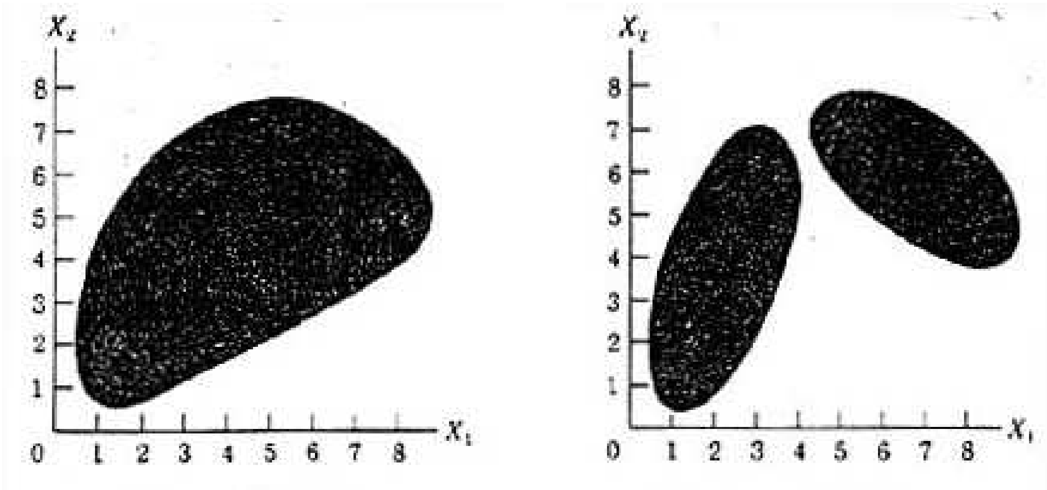
3. Ponoviti korak 2 sve dok nestane konvergencije. Treba nastaviti sve dok se ne ispuni ceo krug, tj. do poslednje observacije.

Kriterijumi za zaustavljanje su :

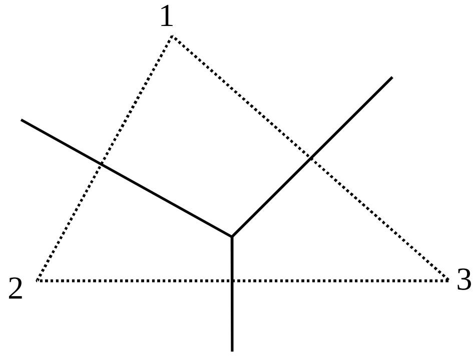
1. Ponavljanje sukcesivnih rešenja
2. Maksimalni broj iteracija
3. Zadata homogenost.

U primeni nehijerarhijskih metoda klasifikovanja treba imati u vidu i to da su oni, kao uostalom i drugi metodi klasifikovanja, osetljivi na prisustvo nestandardnih opservacija. To znači da se u takvim slučajevima može dobiti klasa sa veoma različitim objektima. Metode nehijerarhijskog klasifikovanja, uglavnom se koriste za velike probleme, sa velikim brojem observacija, gde nije potrebno računati i čuvati matricu sličnosti ili čuvati observacije.

Nehjierarhijske metode klasifikovanja se od hijerarhijskih razlikuju u unapred određenom broju klastera, ili bar unapred određenim brojem inicijalnih klastera. Sve metode se uglavnom zasnivaju na poboljšanju neke unapred određene klasifikacije.

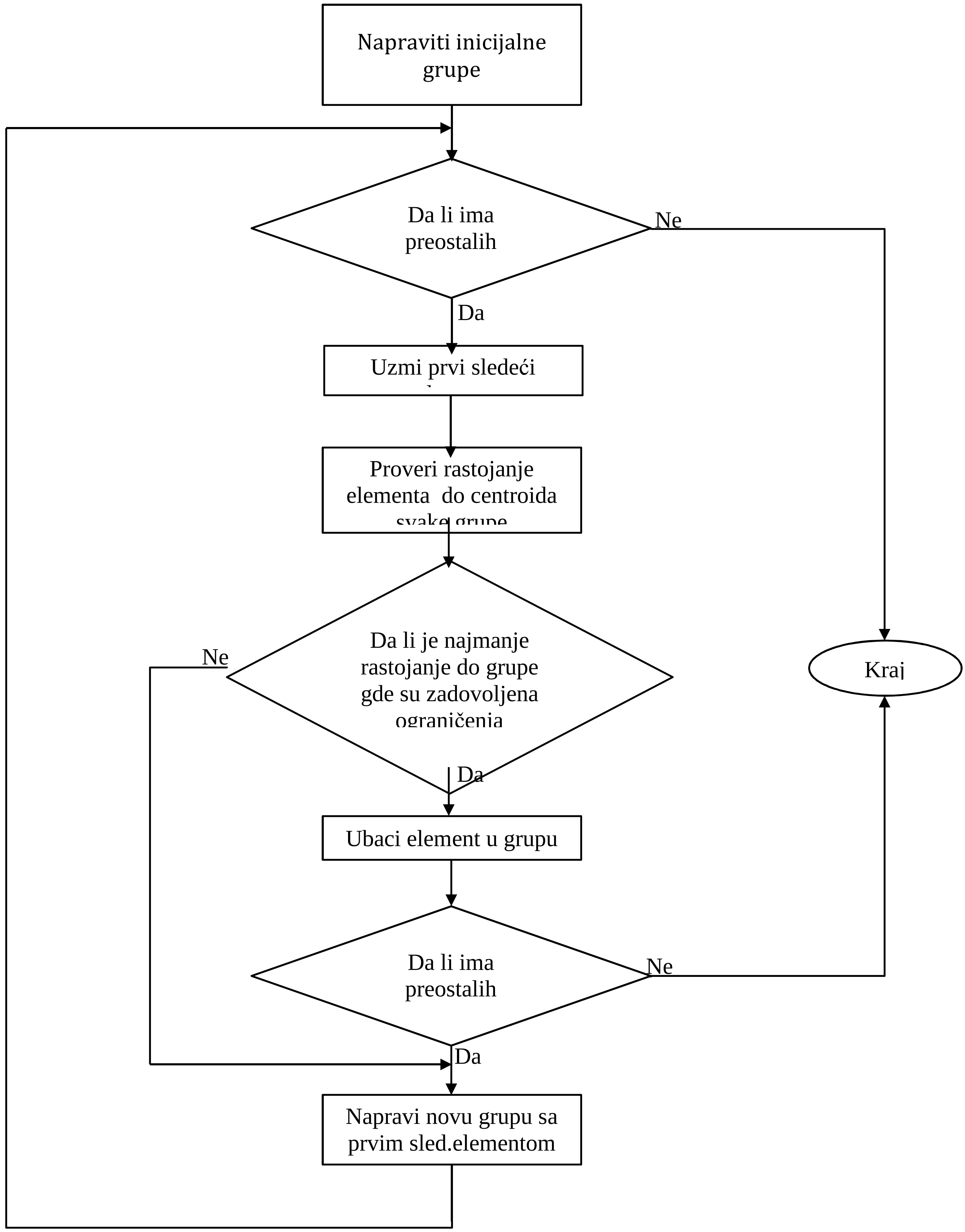


Slika 3.8. Primer razdvajanja grupa kod nehijerarhijske klasifikacije



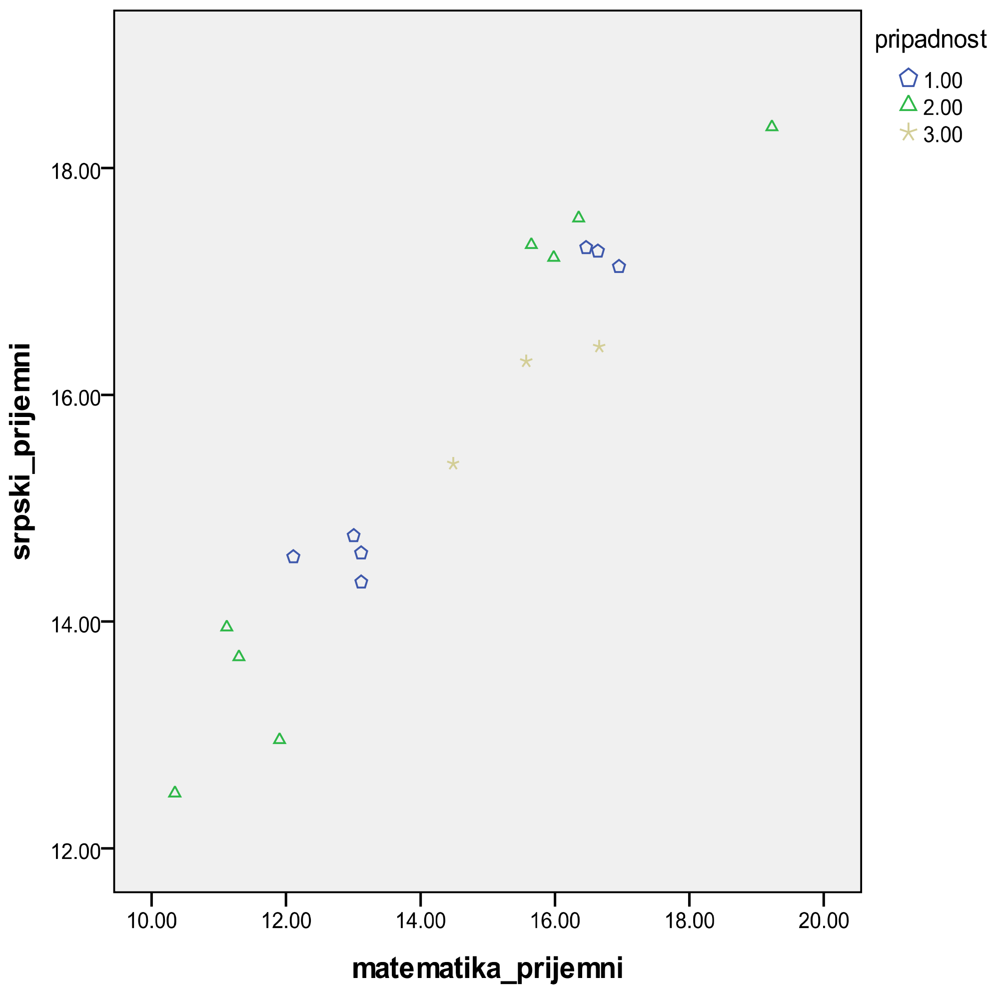
Slika 3.9. Granice klastera (ekvidistance od početnih tačaka)

Pored prethodno navedenog K-means algoritma za klasterovanje, poseban deo u savremenoj literaturi sigurno se mora posvetiti modifikacijama ovog algoritma. Naime, moguće je u proces klasterovanja inkorporirati određena ograničenja. U kontekstu particionih algoritama, ograničenja su mogu iskazati kroz apriori znanje o tome da li neke instance, objekti, entiteti mogu ili ne mogu biti grupisani u isti klaster. Na osnovu toga, možemo identifikovati dva tipa ograničenja: dozvoljena veza- ograničenje koje specificira koja dva entiteta moraju biti u istom klasteru i zabranjena veza-ograničenje koje određuje koja dva entiteta ne smeju biti grupisani u isti klaster. Dozvoljene veze su organičenja koja definišu tranzitivnu binarnu relaciju između elemenata. Algoritam za klasterovanje pri odgovarajućim ograničenjima je predstavljen na sledeći način:



Slika 3.10. Modifikovani MacQueen k-means algoritam sa unapred definisanim ograničenjima

U okviru razmatranja ovog problema došlo se do ideje da se jedna modifikacija ovog problema primeni na odgovarajuće grupisanje škola. Dakle, ograničenja koja su unapred definisana u smislu koji entiteti mogu, a koji ne mogu da se grupišu u iste klastere, uzeta su u obzir da bi se primenio jedan postupak koji je modifikacija K-mean algoritma. Dakle, 18 osnovnih škola u Beogradu su uzete u obzir za razmatranje, odnosno rezultati koje su učenici tih škola postigli na prijemnom ispitu za srednje škole na testu iz matematike i srpskog jezika. Sada je trebalo grupisati škole, ali tako da u istoj grupi ne mogu biti škole koje nisu iz istog dela grada. Naime, unapred smo definisali ograničenje pa smo posmatrali škole iz užeg gradskog jezgra(2), šireg gradskog jezgra(1) i sa perifernih delova(3).



Slika 3.11 Rezultati klasifikacije dobijeni modifikovanim MacQueen k-means algoritmom

Generalno su učenici iz škola koje pripadaju centralnom gradskom jezgru prikazali bolje rezultate na testu, dok su škole iz prigradskih područja u zlatnoj sredini. U drugoj fazi ovog algoritma se detaljnom analizom dobijenih rezultata, a u zavisnosti od prirode posmatranog problema, izvršavaju određena ukrupnjavanja (spajanja dva klastera u jedan, kada je to moguće zbog uslova ograničenja) dobijenih rezultata grupisanja, a sve u cilju bolje interpetacije rešenja našeg problema. Programsko rešenje ovog načina klasifikacije dato je u Prilogu disertacije.

3.4. Analiza obavijanja podataka

Analiza obavijanja podataka (DEA- Data Envelopment Analysis) je najpoznatija metoda za merenje efikasnosti organizacionih jedinica. Metoda je posebno pogodna za merenje efikasnosti entiteta, gde su u razmatranju uzeti više ulaza i izlaza koji su po svojoj prirodi raznorodni (finansijski, tehnički, tehnološki, ekološki, socijalni, itd.) i izražavaju se u različitim mernim jedinicama. U cilju kreiranja sumarnog sintetičkog pokazatelja koji će uzeti u obzir sve značajne višestruke rezultate i sve resurse koji su korišćeni za njihovo ostvarivanje definisana je sledeća mera efikasnosti:

$$Efikasnost = \frac{težinska\ suma\ izlaza}{težinska\ suma\ ulaza}$$

Prethodna definicija omogućava agregaciju posmatranih ulaza (izlaza) u jedan virtuelni ulaz (izlaz) koji predstavljaju sumu proizvoda težinskih koeficijenata i vrednosti ulaza, odnosno izlaza kome su dodeljeni. Izračunanje indeksa efikasnosti kao količnika virtuelnog izlaza i virtuelnog ulaza podrazumeva rešavanje problema koji se odnosi na izražavanje ulaznih i izlaznih podataka u opsezima vrednosti koje su međusobno uporedive (problem skaliranja). Sledeći problem se odnosi na određivanje težinskih koeficijenata ili ponderisanje, pojedinih ulaza odnosno izlaza.

Osim prethodno pomenutih, problem koji se takođe javlja odnosi se na određivanje efikasnosti više različitih jedinica koje koriste iste vrste ulaza i proizvode iste vrste izlaza. Za svaku od posmatranih jedinica, na osnovu prethodne definicije, moguće je izračunati efikasnost i tako izračunate efikasnosti se mogu iskoristiti za utvrđivanje redosleda jedinica.. Očigledno je da na ovaj način izvršeno rangiranje zavisi od vrednosti ulaza i izlaza jedinica, ali i od vrednosti koje su dodeljene za težinske koeficijente. Različite subjektivne metode višekriterijumske analize podrazumevaju *a priori* određivanje težina od strane donosilaca odluka koje je vezano sa njihovim preferencijama i ciljevima (Čupić, Tummala, & Suknović, 2003). U praksi je veoma teško izvršiti vrednovanje ulaza i izlaza i doći do zajedničkog skupa težinskih koeficijenata, jer pojedine jedinice na različite načine tretiraju važnosti njihovim ulazima i izlazima. Na primer, ako se procenjuje efikasnost škola onda se može uočiti da neke škole dostignuća u muzici i u sportu vrednuju na drugačiji način u odnosu na ostale škole.

Tvorci DEA metode (Charnes, Cooper, & Rhodes, 1978) su pošli od pretpostavke da pri oceni efikasnosti jedinica ne mora postojati objektivan postupak za određivanje vrednosti težinskih koeficijenata. Za sve jedinice čija se efikasnost procenjuje treba odrediti koji su to ulazi i izlazi koje treba uzeti u obzir i koje su najmanje dozvoljene vrednosti za težinske koeficijente.Problem skaliranja se jedinstveno rešava na taj način što se efikasnost izražava kao broj između 0 i 1. Naknadnom analizom moguće je pokazati koje su od razmatranih jedinica efikasne, a koje nisu.

Imajući u vidu podatke o ulazima i izlazima, DEA metoda ocenjuje da li je jedinica koja se posmatra efikasna ili nije u odnosu na preostale jedinice uključene u analizu, odnosno da li se nalazi na granici efikasnosti. Rešenje ovog problema ogleda se u posmatranju raspodele skupa tačaka i konstruisanju linija oko njih koja ih obavija – "obvojnica" (*envelope*). Odatle potiče i naziv metode - Analiza obavijanja podataka. Maksimum izlaza koji svaka jedinica može ostvariti sa datim ulazima u ekonomskom smislu predstavlja granicu efikasnosti i ona se za neefikasne jedinice ponaša kao obvojnica. Metoda analizira svaku jedinicu odlučivanja – DMU (Decision Making Unit) i proverava da li je njene ulaze moguće obaviti odozdo (dati izlaz moguće je postići sa manjom količinom ulaza) kao i da li je moguće njene izlaze obaviti odozgo (sa datim ulazom moguće je proizvoditi veći izlaz). Ako je moguće jedinicu obaviti ona je relativno neefikasna, a ako nije ona učestvuje u formiranju granice efikasnosti .

Dakle, DEA je tehnika matematičkog programiranja koja omogućuje da se utvrdi da li je entitet, na osnovu podataka o njegovim ulazima i izlazima, efikasan ili nije, relativno prema drugim entitetima uključenim u analizu. To je neparametarski pristup jer ne zahteva *a priori* pretpostavku o analitičkoj formi funkcije proizvodnje. Za svaku jedinicu odlučivanja se izračunava maksimalna mera performansi u odnosu na sve druge jedinice u posmatranoj populaciji koje moraju zadovoljiti uslov da "leže" na ili ispod ekstremne granice, koja se naziva granica efikasnosti. Mera efikasnosti koju DEA daje je relativna, jer zavisi od toga koji su i koliki broj entiteta je uključeno u analizu, kao i od broja i strukture ulaza i izlaza.

Glavna karakteristika DEA metode je da ona svaku jedinicu odlučivanja procenjuje kao relativno efikasnu ili relativno neefikasnu. Autori DEA metode navode da se jedna jedinica odlučivanja može okarakterisati kao efikasna samo ako nisu ispunjena sledeća 2 uslova:

1. *Moguće je povećati joj bilo koji izlaz bez povećanja bilo kog od ulaza i bez smanjenja bilo kog drugog izlaza;*
2. *Moguće je smanjiti joj bilo koji ulaz bez smanjenja bilo kog od izlaza i bez povećanja bilo kog drugog ulaza.*

Nivo neefikasnosti određen je upoređivanjem sa jednom referentnom DMU ili sa konveksnom kombinacijom drugih referentnih DMU koje se nalaze na granici efikasnosti i koje koriste proporcionalno isti nivo ulaza, a proizvode proporcionalno isti ili veći nivo izlaza (Athanassopoulos & Curram, 1996). DEA metoda je uspešan i nov način za empirijsko određivanje najbolje praktične granice proizvodnje. Autori u (Charnes, Cooper, Lewin, & Seiford, 1994), posebno ističu sledeće osobine DEA metode:

- *fokus je na pojedinačnim opservacijama nasuprot populacionim usrednjavanjima;*
- *u analizu su uključene vrednosti za više ulaza i izlaza koje su izražene u njihovim prirodnim jedinicama;*
- *određuje se pojedinačna sumarna mera za svaku DMU na osnovu vrednosti ulaznih faktora pri proizvodnji željenih izlaza;*
- *ukazuje se na potrebne promene ulaza i/ili izlaza da bi DMU ispod granice efikasnosti (neefikasan DMU) bio projektovan na granicu efikasnosti;*
- *potpuno jednaki kriterijumi se primenjuju u ocenjivanju svake DMU.*

Čarnsu, Kuperu i Roudsu su razvili DEA modele, koji su vremenom modifikovani i proširivani. Ako raspolažemo podacima o ulazima i izlazima za svaku od *n* DMU čiju efikasnost treba proceniti, onda pri selekciji DMU treba voditi računa o sledećim pretpostavkama (Cooper, Seiford, & Tone, 2000):

- Podaci o ulazima i izlazima su raspoloživi za svaki ulaz i izlaz i imaju pozitivne vrednosti za svaku DMU;
- Svi podaci koji izražavaju interese menadžera ili analitičara su uključeni u analizu efikasnosti;
- U principu teži se smanjenju ulaza i povećanju izlaza i indeks efikasnosti treba da odražava ovaj princip;
- Merne jedinice ulaza i izlaza ne moraju biti jednorodne. One mogu uključivati broj časova, površinu radnog prostora, novac, itd.

Neka je x_{ij} - posmatrani iznos ulaza *i* –te vrste za DMU_j ($x_{ij} > 0$, $i = 1,2,...,m$, $j = 1,2,...,n$), a y_{rj} – posmatrani iznos izlaza *r*-te vrste za DMU_j ($y_{rj} > 0$, $r = 1,2,...,s$, $j = 1,2,...,n$). Charnes et al. (1978) su predložili da se za svaku DMU_k, $k = 1,2,...,n$, reši optimizacioni zadatak (u literaturi poznat kao CCR racio model):

MODEL (M1)

$$(\text{Max})\ h_k = \frac{\sum_{r=1}^{s} u_r y_{rk}}{\sum_{i=1}^{m} v_i x_{ik}} \qquad \text{(M1.1)}$$

p.o.

$$\frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1, \quad j = 1, 2, \ldots, n \qquad \text{(M1.2)}$$

$$u_r \geq 0, \quad r = 1,\ 2, \ldots, s \qquad \text{(M1.3)}$$

$$v_i \geq 0, \quad i = 1,\ 2, \ldots, m \qquad \text{(M1.4)}$$

gde su*:* h_k *– relativna efikasnost* $k-te$ *DMU,* n *- broj DMU koje treba porediti,* m *- broj ulaza ,* s *- broj izlaza,* u_r *- težinski koeficijent za izlaz* r *,* v_i *- težinski koeficijent za ulaz* i .

U modelu se teži maksimizaciji vrednosti h_k na taj način što se svakoj DMU subjektivno dodeljuju vrednosti upravljačkim promenljivim u_r i v_i .U ovako definisanom modelu pretpostavlja se konstantni prinos na obim, odnosno da povećanje vrednosti angažovanih ulaza treba da rezultuje u proporcionalnom povećanju ostvarenih izlaznih nivoa. Vrednost h_k je invarijantna u odnosu na merne jedinica ulaza i izlaza, pri čemu su naravno merne jedinice iste za sve DMU. Ako je vrednost za h_k u funkciji cilja jednaka 1, onda je $k-ta$ DMU relativno efikasna, a ako je manja od 1, onda je ta jedinica odlučivanja relativno neefikasna i vrednost h_k pokazuje za koliko procentualno ova jedinica treba da smanji svoje ulaze. Uslov dat u relaciji (M1.4) važi za sve DMU i označava da svaka od njih leži na ili ispod granice efikasnosti.

Težinski koeficijenti u_i i v_j (nepoznate u modelu) pokazuju stepene važnosti svakog ulaza i izlaza koje svaka jedinica bira tako da bude što je moguće efikasnija. Ako tada ne postoji neka druga jedinica koja sa istim angažovanim ulazima proizvodi veći izlaz onda je posmatrana jedinica efikasna. Dakle, DMU$_k$ bira vrednosti težina za ulaze i izlaze tako da se njena efikasnost maksimizira, ali vrednosti težina moraju biti dopustive za sve DMU uključene u merenje efikasnosti i zadovoljavati uslov da je za svaku DMU odnos težinske sume izlaza i težinske sume ulaza manji ili jednak od 1. Dobijene vrednosti za težinske faktore zavise od skale merenja vrednosti za ulaze i izlaze i nisu pogodne za međusobno poređenje. Udeo i važnost svakog ulaza (izlaza) u dobijenom indeksu efikasnosti pokazuje proizvod vrednosti tog ulaza (izlaza) i dodeljenog težinskog koeficijenta koji se naziva virtuelni ulaz (izlaz). Ograničenja data relacijama (M1.3) i (M1.4) označavaju da težinski koeficijenti mogu imati samo nenegativne vrednosti, a daljim razvijanjem modela su modifikovana u sledeća ograničenja:

$$u_r \geq \varepsilon, \quad r = 1,2,\ldots,s \tag{M1.5}$$

$$v_i \geq \varepsilon, \quad i = 1,2,\ldots,m \tag{M1.6}$$

gde je: ε - mala pozitivna vrednost.

Na ovaj način se sprečava potpuno ignorisanje uticaja pojedinih ulaza i izlaza zato štoneka DMU može da bude "lažno" klasifikovana kao relativno efikasna samo na osnovu vrednosti jednog ulaza i jednog izlaza, za koje će izabrati pogodne vrednosti težinskih faktora.

Zadatak opisan relacijama (M1.2)–(1.5) je nelinearan, nekonveksan sa linearno-razlomljenom funkcijom cilja i linearno-razlomljenim ograničenjima. Cooper et al. (2000) zadatak linearnog razlomljenog programiranja pomoću transformacija su sveli na ekvivalentan linearni problem.

MODEL (M2)

$$(\text{Max})\ h_k = \sum_{r=1}^{s} u_r y_{rk} \tag{M2.1}$$

p.o.

$$\sum_{i=1}^{m} v_i x_{ik} = 1 \qquad \text{(M2.2)}$$

$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0, \quad j = 1,\ 2 \ldots, n \qquad \text{(M2.3)}$$

$$u_r \geq \varepsilon, \quad r = 1,2,\ldots,s \qquad \text{(M2.4)}$$

$$v_i \geq \varepsilon, \quad i = 1,2,\ldots,m \qquad \text{(M2.5)}$$

U modelu M2 za $k-tu$ DMU maksimizira se virtuelni izlaz, a njen virtuelni ulaz je jednak 1. Ograničenja koja su data sa (M2.3) označavaju da optimalne težine za $k-tu$ DMU moraju zadovoljavati uslov da za svaku od n DMU njen virtuelni izlaz ne može biti veći od njenog virtuelnog ulaza. Ako je vrednost funkcije cilja jednaka 1, onda za sve preostale jedinice njihov virtuelni izlaz biće manji od virtuelnog ulaza, a ako je vrednost funkcije cilja manja od 1, onda one jedinice kod kojih virtuelni izlaz bude jednak njihovom virtuelnom ulazu čine uzorne ili referentne jedinice za $k-tu$ DMU i obrazuju ivicu granice efikasnosti u odnosu na koju je izmeren njen nivo efikasnosti.

Imajući u vidu da je broj DMU koje se ocenjuju u najvećem broju slučajeva dosta veći od ukupnog broja ulaza i izlaza, u praksi se, najčešće rešava njegov dualni model. Dualni CCR DEA model glasi:

MODEL (M3)

$$\text{(Min)}\ Z_k - \varepsilon\Big(\sum_{r=1}^{s} s_r^+ + \sum_{i=1}^{m} s_i^-\Big) \qquad \text{(M3.1)}$$

p.o.

$$\sum_{j=1}^{n} \lambda_j . y_{rj} - s_r^+ = y_{rk}, \qquad r = 1,2,\ldots,s \qquad \text{(M3.2)}$$

$$Z_k . x_{ik} - \sum_{j=1}^{n} \lambda_j x_{ij} - s_i^- = 0, \quad i = 1,2,\ldots,m \qquad \text{(M3.3)}$$

$\lambda_j, s_r^+, s_i^- \geq 0;\ j=1,2,\ldots,n,\ r=1,2,\ldots,s,\ i=1,2,\ldots,m,\ Z_k$ - neograničeno(M3.4)

Funkcija cilja u ovako definisanom modelu pokazuje sa kojom minimalnom vrednošću ulaza je moguće ostvariti postojeći nivo izlaza posmatrane jedinice odlučivanja. Promenljiva Z_k naziva se faktor intenziteta i pokazuje koliki je nivo smanjenja izlaza koji posmatrana jedinica treba da pretrpi da bi postala efikasna. Dualne promenljive s_i^- i s_r^+ govore o tome koliko treba biti smanjenje $i-tog$ ulaza i povećanje $r-tog$ izlaza $k-te$ DMU da bi postala efikasna. S obzirom da one predstavljaju dopunu do jednakosti u relacijama (M3.2) i (M3.3), one se nazivaju dopunske promenljive.

Dualna promenljiva λ_j predstavlja dualnu težinu koja pokazuje važnost koja je dodeljena $j-toj$ DMU $(j=1,2,\ldots,n)$ pri definisanju ulazno-izlaznog miksa hipotetičke kompozitne jedinice sa kojom će se DMU_k direktno porediti. Vrednosti za promenljive λ_j $(j=1,2,\ldots,n)$ se biraju tako da svaki od s izlaza hipotetičke kompozitne jedinice $\left(\sum_{j=1}^{n}\lambda_j y_{rj}, r=1,2,\ldots,s\right)$ ne bude manji od odgovarajućeg stvarnog izlaza DMU_k, a da svaki od ulaza kompozitne jedinice $\left(\sum_{j=1}^{n}\lambda_j x_{ij}, i=1,2,\ldots,m\right)$ ne bude manji od odgovarajućeg stvarnog ulaza DMU_k (Savić, 2011). Ako od svih λ_j (j= 1, 2,..., n) samo λ_k ima pozitivnu vrednost onda je faktor intenziteta $Z_k=1$, što znači da je DMU_k angažovala minimalnu količinu ulaznih faktora i granična je tačka(u suprotnom je $k-ta$ DMU neefikasna. One organizacione jedinice koje imaju pozitivnu vrednost za λ_j nazivaju se referentne ili uzorne za $k-tu$ DMU. Najkraće rastojanje između neefikasne DMU i granice efikasnosti je upravo rastojanje do kompozitne jedinice. Znači, ako je $Z_k<1$, onda je DMU_k relativno neefikasna i treba proporcionalno za $(1-Z_k)\cdot 100\%$ da smanji sve ulaze da bi postala efikasna sa postojećim nivoom izlaza.

Za svaku DMU_j (j=1,…,n) uzetu kao DMU_k rešava se odgovarajući problem linearnog programiranja. Zbog povezanosti problema (M2) i (M3), kao i zbog teoreme dualnosti koja je opštevažeća u linearnom programiranju, DMU_k je efikasna, ako i samo ako, su za optimalno rešenje $(\lambda^*, s^{+*}, s^{-*}, Z_k^*)$) problema (M3) ispunjeni uslovi:

$$Z_k^* = 1 \tag{M3.5}$$

$$s^{+*} = s^{-*} = 0 \tag{M3.6}$$

Da bi $k-ta$ DMU bila efikasna neophodan uslov je da joj faktor intenziteta bude jednak 1, kao i da sve dopunske promenljive budu jednake nuli. Ako je faktor intenziteta Z_k jednak 1, a neka od dopunskih promenljivih je pozitivna, DMU_k nije efikasna granična tačka. Za takvu jedinicu se kaže i da je “slabo efikasna”. BCC model meri čistu tehničku efikasnost, odnosno daje meru efikasnosti koja ignoriše uticaj obima poslovanja. Efikasnost obima, koja pokazuje da li posmatrana jedinica posluje sa optimalnim obimom operacija, može se dobiti kada se mera efikasnosti koju daje CCR model (ukupna tehnička efikasnost) podeli sa merom efikasnosti koju daje BCC model (čista tehnička efikasnost). U odnosu na primalni CCR model, primalni BCC model sadrži dodatnu promenljivu u_* koja definiše položaj pomoćne hiperravni koja leži na ili iznad svake DMU uključene u analizu (Martić,1999). Specijalno kada je $u_* = 0$, onda se BCC model svodi na CCR model (M1.1)-(M1.4). Banker et al. (1984) su predložili primalni BCC DEA model koji ima sledeći oblik

MODEL (M4)

$$(\text{Max})\ h_k = \sum_{r=1}^{s} u_r y_{rk} + u_* \tag{M4.1}$$

p.o.

$$\sum_{i=1}^{m} v_i x_{ik} = 1 \tag{M4.2}$$

$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} + u_* \leq 0, \quad j = 1,2 \ldots, n \tag{M4.3}$$

$$u_r \geq \varepsilon, \quad r = 1, 2, \ldots, s \tag{M4.4}$$

$$v_i \geq \varepsilon, \quad i = 1, 2, \ldots, m \tag{M4.5}$$

Ideja na kojoj se zasnivaju BCC modeli lakše se može razumeti na dualnom DEA modelu. Dualni BCC model se dobija ako se u dualni CCR model doda ograničenje konveksnosti $\sum_{j=1}^{n} \lambda_j = 1$.

Modeli prikazani u prethodnom delu (M1)-(M4) su dizajnirani sa ciljem da se minimiziraju ulazi potrebni za proizvodnju tražene količine izlaza. Ovakvi modeli se najčešće nazivaju *ulazno orijentisani* modeli. DMU_k se smatra relativno neefikasnom ako joj je moguće smanjiti bilo koji ulaz bez smanjenja bilo kog izlaza i bez uvećanja nekog od preostalih ulaza. Neefikasna jedinica može postati efikasna smanjujući svoje ulaze (proporcionalno faktoru intenziteta *Z* u dualnom modelu) dok se njeni izlazi ne menjaju. Nasuprot ulaznoj orijentaciji, u *izlazno orijentisanom* modelu cilj je da se maksimizira izlaz pri zadatom nivou ulaza, a neefikasna jedinica postaje efikasna kroz povećanje svojih izlaza (proporcionalno faktoru intenziteta θ u dualnom modelu). DMU_k je relativno neefikasna ako joj je moguće povećati bilo koji izlaz bez povećanja bilo kog ulaza i smanjenja nekog od preostalih izlaza. Pored ove dve striktno određene orijentacije modela u literaturi se često pominju i *neorijentisani* (Cooper, Seiford, & Tone, 2000) ili *kombinovani* modeli (Thanassoulis & Emrouznejad, 1995)). Kod ovih modela se razmatra mogućnost da se vrši simultano smanjenje ulaza i povećanje izlaza da bi posmatrana jedinica postala efikasna.

4. IVANOVIĆEVO ODSTOJANJE

Za ocenjivanje "veličine" neke pojave i uspostavljanja međusobnih odnosa između složenih pojava (sistema) mogu se koristiti različite promenljive, gde svaka promenljiva daje delimičnu predstavu veličine pojave. Osnovno pitanje i definicija problema je da li možemo kombinovanjem tih promenljivih iz skupa X (varijabli), formirati jedan potpuniji, globalniji indeks "veličine" pojava. Ako bi se radilo o jednoj merljivoj veličini, mogla bi se ustanoviti jedna redosledna klasifikacija posmatranog skupa prema "veličini", tj. mogli bi da uspostavimo rang, a ujedno i međusobne odnose između entiteta. Ako je faktor F merljiva veličina i ako se njena vrednost izračunava preko skupa obeležja X, moguće je odrediti rang listu elemenata skupa P u odnosu na F (Ivanović, 1977; Bogosavljević, 1997).

Međutim, postoje brojne prepreke koje otežavaju konstrukciju jednog takvog indeksa. Statistička obeležja veličine pojave iskazana su u različitim jedinicama mere, tako da se ne može govoriti o određivanju jednog sintetičkog broja koji bi na jedan apsolutni način iskazivao "veličinu". Zato bi se u skupu posmatranih pojava mogao odrediti jedan globalni indeks "veličine" jedino kao relativni odnos te pojave prema ostalim pojavama posmatranog skupa (Radojičić et al., 1995).

Takođe, neka obeležja sadrže veću, a neka manju količinu informacije o veličini pojave, tako da sva obeležja nemaju isti značaj. Postavlja se pitanje kako izvršiti izbor obeležja i na koji način ih ponderisati kako bi se izbeglo da neka od njih dobiju suviše veliki značaj. Isto tako, treba voditi računa o varijabilitetu svakog obeležja (Birch, 1964; 1965). Odstupanje između dve pojave, koje postoji u odnosu na jedno obeležje, značajnije je ukoliko je njegova varijansa u posmatranom skupu pojava manja.

Napomenimo da su obeležja međusobno zavisna. Informacija koju pruža jedno obeležje, biće delimično sadržana i u ukupnoj informaciji koju pružaju ostala obeležja (Bogosavljević, 1985). Ivanovićevo I-odstojanje definisano je sa idejom da se izbegnu dupliciteti istih informacija koje nosi niz srodnih obeležja (Ivanović, 1977).

Označimo sa $X = x_1, x_2, ..., x_k$ izabrani skup obeležja, a sa $P = p_1, p_2, ..., p_n$ skup pojava kod kojih ispitujemo i upoređujemo "veličinu".

Uočimo ma koje dve pojave P_r i P_s i uporedimo njihove odgovarajuće vrednosti svih obeležja iz X. Ako su sve razlike tih vrednosti jednake nuli, nema razloga da tvrdimo da postoji neka razlika u "veličini" između ove dve pojave. Ta situacija se može promeniti ako se uvedu nova obeležja.

Ako nam u datim uslovima naknadne informacije nisu dostupne, usvojićemo da su za

$$\forall i \; (i \in \{1,2,...,k\} \Rightarrow \; x_{ir} = x_{is})$$

pojave P_r i P_s iste "veličine". Suprotno, ako je bar jedna od tih razlika različita od nule, ne može se više tvrditi da su pojave jednake "veličine".

Razlika $d_i(r,s) = x_{ir} - x_{is,}$ definiše *diskriminacioni efekat* obeležja X_i u uređenom paru pojava $\langle P_r, P_s \rangle$. Diskriminacioni efekat skupa obeležja X u uređenom paru pojava $\langle P_r, P_s \rangle$ je vektor

$$d_x(r,s) = \langle d_1(r,s),...,d_k(r,s) \rangle, \text{ dok matrica}$$

$$\boldsymbol{d}_x(\boldsymbol{P}) = \begin{bmatrix} 0 & \boldsymbol{d}_x(1,2) & \cdots & \boldsymbol{d}_x(1,\boldsymbol{n}) \\ -\boldsymbol{d}_x(1,2) & 0 & \cdots & \boldsymbol{d}_x(2,\boldsymbol{n}) \\ \vdots & \vdots & 0 & \vdots \\ -\boldsymbol{d}_x(1,\boldsymbol{n}) & -\boldsymbol{d}_x(2,\boldsymbol{n}) & \cdots & 0 \end{bmatrix}$$

predstavlja *efekat diskriminacije* od X u P.

Veliki broj obeležja otežava problem rangiranja ili grupisanja prema "veličini". Naime, ako posebno za svako izabrano obeležje upoređujemo odgovarajuće vrednosti za dva entiteta P_r i P_s može se desiti da jedna posmatrana pojava bude veća od druge u odnosu na jedno obeležje, a manja u odnosu na druga obeležja (Radojičić et al., 1998).

Priroda problema ne dozvoljava da se konstruiše jedan globalni indeks koji bi na jedan apsolutan način iskazao "veličinu" pojave. Međutim, ono što bismo mogli odrediti je relativni položaj jedne pojave u odnosu na ostale pojave iz posmatranog skupa P. Tako dolazimo do pojma "odstojanja" između dve pojave u odnosu na njihovu "veličinu".

Ovo odstojanje treba da zadovolji čitav niz uslova. Neka je $D(r,s)$ odstojanje između elemenata P_r i P_s. Svaki elemenat (pojavu) možemo predstaviti u vidu jedne tačke topološkog prostora. Da bi taj prostor bio metričan, potrebno je da odstojanje zadovoljava sledeće uslove:

- ***Nenegativnost***. Odstojanje je nenegativan realan broj, tj.

$$D(r,s) \geq 0 \text{ i } D(r,r) = 0$$

- ***Komutativnost***. Odstojanje između P_r i P_s jednako je odstojanju između P_s i P_r,

$$D(r,s)=D(s,r)$$

- ***Triangularnost***. Za ma koje tri pojave P_s, P_r i P_q, mora da važi sledeća relacija:

$$D(r,s) + D(s,q) \geq D(r,q)$$

- ***Uslov homogenosti***. Odstojanje između dve pojave je homogena funkcija razlika između odgovarajućih vrednosti njihovih izabranih obeležja. Zato će biti $D(r,s) = 0$ ako i samo ako su sve te razlike jednake nuli.
- ***Uslov rasta***. Odstojanje je neopadajuća funkcija svih tih razlika.

- ***Uslov varijabiliteta***. Razlike d_i (r,s), $i \in \{i,...,k\}$ treba da budu tako ponderisane da je njihovo učešće u odstojanju $D(r,s)$ obrnuto srazmerno standardnoj devijaciji odgovarajućih obeležja X_i, $i \in \{1,...,k\}$. Razlike d_i (r,s) pojavljivaće se zato u obliku

$$\frac{|d_i(r,s)|}{\sigma_i} ili \frac{d_i^2(r,s)}{\sigma_i^2}$$

- ***Anuliranje dupliciteta u informaciji***. Odstojanje $D(r,s)$ trebalo bi konstruisati tako da ponavljanja budu isključena i da samo čist deo informacije svakog obeležja učestvuje u izračunavanju ukupne vrednosti odstojanja.
- ***Uslov asimetrije***. Pošto sva obeležja nemaju isti značaj, potrebno je da se odredi njihova rang lista prema količini informacije koju ona pružaju. Odstojanje će se konstruisati tako da snižavanju ranga jednog obeležja odgovara smanjenje njegovog učešća u odstojanju i to za onu količinu informacije koju daju obeležja višeg ranga.
- ***Uslov nezavisnosti***. Ako su sva obeležja među sobom nezavisna neće doći do ponavljanja istih količina informacija. Zato bi tada izraz za odstojanje trebalo da ima oblik:

$$D(r,s) = \sum_{i=1}^{k} \frac{|d_i(r,s)|}{\sigma_i} \qquad ili \qquad D^2(r,s) = \sum_{i=1}^{k} \frac{d_i^2(r,s)}{\sigma_i^2}$$

- ***Uslov linearne zavisnosti***. Ako između svih obeležja postoji linearna zavisnost, izraz za odstojanje će se svesti na:

$$D(r,s) = \frac{|d_1(r,s)|}{\sigma_1} \qquad ili \qquad D^2(r,s) = \frac{d_1^2(r,s)}{\sigma_1^2}$$

- ***Uslov nezavisnih grupa***. Ako je jedna grupa od m obeležja nezavisna od preostalih k-m obeležja, potrebno je da postoji relacija:

$$D_k(r,s) = D_m(r,s) + D_{k-m}(r,s)$$

  U tom slučaju, odstojanje između pojava P_r i P_s možemo da izračunamo nezavisno, jednom na osnovu prvih m obeležja, a jednom na osnovu preostalih k-m obeležja. Traženo odstojanje, bazirano na svih k obeležja, biće tada jednako zbiru prethodna dva.

- ***Nezavisnost od početka***. Uvek možemo konstruisati dve fiktivne pojave P_+ i P_- čije su odgovarajuće vrednosti obeležja X_i^+ i X_i^- proizvoljno izabrane, ali tako da je za svaku posmatranu pojavu i svako izabrano obeležje:

$$X_i^- \leq X_{ir} \leq X_i^+ \qquad i \in \{1,\dots,k\}$$

- ***Tehnički uslov.*** Ako je na osnovu k obeležja, izračunato odstojanje $D_k(r,s)$ između pojava P_r i P_s i ako se naknadno doda još jedno obeležje, poželjno je da novo odstojanje $D_{k+1}(r,s)$ bude jednako zbiru prethodnog, već izračunatog, odstojanja i jedne dodatne veličine koja odgovara uticaju novog obeležja X_{k+1}. Odnosno, treba da bude

$$D_{k+1} = D_k + E_{k+1},$$

  gde je E_{k+1} dodatak koji se odnosi na novo obeležje. Za dobijanje vrednosti D_{k+1}, dovoljno je tada izračunati samo E_{k+1} i tome dodati već poznatu vrednost D_k.

Neka je izabrano k obeležja sa sledećim redosledom po značaju informacije koje pružaju o "veličini" pojave $X = \langle X_1, \dots, X_k \rangle$.

Ako je $P = \{P_1, \dots, P_n\}$ posmatrani skup pojava (Radojičić, 2001), raspolagaćemo sledećom tabelom:

obeležja / objekti	X_1	X_2	...	X_k
P_1	x_{11}	x_{21}	...	x_{k1}
P_2	x_{12}	x_{22}	...	x_{k2}
...	...	...	...	...
P_n	x_{1n}	x_{2n}	...	x_{kn}

Izračunavanje statističkih parametara obeležja X_i zahteva poznavanje koeficijenata ponderacije osnovnih elemenata x_{ij}. Za različita obeležja, koeficijenti ponderacije ne moraju biti isti.

Ako sa f_i^r označimo relativni koeficijent ponderacije od x_{ir}, imaćemo tabelu:

X / P	X_1	X_2	...	X_k
P_1	f_1^1	f_2^1	...	f_k^1
P_2	f_1^2	f_2^2	...	f_k^2
...	...	...	...	...
P_n	f_1^n	f_2^n	...	f_k^n

pri čemu pojedine kolone mogu biti identične.

Aritmetička sredina i varijansa obeležja X_i biće

$$\bar{x}_i = \sum_{r=1}^{n} f_i^{\,r} x_{ir} \qquad i \in \{1, \dots , k\}; \; \sigma_i^2 = \sum_{r=1}^{n} f_i^{\,r} x_{ir}^2 - \bar{x}_i^2 \qquad i \in \{1, \dots ,k\}.$$

Izračunavanje kovarijanse w_{ij} zahteva poznavanje dvodimenzionalnih koeficijenata ponderacije f_{ij}^r u odnosu na obeležja X_i i X_j. Međutim, u praksi retko raspolažemo dvodimenzionalnim rasporedima *[f_{ij}^r]* i zato se tada obično zadovoljavamo aproksimativnim ocenama

$$\left(f_{ij}^r\right)^* = \frac{\sqrt{f_i^r f_j^r}}{F_{ij}};\ F_{ij} = F_{ji} = \sum_{r=1}^{n} \sqrt{f_i^r f_j^r}\ ;\ i \in \{1, \dots, k\};\ j \in \{1, \dots, k\}.$$

Odgovarajuća aproksimativna vrednost kovarijanse biće

$$w_{ij} = \frac{1}{F_{ij}} \sum_{r=1}^{n} \sqrt{f_i^r f_j^r} \left(x_{ir} - \overline{x}_i\right)\left(x_{jr} - \overline{x}_j\right)$$

a običnog koeficijenta korelacije

$$r_{ij} = \frac{w_{ij}}{\sigma_i \sigma_j};\ i \in \{1, \dots, k\};\ j \in \{1, \dots, k\}\text{(Ivanović, 1977).}$$

Preko elemenata korelacione matrice $\boldsymbol{R} = \lfloor \boldsymbol{r_{ij}} \rfloor$ možemo izračunati parcijalne koeficijente korelacije

$$r_{ji.t} = \frac{r_{ij} - r_{jt} r_{it}}{\sqrt{\left(1 - r_{jt}^2\right)\left(1 - r_{it}^2\right)}};\ i > j;\ \{j,i\} \in \{1, \dots, k\};\ t \notin \{j,i\}.$$

Iterativnim postupkom možemo izračunati i sledeće parcijalne koeficijente korelacije

$$r_{ji.12\ldots j-1} = \frac{r_{ji.12\ldots j-2} - r_{j-1,i.12\ldots j-2}\, r_{j-1,j.12\ldots j-2}}{\sqrt{\left(1 - r_{j-1,i.12\ldots j-2}^2\right)\left(1 - r_{j-1,j.12\ldots j-2}^2\right)}}$$

Na taj način se formira matrica parcijalnih korelacija

$$\boldsymbol{R}_{.} = \begin{bmatrix} 1 & \boldsymbol{r}_{12} & \boldsymbol{r}_{13} & \cdots & \boldsymbol{r}_{1\boldsymbol{k}} \\ \boldsymbol{r}_{12} & 1 & \boldsymbol{r}_{23.1} & \cdots & \boldsymbol{r}_{2\boldsymbol{k}.1} \\ \boldsymbol{r}_{13} & \boldsymbol{r}_{23.1} & 1 & \cdots & \boldsymbol{r}_{3\boldsymbol{k}.12} \\ \vdots & \vdots & \vdots & 1 & \vdots \\ \boldsymbol{r}_{1\boldsymbol{k}} & \boldsymbol{r}_{2\boldsymbol{k}.1} & \cdots & \cdots & 1 \end{bmatrix}.$$

Prema tipu podataka i odstojanja po pojedinačnim obeležjima razlikuju se tri vrste I-odstojanja:

- Obično I-odstojanje;
- Kvadratno I-odstojanje i
- Strukturno I-odstojanje.

4.1. Obično I-odstojanje

Za odabrani skup obeležja $X=\{x_1, x_2, \ldots, x_k\}$, rangiranih prema značajnosti informacije koju pružaju, I-odstojanje između P_r i P_s definiše se izrazom

$$D(r,s) = \sum_{i=1}^{k} \frac{|d_i(r,s)|}{\sigma_i} \prod_{j=1}^{i-1} \left(1 - r_{ji.12\ldots j-1}\right)$$

gde je $d_i(r,s)$ odstojanje između vrednosti obeležja X_i za P_r i P_s , tj. $d_i(r,s) = x_{ir} - x_{is}$, $i \in \{1, \ldots, k\}$, σ_i standardna devijacija od X_i a $r_{ji.12\ldots j-1}$ koeficijent parcijalne korelacije između X_i i X_j, $(j<i)$.

Konstrukcija I-odstojanja je postupna. Počinje se sa integracijom celokupnog diskriminacionog efekta obeležja X_1, tj. obeležja koje sadrži najveću količinu informacije o pojavi koja se rangira. Zatim se dodaje onaj deo diskriminacionog efekta drugog (po rangu) obeležja koji nije bio već uključen u diskriminacionom efektu prvog obeležja, pa onaj deo diskriminacionog efekta trećeg obeležja koji nije bio već uključen u diskriminacionom efektu prva dva obeležja itd. (Ivanović, 1977).

Ovako definisano I-odstojanje zadovoljava svih 13 uslova, koje po Ivanoviću jedno odstojanje treba da zadovoljava.

4.2. Kvadratno I-odstojanje

Kvadratno I-odstojanje dato je sa

$$D^2(r,s) = \sum_{i=1}^{k} \frac{d_i^2(r,s)}{\sigma_i^2} \prod_{j=1}^{i-1} \left(1 - r_{ji.12\ldots j-1}^2\right)$$

Kako važi (Croxton et al., 1967)

$$\prod_{j=1}^{i-1} \left(1 - r_{ji.12\ldots j-1}^2\right) = 1 - r_{i.12\ldots i-1}^2$$

kvadratno I-odstojanje je prikladnije izraziti u obliku:

$$D^2(r,s) = \sum_{i=1}^{k} \frac{d_i^2(r,s)}{\sigma_i^2} \left(1 - r_{i.12\ldots i-1}^2\right)$$

Kvadratno I-odstojanje nije jednako kvadratu običnog I-odstojanja. Kvadratno I-odstojanje se koristi ako je broj izabranih obeležja veliki, pa se u običnom I-odstojanju gubi uticaj jednog broja obeležja nižeg ranga. Veoma često treba na isti način i sa istim obeležjima analizirati više skupova ili isti skup u više vremenskih trenutaka. Tada se može desiti da je nemoguće postići jednakosmernost svih obeležja u svim skupovima, pa se mogu javiti negativni koeficijenti korelacije i negativni koeficijent parcijalne korelacije. Zato se u takvim slučajevima upotrebljava kvadratno, umesto običnog I-odstojanja. Osim toga, kvadratno I-odstojanje zahteva manji broj operacija, a time i manje kompjuterskog vremena od običnog, pa je i to razlog da se koristi kvadratno I-odstojanje (Bogosavljević, 1984).

4.3. Strukturno I-odstojanje

Za konstrukciju strukturnog I-odstojanja polazi se od osobina običnog I-odstojanja između dva skupa P_r i P_s:

$$D(r,s) = \sum_{i=1}^{k} \frac{\left|\overline{x}_{ri} - \overline{x}_{si}\right|}{\sigma_i} \prod_{j=1}^{i-1} \left(1 - r_{ji.12\ldots j-1}\right)$$

Umesto apsolutnih vrednosti razlike aritmetičkih sredina obeležja $X_i$skupova P_r i P_s, koristi se zbir apsolutnih vrednosti razlika odgovarajućih frekvencija:

$$\left|\overline{x}_{ri} - \overline{x}_{si}\right| \approx \sum_{l=1}^{m} \left| f_{ril} - f_{sil} \right|$$

Ovaj zbir apsolutnih razlika odgovarajućih relativnih frekvencija predstavlja odstojanje između struktura skupova P_r i $P_s$za obeležje X_i, a u odnosu na klasifikaciju K_i (Ivanović, 1972).

Ako su sve razlike jednake nuli, odstojanje će biti jednako nuli i obe strukture će biti identične u odnosu na K_i. Ako je jedna od razlika različita od nule, tada postoji bar još jedna razlika koja je takođe različita od nule a suprotno označena, jer je

$$\sum_{l=1}^{m} f_{ril} = \sum_{l=1}^{m} f_{sil} = 1$$

Umesto obične standardne devijacije σ_i, koja se pojavljuje u običnom I-odstojanju, sada će se pojaviti standardna devijacija strukture u odnosu na X_i, odnosno, varijansa strukture obeležja X_i, u okviru skupa P, data je sa:

$$S_i^2 = \begin{vmatrix} \sum_{r=1}^{n} f_{ri1}^2 & \sum_{r=1}^{n} f_{ri1} f_{ri2} & \cdots & \sum_{r=1}^{n} f_{ri1} f_{rim} \\ \sum_{r=1}^{n} f_{ri2} f_{ri1} & \sum_{r=1}^{n} f_{ri2}^2 & \cdots & \sum_{r=1}^{n} f_{ri2} f_{rim} \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{r=1}^{n} f_{rim} f_{ri1} & \sum_{r=1}^{n} f_{rim} f_{ri2} & \cdots & \sum_{r=1}^{n} f_{rim}^2 \end{vmatrix}$$

I umesto parcijalnog koeficijenta korelacije $r_{ji.12...j-1}$ trebalo bi uzeti odgovarajući parcijalni kolektivni koeficijent korelacije $R_{ji.12...j-1}$. Zbog komplikovanosti ovog koeficijenta, koristi se apsolutna vrednost običnog kolektivnog koeficijenta korelacije R_{ji}, koji je definisan obrascem

$$R_{ij}^2 = \frac{Q_{ij}^2}{S_i^2 S_j^2}$$

Tako dolazimo do strukturnog I-odstojanja između skupova P_r i P_s za obeležja $X_1, \dots, X_k$ a u odnosu na respektivne klasifikacije $K_{1, \dots,} K_k$:

$$D(r,s) = \sum_{i=1}^{k} \sum_{l=1}^{m} \left| f_{ril} - f_{sil} \right| \frac{1}{S_i} \prod_{j=1}^{i-1} \left(1 - \left| R_{ji} \right|\right)$$

Ovako dobijeni rezultati, važiće jedino ako su klasifikacije kvalitativne prirode. Kod strukturnog I-odstojanja se međusobno upoređuju strukture u odnosu na pojedina obeležja, a ne vrednosti. Zato je logično da se porede frekvencije onih klasa koje sadrže srednje vrednosti od X_i a zatim odgovarajuće leve, odnosno desne frekvencije (Ivanović, 1977).

4.4. Redosledna klasifikacija i I-odstojanje

Metoda I-odstojanja omogućava da se obrazuje rang lista posmatranih jedinica. Potrebno je najpre fiksirati jednu jedinicu koja će poslužiti kao reperna tačka na skali. U praksi se obično za repernu tačku uzima fiktivna jedinica čije su vrednosti obeležja odgovarajuće minimalne vrednosti u posmatranom skupu. Odnosno, vrednost fiktivnog baznog elementa P_- definisana je sa:

$$x_i^- = \min_{1 \le r \le n} \{x_{ir}\}, \qquad i \in \{1, 2, \ldots, k\}$$

I-odstojanje između pojave P_r i fiktivno najmanje pojave P_- definiše veličinu pojave P_r. Obrazac za I-odstojanje sada se svodi na

$$D_r^- = \sum_{i=1}^{k} \frac{x_{ir} - x_i^-}{\sigma_i} \prod_{j=1}^{i-1} \left(1 - r_{ji.12\ldots j-1}\right)$$

Na taj način možemo odrediti odstojanje za svaki element skupa P, pa ako zatim uredimo sve elemente prema veličini njihovih tako dobijenih I-odstojanja, dobićemo rang listu pojava prema veličini. Za repernu tačku, baznu jedinicu, može se uzeti i fiktivno najveća pojava P_+ unutar skupa P, tj. pojava sa vrednostima obeležja

$$x_i^+ = \max_{1 \le r \le n} \{x_{ir}\}, \qquad i \in \{1,2,\ldots,k\}$$

Odgovarajuće I-odstojanje bi bilo

$$D_r^+ = \sum_{i=1}^{k} \frac{x_i^+ - x_{ir}}{\sigma_i} \prod_{j=1}^{i-1} \left(1 - r_{ji.12\ldots j-1}\right)$$

Ako sada poređamo pojave prema veličini njihovih I-odstojanja, dobijeni redosled biće inverzan prethodnom. Takođe možemo uzeti za repernu tačku i element koji će imati prosečne vrednosti obeležja (Ivanović, 1977; Bogosavljević, 1984). Redosled uključivanja obeležja u obrazac za I-odstojanje treba da odgovara količini informacije koju to obeležje pruža. Za utvrđivanje redosleda obeležja koriste se dve metode:

1. *Subjektivna metoda.* Posle detaljne analize svakog obeležja, poželjno bi bilo da se svi zainteresovani korisnici slažu sa ocenom koje je obeležje u svakom paru $\{X_i, X_j\} \subseteq X$ značajnije u pogledu ocene veličine posmatrane pojave. Tada bi se mogao neposredno odrediti redosled svih obeležja, a izračunavanje I-odstojanja vršilo bi se prema tom redosledu. Retko se dešava da su svi zainteresovani korisnici saglasni sa tako formiranom rang listom obeležja. Svaki od korisnika pri određivanju redosleda obeležja koristi svoja lična znanja i iskustva koja su subjektivne prirode i koja se razlikuju od saznanja i iskustva drugih korisnika. Zato nije pogodno koristiti subjektivnu metodu, već treba koristiti neku objektivnu metodu za utvrđivanje redosleda obeležja.

2. *Objektivna metoda.* Ako postoji potpuna linearna zavisnost između obeležja X_i i veličine pojave, rang liste pojava prema obeležju X_i i prema veličini biće identične. Tada će biti indiferentno po kom od ovih kriterijuma ćemo rangirati pojave. Osnovna ideja za objektivno rangiranje obeležja, počiva na

korelacijama između efektivno korišćenih obeležja i globalnog indeksa koji sadrži maksimalnu količinu informacije.

Mogu se uočiti dva ekstremna slučaja koja mogu nastupiti kod izračunavanja I-odstojanja. U prvom slučaju sva su obeležja nezavisna i do ponavljanja istih informacija neće doći. Tada sva obeležja ulaze u igru sa jednakim pravom, a I-odstojanje se svodi na

$$F_r = \sum_{i=1}^{k} \frac{d_i(r)}{\sigma_i}$$

U drugom ekstremnom slučaju, sva obeležja potpuno zavise od jednog jedinog obeležja. Označimo dominirajuće obeležje sa X_1 i primetimo da su informacije, koje pružaju sva ostala obeležja, sadržane u informaciji obeležja X_1. Tada se I-odstojanje svodi na samo jedan član

$$D_r^{(1)} = \frac{d_1(r)}{\sigma_1}$$

U opštem slučaju I-odstojanje se nalazi između vrednosti ova dva ekstremna slučaja. Ako se u skupu izabranih obeležja *X*, pojavilo neko sa suprotnim smerom u odnosu na ostale, najčešće se ono može zameniti sa svojim komplementarnim obeležjem koje će imati pozitivne korelacije sa ostalim obeležjima iz *X.* Međutim, u praksi se dešava da je neophodno suprotnosmerno supstituisati istosmernim obeležjem. Tada treba izvršiti odgovarajuću korekciju I-odstojanja (Ivanović, 1977). Za repernu tačku može se uzeti fiktivna jedinica $P^*=\langle x_1^*, x_2^*, \ldots, x_k^* \rangle$, pri čemu je $x_i^* = x_i^-$ kada je X_i istosmerno sa ostalim obeležjima, odnosno $x_i^* = x_i^+$, kada je X_i suprotnosmerno sa ostalim obeležjima. Adaptirano I-odstojanje između P_r i P^* je tada:

$$D_r = \sum_{i=1}^{k} \frac{\left| x_{ir} - x_i^* \right|}{\sigma_i} \prod_{j=1}^{i-1} \left(1 - r_{ji.12\ldots.j-1}\right).$$

4.5 Raspodela kvadratne forme slučajnih promenljivih koje imaju normalnu raspodelu

<u>Teorema 4.5.1</u> Neka je $A = [a_{ii}]$ nesingularna dijagonalna matrica n-tog reda, X standardizovana normalno raspoređena n-dimenzionalna slučajna promenljiva, a

$$H_{n;\,A,b}(t) = P\left\{ (X-b)^{'} A (X-b) \le t \right\}$$

gde je b dati vektor.

I Verovatnoća $H_{n;\,A,b}(t)$ može se izraziti u vidu sledećeg uniformno konvergentnog reda

(4.5.1)

$$H_{n;\,A,b}(t) = \sum_{j=0}^{\infty} c F_{n+2j}(t/p) \quad , \quad t < \infty \quad ,$$

gde su $F_k(t)$ funkcije x^2 raspodele sa k stepena slobode, $p > 0$ unapred data konstanta, a koeficijenti c_j dati su izrazima

(4.5.2)

$$c_j = |A|^{-1/2} e^{-\frac{1}{2} b^{'} b} p^{n/2+j} E\left\{ Q^j H_{2j}\left(L/Q^{1/2}\right)\right\} \; / \; (2j)! \quad ,$$

gde smo označili sa

(4.5.3)

$$L = b^{'} A^{-1/2} X$$

(4.5.4)

$$Q = X'\left(A^{-1} - \frac{1}{p}I\right)X \quad .$$

II Funkcija generatrise niza $\{c_j\}$ je oblika (Breslow ,1982)

(4.5.5)

$$V(s) = \prod_{i=1}^{n}\left\{(p/a_{ii})^{1/2}\left[1-(1-p/a_{ii})z\right]^{-\frac{1}{2}}\right\}\exp\left\{-1/2\sum b_1^2 \frac{1-z}{1-\left(1-\frac{p}{a_{ii}}\right)}\right\}$$

III. Između članova niza $\{a_j\}$ postoji sledeća rekurentna veza

$$c_0 = e^{-\frac{1}{a}b'b}\prod_{i=0}^{n}(p/a_{ii})^{1/2}$$

(4.5.6)

$$c_j = 1/(2j)\sum_{r=c}^{j-1} g_{j-r}c_r \quad ,$$

gde smo označili sa

(4.5.7)

$$g_m = \sum_{i=1}^{n}(1-p/a_{ii})^m + mp\sum_{i=1}^{m}\frac{b_i^2}{a_{ii}}(1-p/a_{ii})^{m-1} \quad ,$$

$j, m = 1, 2, \ldots$

Teorema 4.5.2 Uz početne pretpostavke teoreme 4.5.1,

I Kvadratna forma $X'AX$ ima za funkciju raspodele

(4.5.8)

$$H_{n;A,\,0}(t)=\sum_{j=0}^{\infty} c_j\, F_{n+2j}(t/p) \quad ,$$

gde je $p > 0$ unapred data konstanta, a koeficijenti c_j zavise od n, A i p i dati su sa izrazima

(4.5.9)

$$c_j = |A|^{-1/2}\, p^{n/2+j}\, E\left\{(-c)^j\right\} / \left[2^j \;\; j\,!\right] \quad .$$

II Red u (4.5.8) je uniformno konvergentan u svakom konačnom razmaku.

III Funkcija generatrise za niz $\{c_j\}$ je oblika

(4.5.10)

$$\psi(z)=\prod_{i=1}^{n}\left\{(p/a_{ii})^{1/2}\left[1-(1-p/a_{ii})\,z\right]^{-1/2}\right\} \quad .$$

IV Članovi niza $\{c_j\}$ zadovoljavaju rekurentnu vezu

(4.5.11) $$c_j = \frac{1}{2j}\sum_{r=0}^{j-1} g_{j-r}\, c_r \quad ,$$

gde smo označili sa

(4.5.12) $$g_m = \sum_{i=1}^{n}(1-p/a_{ii})^m \quad ,$$

za svako $m, j = 1, 2, \dots$.

Posledica 4.5.1 Necentralna x^2-raspodela se može izraziti kao mešavina raspodela centralnih x^2-raspodela i Poisson-ove raspodele.

Posledice 4.5.2 Centralna x^2-raspodela se može izraziti u obliku

(4.5.11)

$$F_n(t) = p^{n/2} \sum_{j=0}^{\infty} \frac{n}{2}\left(\frac{n}{2}-1\right)\dots\left(\frac{n}{2}-j-1\right) / \ j\ ! \ (1-p)^j \ F_{n+2j}(t/p) \quad .$$

Teorema 4.5.3 Uz početne pretpostavke teoreme 4.5.1

I Kvadratna forma $X'AX$ ima za funkciju raspodele

$$H_{n;A,0}(t) = \sum_{j=0}^{\infty} c_j \, t^{n/2+j} \quad ,$$

koeficijenti c_j su dati sa

$$c_j = \frac{|A|^{-1/2} \ E(Q^* \ j)}{2^{n/2+2j} \ \Gamma(n/2+j)\, j\ !} \quad ,$$

gde smo označili sa

$$Q^* = X'A^{-1} X \quad .$$

Do sada smo pretpostavljali da je X bila standardizovana n-dimenzionalna slučajna promenljiva koja ima normalnu raspodelu.. Pretpostavimo sada da n-dimenzionalna slučajna promenljiva Y ima raspodelu $N(\mu, W)$, i potražimo funkciju raspodele kvadratne forme $Y'CY$. Bez gubitka u opštosti uzećemo da je $C = [c_{ij}]$ simetrična pozitivno defintna matrica reda $n \times n$. U protivnom možemo posmatrati matricu $C_1 = [(c_{ij} + c_{ji})/2]$ koja je simetrična, jer je

$$Y'CY = Y'C_1Y.$$

Zbog simetričnosti matrice W i C postoji matrica $B = B_1B_2$ takva da je

(4.5.12)

$$B = IB'W^{-1}$$

i

(4.5.13)

$$B'C\,B = A \quad ,$$

gde je A dijagonalna matrica čiji su elementi karakteristični koreni matrice $B_1'\ C\ B_1$, a B_2 je ortogonalna matrica, dok je B_1 takva matrica da je

$$B_2^1 W^{-1} B_2 = I \quad .$$

Ako izvršimo linearnu transformaciju

$$Y - \mu = BX \quad ,$$

dobijamo da je

$$P\{Y'CY \le t\} = P\left\{\left(X-b\right)' A(X-b) \le t\right\} \quad ,$$

gde smo stavili da je

(4.5.14)

$$b = -B^{-1} \ \mu \quad ,$$

gde X ima raspored $N(O,I)$, pa je dakle

$$P\ (Y'CY \le t) = H_{(n;A,b)}(t),$$

gde su A i b dati sa (4.5.13) i (4.5.14).

4.6 Ocena I^2- odstojanja

I^2 – odstojanje između statističkih skupova definisano je sa

(4.6.1)

$$I^2 = \sum_{i=1}^{n} \frac{\left(\mu_1^i - \mu_2^i\right)^2}{\sigma_i^2} \prod_{j=1}^{i-1}\left(1-r_{ij}\right) \quad ,$$

gde je

$\mu'_1 = \left(\mu_1^1 \ ,\ldots,\mu_1^k\right)$ sredina skupa π_1 , a

$\mu'_2 = \left(\mu_2^1 \ ,\ldots,\mu_2^k\right)$ sredina skupa π_2 ,

r_{ij} su koeficijenti korelacije između obeležja X_i i X_j ,

σ_i^2 varijansa obeležja X_i .

Za izvođenje ocene I^2– odstojanja dokazaćemo teoremu koja nam je potrebna u daljem radu.

<u>Teorema 4.6.1</u> Aritmetička sredina uzorka od n elemenata

(4.6.2)

$$\bar{x} = \frac{1}{n}\sum_{\alpha=1}^{n} x_\alpha$$

izvučenog iz k-dimenzijalnog osnovnog skupa $N(\mu, W)$, je nepristrasna ocena i ocena najveće verodostojnosti aritmetičke sredine μ , k-dimenzionalna slučajna promenjiva $\bar{x}$ ima raspodelu $N(\mu, \frac{1}{n} W)$

<u>Dokaz</u>. – Nepristrasnost ocene (4.6.2) sledi iz činjenice da je x_α , k-dimenzionalna slučajna promenljiva sa normalnom raspodelom $N(\mu,W)$ za svako $\alpha = 1, \dots, n$ i da je

$$E\left\{\sum_{i=1}^{n} x_\alpha\right\} = \sum_{i=1}^{n} E(x_\alpha) \quad ,$$

odnosno

$$E\left\{\sum_{i=1}^{n} x_\alpha\right\} = n\mu \quad .$$

Kako su x_α nezavisne k-dimenzionalne slučajne promenljive sa istom raspodelom $N(\mu, W)$, to je funkcija najveće verodostojnosti data sa

$$L = (2\pi)^{-kn/2}\, |W|^{-n/2} \exp\left\{ -\frac{1}{2} \sum_{\alpha=1}^{n} (x_\alpha - \mu)'\; W^{-1}\, (x_\alpha - \mu) \right\} \quad ,$$

gde je μ parametar koji treba odrediti tako da funkcija L definiše svoj maksimum. Funkcija $log\, L$ je rastuća funkcija od L pa ima maksimum u istoj tački u kojoj i funkcija L ima maksimum. Zato ćemo potražiti maksimum funkcije

(4.6.3)

$$Log\ L = -\frac{nk}{2}\log 2\pi - \frac{n}{2} log|W| - \frac{1}{2}\sum_{\alpha=1}^{n}(x_\alpha - \mu)' W^{-1}(x_\alpha - \mu)$$

u poslednjem članu u (4.6.3) dodavanjem i oduzimanjem u zagradama $\bar{x}$ dobijamo

$$\sum_{\alpha=1}^{n}(x_\alpha - \mu)' W^{-1}(x_\alpha - \mu) = \sum_{\alpha=1}^{n}(x_\alpha - \bar{x})' W^{-1}(x_\alpha - \bar{x}) + n(\bar{x} - \mu)' W^{-1}(\bar{x} - \mu) +$$

$$+(\bar{x} - \mu)' W^{-1}\sum_{\alpha=1}^{n}(x_\alpha - \bar{x}) + \left\{\sum_{\alpha=1}^{n}(x_\alpha - \bar{x})\right\}' W^{-1}(\bar{x} - \mu) \quad .$$

Iz (4.6.2) imamo da je

$$\sum_{\alpha=1}^{n} x_\alpha - n\bar{x} = 0 \quad ,$$

pa se (4.6.3) može napisati u obliku

(4.6.4)

$$logL = -k\frac{n}{2} - log 2\pi - \frac{n}{2} log\,|W\,| - \frac{1}{2}\sum_{\alpha=1}^{n}(x_\alpha - \bar{x})' W^{-1}(x_\alpha - \bar{x}) - \frac{n}{2}(\bar{x} - \mu)' W^{-1}(\bar{x} - \mu)$$

U ovom izrazu vidimo da samo poslednji član zavisi od μ, pa će funkcija $logL$ imati maksimum za ono μ za koje taj poslednji clan ima minimum. Matrica W je pozitivno definitna, pa kvadratna forma poslednjeg člana u (4.6.4) ima minimalnu vrednost sa

$$\bar{x} - \mu = 0 \quad ,$$

$$\bar{x} - \mu = 0 \quad ,$$

odakle sledi da je $\bar{x}$ ocena najveće verodostojnosti. Da bismo našli zakon verovatnoće promenljive $\bar{x}$, naći ćemo karakterističnu funkciju. Pošto je raspored od $x_\alpha : N(\mu, W)$ to je

$$\varphi_{x_\alpha}(n) = exp(i\,u'u - \frac{1}{2}u'Wu).$$

Zbog nezavisnosti promenljivih x_α dobijemo da je

$$\varphi_{\sum_{\alpha=1}^{n} x_\alpha}(u) = exp\left\{n\,i\,u'\mu - \frac{n}{2}u'W\,u\right\}$$

karakteristična funkcija promenljive $n\,\bar{x}$, pa je

$$\varphi_{\bar{x}} = \exp\left\{i\,u'\mu - \frac{1}{2}u'\left(\frac{1}{2}W\right)u\right\} \quad ,$$

Odakle sledi da $\bar{x}$ ima raspodelu $N\left(\mu, \frac{1}{2}W\right)$

Označimo sa

(4.6.5)

$$\alpha_i = \frac{1}{\sigma_i^2} \prod_{j=1}^{i=1} \left(1 - r_{ij}\right)$$

elemente dijagonalne matrice D. Tada I^2 odstojanje možemo napisati u obliku sledeće kvadratne forme

(4.6.6)

$$I^2 = \left(\mu_1 - \mu_2\right)' D\left(\mu_1 - \mu_2\right) \quad .$$

Pretpostavimo da su oba statistička skupa π_1 i π_2 normalno raspoređena sa poznatom zajedničkom dispersionom matricom, tj.

$$\pi_1 \ : N(\mu_1 \ ,W)$$
$$\pi_2 \ : N(\mu_2 \ ,W)$$

Za određivanje ocene I^2 odstojanja na osnovu dva uzorka iz skupova π_1 i π_2, iskoristićemo aritmetičke sredine uzoraka umesto aritmetičkih sredina osnovnih skupova, koje su njihove ocene najveće verodostojnosti.

Označimo sa

(4.6.7)

$$\overline{x}_1 = \frac{1}{n_1} \sum_{\alpha=1}^{m_1} x_\alpha$$

aritmetičku sredinu uzorka od n_1 elemenata iz skupova π_1 , a sa

(4.6.8)

$$\overline{x}_2 = \frac{1}{n_2}\sum_{\beta=1}^{n_2} x_\beta$$

aritmetičku sredinu uzorka od n_2 elemenata iz skupova π_2 .

Ocena I^2- odstojanja definisana je sa

(4.6.9)

$$\hat{I}^2 = (\hat{x}_1 - \hat{x}_2)' D (\hat{x}_1 - \hat{x}_2).$$

Teorema 4.6.2 $\hat{I}^2$ je nepristrasna ocena I^2- odstojanja.

Dokaz. - $\bar{x}_1$ i $\bar{x}_2$ su nezavisno k-dimenzionalne slučajne promenljive, pa iz teoreme 4.6.1 sledi

$$\overline{x}_1 \; : N(\mu_1 \; , \frac{1}{n_1} W)$$

$$\overline{x}_2 \; : N\left(\mu_2 \; , \frac{1}{n_2} W \right) \; ,$$

pa je

(4.6.10)

$$\overline{x}_1 - \overline{x}_2 \; : N\left(\mu_1 - \mu_2 \; , V \right) \; ,$$

gde smo kratkoće radi stavili da je

(4.6.11)

$$V = \left(\frac{1}{n_1} + \frac{1}{n_2} \right) W \quad .$$

Matrica V je simetrična matrica, pa postoji ortogonalna matrica C takva da je

$$C'VC = D_1 \quad ,$$

gde je D_1 dijagonalna matrica čiji su elementi karakteristični koreni matrice V.

Kada izvršimo linearnu transformaciju

$$(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2) = CY \quad ,$$

Dobićemo da je

(4.6.12)

$$Y \; : N(0, D_1) \quad ,$$

a $\hat{I}^2$ se svodi na

$$\hat{I}^2 = \left[CY + (\mu_1 - \mu_2)\right]' D \left[CY + (\mu_1 - \mu_2)\right] \quad ,$$

odnosno

(4.6.13)

$$\hat{I}^2 = Y'C'DCY + (\mu_1 - \mu_2)' DCY + Y'C'D(\mu_1 - \mu_2) + I^2 \quad .$$

Iz (4.6.12) sledi da su $Y_1, \ldots, Y_k$ nezavisne centrirane normalno raspoređene slučajne promenljive, pa je

$$E(Y_i) = 0$$

i

$$E(Y_i, Y_j) = 0 \quad ,$$

za svako $i, j = 1, \dots, k$, tako da je

$$E\left(Y'C'DCY\right) = 0$$

(4.6.14)

$$E[(\mu_1 - \mu_2)'DCY] = 0$$

$$E[Y'C'D(\mu_1 - \mu_2)] = 0,$$

Iz (4.6.13) i (4.6.12) sledi

$$E\left(\hat{I}^2\right) = I^2 \quad .$$

4.7. Raspodela $\hat{I}^2$ – odstojanja

<u>Teorema4.7.1</u> Neka su π_1 i π_2 k-dimenzionalni statistički skupovi koji imaju normalnu raspodelu sa sredinama $\mu_1 \neq \mu_2$ i zajedničkom disperzionom matricom W.

I Funkcije raspodele $\hat{I}^2$ – odstojanja mogu se izraziti u vidu sledećeg uniformno konvergentnog reda

(4.7.1)

$$P\left\{\hat{I}^2 \leq t\right\} = \sum_{j=0}^{\infty} c_j \, F_{n+2j}\left(t / p\right) \quad , \qquad t < \infty \quad ,$$

gde je $p > 0$ unapred data konstanta, $F_n(t)$ je x^2 funkcija raspodele sa n stepena slobode, koeficijenti c_j dati su sa

(4.7.2)

$$c_j = |A|^{-\frac{1}{2}} e^{-\frac{b'b}{2}} \, p^{\frac{n}{2}+j} \, E\left\{Q^j \, H_{2j}\left(L / Q^{\frac{1}{2}}\right)\right\} / \left(2j\right)! \quad ,$$

gde smo označili sa

$$L = b' A^{-1/2} Y$$

(4.7.3)

$$Q = Y' \left(A^{-1} - \frac{1}{2} I \right) Y \quad ,$$

a Y je k-dimenzionalno standardizovana normalno raspoređena slučajna promenljiva.

$$A = C' D C$$

(4.7.4)

$$b = -C^{-1} \left(\mu_1 - \mu_2 \right) \quad ,$$

a C je matrica takva da je

(4.7.5)

$$C' V^{-1} C = I$$

$$V = \left(\frac{1}{n_1} + \frac{1}{n_2} \right) W \quad .$$

II Fukcija generatrise niza $\{c_j\}$ je

(4.7.6)

$$\psi(z) = -\prod_{i-1}^{n} \left\{ \left(p / a_{ii} \right)^{\frac{1}{2}} \left[\left(1 - p / a_{ii} \right) z \right]^{-\frac{1}{2}} \right\} \quad .$$

III Koeficijenti c_j zadovoljavaju rekurentnu vezu

$$c_0 = e^{-b'b/2} \prod_{i=1}^{n} \left(p / a_{ii} \right)^{\frac{1}{2}}$$

$$c_j = \frac{1}{2J} \sum_{r=0}^{J-1} g_{j-r} \, c_r \quad ,$$

gde smo stavili da je

$$g_m = \sum_{i=1}^{k} \left(1 - p / a_{ii}\right)^m + mp \sum_{i=1}^{k} \frac{b_i^2}{a_{ii}} \left(1 - p / a_{ii}\right)^{m-1}$$

za svako $m, j = 1\,, 2, \ldots$.

Dokaz. Iz (3.10), koristeći oznake iz (4.5) dobijamo da je zakon verovatnoće k-dimenzionalne slučajne promenljive $\bar{x}_1 - \bar{x}_2$

(4.7.7)

$$f\left(\bar{x}_1 - \bar{x}_2\right) = c_k \exp\left\{ \left[\bar{x}_1 - \bar{x}_2 - \left(\mu_1 - \mu_2\right) \right]' V^{-1} \left[\bar{x}_1 - \bar{x}_2 - \left(\mu_1 - \mu_2\right) \right] \right\} \quad ,$$

gde smo kratkoće radi stavili da je

$$C_k = (2\pi)^{(-k/2)} \, | V |^{(-1/2)} \, .$$

Kako je V simetrična matrica, to postoji matrica C_1 takva da je

$$C_1' V^{-1} C_1 = I \quad .$$

Isto tako zbog simetričnosti matrice $C'_1 \, D \, C_1$ postoji ortogonalna matrica C_2 takva da je $C'_2 \, (C'_1 \, D \, C_1) C_2 = A$, gde je A dijagonalna matrica čiji su elementi karakteristični koreni matrice $C'_1 \, D \, C_1$.

Označimo sa

$$C = C_1 C_2 \quad ,$$

i izvršimo linearnu transformaciju

(4.7.8)

$$\overline{x}_1 - \overline{x}_2 - (\mu_1 - \mu_2) = CY \quad ,$$

Dobićemo da je zakon verovatnoća za Y

(4.7.9)

$$f(y) = (2\pi)^{-k/2} \exp\left\{-\frac{1}{2} y'y\right\} \quad ,$$

a $\hat{I}^2$ se svodi na

(4.7.10)

$$\hat{I}^2 = \left[CY + (\mu_1 - \mu_2)\right]' D \left[CY + (\mu_1 - \mu_2)\right] \quad .$$

Kad u (4.7.10) uvrstimo oznake date u (4.7.4) imaćemo

(4.11)

$$\hat{I}^2 = (Y - b)' A (Y - b) \quad .$$

Iz (4.7.9) i (4.7.11) vidi se da su ispunjeni uslovi teoreme 2.1, pa je

(4.7.12)

$$P\left\{\hat{I}^2 \le t\right\} = P\left\{(Y - b)' A\, (Y - b) \le t\right\} = H_{k;A,b}(t) \quad ,$$

odakle se dobija (4.7.1), (4.7.6) i rekurentne veze za članove niza $\{c_j\}$.

Teorema 4.7.2 Uz početne pretpostavke teoreme 4.1, funkcija raspodele $\hat{I}^2$ – odstojanja može se izraziti preko stepenog reda

(4.7.13)

$$P\left\{\hat{I}^2 \leq t\right\} = \sum_{j=0}^{\infty} c_j^* \, t^{\frac{k}{2}+j} \quad .$$

Koeficijenti c_j^* su dati sa

(4.7.14)

$$c_j^* = |A|^{-\frac{1}{2}} e^{-b'b} \frac{E\left\{Q^{*j} H_{2j}\left(L / Q^{*\frac{1}{2}}\right)\right\}}{\Gamma\left(\frac{k}{2}+j+1\right) 2^{\frac{k}{2}+j} (2j)!} \quad ,$$

gde smo stavili da je

(4.7.15)

$$Q^* = Y'A^{-1}Y \quad ,$$

a Y, A, b su dati kao i u teoremi 4.1.

Teorema 4.7.3 Neka su statistički skupovi π_1 i π_1 normalno raspoređeni sa jednakim aritmetičkim sredinama $\mu_1 = \mu_2$, i zajedničkom dispersionom matricom W.

I Funkcija raspodele $\hat{I}^2$ odstojanja može se izraziti u vidu sledećeg uniformno konvergentnog reda

(4.7.16)

$$P\left\{\hat{I}^2 \leq t\right\} = \sum_{j=0}^{\infty} h_j \, F_{k+2j}(t / p) \quad , \qquad t < \infty \quad ,$$

gde je $p > 0$ unapred data konstanta, a koeficijenti h_j su dati sa

(4.7.17)

$$h_j = |A|^{-\frac{1}{2}} p^{\frac{k}{2}} E\left\{(-Q)^j\right\} /(2^j \ j\,!) \quad ,$$

a A i Q su dati sa (4.7.5) i (4.7.4).

II Funkcija generatrise za niz $\{h_j\}$ je

(4.7.18)

$$\psi(z) = \prod_{i=1}^{k} \left\{ (p / a_{ii})^{\frac{1}{2}} [1 - (1 - p / a_{ii}) z]^{-\frac{1}{2}} \right\} \quad .$$

III Članovi niza $\{h_j\}$ zadovoljavaju rekurentnu vezu

(4.7.19)

$$h_0 = \prod_{i=1}^{k} (p / a_{ii})^{1/2}$$

$$h_j = \frac{1}{2j} \sum_{r=0}^{\infty} g_{j-r} c_r \quad ,$$

gde smo označili sa

$$g_m = \sum_{i=1}^{k} (1 - p / a_{ii})^m \quad ,$$

za svako $m, j = 1, 2, \ldots$

Teorema 4.7.4 Uz početne pretpostavke teoreme 4.7.3

I Funkcija raspodele $\hat{I}^2$ – odstojanja se može izraziti preko stepenog reda

(4.7.20)

$$P\left\{\hat{I}^2 \leq t\right\} = \sum_{j=0}^{\infty} h_j^* \, t^{\frac{k}{2}+j}$$

(4.7.21)

$$h_j^* = |A|^{-\frac{1}{2}} \frac{E\left\{Q^{*j}\right\}}{2^{\frac{k}{2}+2j}\,\Gamma\left(\frac{k}{2}+j\right) j!}$$

II Momente od Q^* izračunavamo iz kumulanti koje su date sa

(4.7.22)

$$K_l = 2^{l-1}(l-1)! \sum_{j=1}^{k} \left(1/a_{jj}\right)^l$$

za $l = 1, 2, \dots$

Da bismo testirali hipotezu da je I^2 – odstojanje između statističkih skupova π_1 i π_1 jednako nuli, potrebno je prvo odrediti matricu C, koja svodi dispersionu matricu $\left(\frac{1}{n_1}+\frac{1}{n_2}\right)W$ na jediničnu, zatim izračunati A i b, pa upotrebom rezultata teorema 4.7.3 i 4.7.4 napraviti tablice za razne vrednosti t. Iz tako dobijenih tablica možemo odrediti t_1 za koje je

$$P\left\{\hat{I}^2 \leq t_1\right\} = \alpha \quad ,$$

i izracunavanjem $\hat{I}^2$ – odstojanja između dva uzorka i videti da li je $\hat{I}^2 \leq t_1$. Ukoliko je $\hat{I}^2 \leq t_1$ prihvatamo hipotezu da je između skupova π_1 i π_1 $I^2 = 0$, sa rizikom α .

4.8. Raspodela $\hat{I}^2$ – odstojanja za slučajne promenljive koje nemaju normalnu raspodelu

Postavlja se pitanje da li Ivanovićevo odstojanje pokazuje slaganje sa normalnom raspodelom, ako se ceo postupak primeni nad vektorima koji imaju različite teoretske raspodele. Naime, prilikom izrade same disertacije, kao i u mnogim drugim problemima koji su rešeni i publikovani u raznim časopisima, proverom rezultata je utvrđeno da je to uvek bio slučaj. Za proveru ovih rezultata korišćena je Bootstrap metoda koja kao svoj sastavni deo podrazumeva primenu Monte-Karlo simulacije. U tabeli su prikazani rezultati generisanih 6 varijabli koje imaju Normalnu, Uniformnu, Eksponenicijalnu, Vejbulovu, Binomnu i Puasonovu raspodelu, izvršeno je reuzorkovanje, a zatim je primenjena metoda I-odstojanja. Slaganje dobijenih vrednosti I-odstojanja sa teoretskom normalnom raspodelom provereno je Kolmogorov- Smirnov testom.

Tabela 4.8.1. Generisan 6-dimvektor različitih raspodela na uzorku obima 100

n5	u5	e5	w5	b5	p5
-1.75339	0.626924	0.177863	0.301622	0.465	1.1
-2.25271	0.632643	1.615341	2.926521	0.475	1.4
-0.26238	0.610855	0.081644	1.241879	0.475	0.7
-0.16917	0.30247	0.484975	1.124464	0.56	1.1
1.671215	0.725348	0.83921	3.72023	0.47	0.7
0.414999	0.875224	0.195324	2.429727	0.525	0.8
0.03611	0.472922	0.663404	2.258105	0.455	0.6
0.063654	0.807523	0.483084	1.443432	0.555	0.8
0.364664	0.912926	0.39958	1.254291	0.48	0.7

-1.38027	0.091059	0.51932	0.480689	0.495	1.6
-0.01179	0.429779	1.165284	0.589863	0.465	0.6
0.562083	0.04402	0.01407	1.588781	0.485	0.9
-0.13827	0.216165	0.25725	1.044094	0.535	0.8
-0.62259	0.020165	1.164491	1.306678	0.52	0.7
-1.05352	0.808375	2.913404	0.921062	0.535	1.3
1.038854	0.212295	0.639777	1.875857	0.555	1
-0.0518	0.06001	1.494142	0.120983	0.46	1.3
1.400479	0.758669	0.741193	1.617622	0.52	0.9
-0.36409	0.639397	0.808959	2.172128	0.48	0.8
-1.5172	0.83731	0.00451	1.535075	0.47	0.6
0.042431	0.585334	0.042605	0.476529	0.455	0.6
-0.90987	0.134957	0.611852	1.037373	0.495	1.5
0.351904	0.357952	0.780203	1.128142	0.53	0.5
1.456563	0.784756	0.750087	0.766682	0.56	0.9
0.176665	0.195252	1.411098	0.034248	0.525	0.9
-0.35957	0.051791	2.596465	0.258563	0.48	1.3
-0.19255	0.426109	0.077334	0.606296	0.475	0.6
0.642001	0.175831	0.057972	2.155949	0.5	1.1
-1.21661	0.137249	0.470219	3.46435	0.485	1.1
0.244878	0.54512	0.3784	0.583716	0.55	1.1
-0.35804	0.390826	2.110279	0.217488	0.45	0.9
0.818553	0.564295	0.845545	0.550368	0.48	0.7
1.142504	0.341865	0.827669	0.198718	0.435	1
0.520532	0.482867	0.130309	0.749208	0.475	0.7
-0.64365	0.673577	2.807745	2.752805	0.43	1.7
0.032486	0.26275	2.219819	1.500497	0.525	0.7
-0.5059	0.897529	1.118264	0.15804	0.545	0.8
-0.29306	0.680849	0.11104	0.125645	0.47	1.3
-0.40705	0.967427	2.387315	1.363066	0.5	0.9
-0.98412	0.487267	3.652764	0.285268	0.555	0.8
-0.97286	0.07577	0.287247	0.321334	0.555	1.8

0.484831	0.695323	1.107877	0.710162	0.475	1
0.912072	0.646955	0.123808	0.144192	0.52	0.6
-0.91143	0.114855	0.118382	0.375365	0.48	0.8
0.874737	0.967848	1.299402	1.031177	0.455	1.3
-0.74526	0.433299	0.433763	2.384951	0.5	0.9
1.770523	0.445762	1.076407	1.453648	0.505	1
-0.73956	0.568904	0.258441	0.459053	0.535	0.7
0.951727	0.613991	0.840751	0.395517	0.53	1.6
-1.13569	0.000711	0.675719	6.680855	0.515	0.8
1.945534	0.575374	2.231271	0.013722	0.53	1.6
-0.00827	0.611708	1.965516	0.593062	0.54	0.7
-0.44014	0.580394	1.521526	1.868735	0.525	0.7
-0.3196	0.453393	0.173667	1.221573	0.515	1
-1.41902	0.195683	0.148922	0.298188	0.525	0.9
-1.61751	0.996897	0.291631	0.730663	0.55	1.2
1.161441	0.16975	1.153077	0.068053	0.435	1.2
-0.74859	0.159351	0.852687	0.372888	0.515	0.6
1.030498	0.198926	2.097089	0.083124	0.5	1.1
-0.52548	0.692451	0.678964	1.135819	0.54	0.7
0.355312	0.971618	0.549736	0.753913	0.545	0.4
-0.68183	0.84961	1.580963	3.190036	0.485	1
0.178896	0.928559	0.625311	1.694759	0.535	1.2
-2.05211	0.844705	1.601454	0.914183	0.535	0.6
0.968116	0.585789	0.727118	1.085728	0.47	0.6
1.428756	0.077135	3.376917	0.974836	0.48	0.8
1.111868	0.959052	0.208605	0.077619	0.465	0.7
-1.27097	0.487447	0.238329	0.548033	0.445	0.8
0.538817	0.801987	5.711671	0.634925	0.505	0.8
-0.37358	0.326449	0.032944	0.67877	0.48	1.4
-1.3701	0.05815	0.514386	1.182574	0.515	1.4
-0.66284	0.263534	2.469349	0.069637	0.46	1.3
2.089505	0.778298	0.074462	1.286895	0.5	1.1

-0.20196	0.186518	0.273922	0.850803	0.555	0.8
-0.30557	0.665182	1.084587	0.125779	0.585	1.3
0.521891	0.419638	2.657096	0.92717	0.515	1.2
0.124906	0.397892	1.967589	2.88542	0.495	1
0.170974	0.773245	0.54568	4.149802	0.49	1.7
-1.53328	0.794008	0.61123	0.371028	0.535	1.1
2.118307	0.544018	1.304914	2.678271	0.51	0.7
-0.7505	0.39935	0.080444	3.995897	0.495	0.9
1.611544	0.449483	0.520052	3.462098	0.47	0.7
-2.07481	0.664344	0.950779	0.155709	0.46	0.8
0.088108	0.343739	0.832299	2.315025	0.515	0.7
0.794812	0.794632	0.689579	3.654226	0.475	1
-0.2324	0.518531	1.050119	0.506351	0.515	1.2
-1.17385	0.598801	0.502224	0.907931	0.475	0.9
0.580255	0.379871	1.806114	0.663292	0.52	0.9
0.667453	0.569105	1.299358	0.005155	0.485	1.2
-0.05323	0.829714	0.53497	0.078022	0.47	0.6
-0.18561	0.144499	0.611325	0.235	0.51	0.8
7.63E-05	0.371541	0.921152	2.911821	0.525	1.1
1.039642	0.73027	0.459066	0.677851	0.5	1.7
-1.98504	0.787752	0.223296	0.161754	0.45	1
0.261119	0.781441	0.111615	0.338301	0.49	1.5
1.271615	0.615541	1.024582	1.484942	0.525	1.8
2.571467	0.349059	1.981766	0.394421	0.495	0.9
0.806993	0.274256	1.624739	0.556002	0.475	1
1.244	0.161299	0.657328	1.271586	0.475	0.6
0.176472	0.59695	0.02556	1.599889	0.485	0.8

Tabela 4.8.2 Test slaganja I-odstojanja sa normalnom raspodelom

Kolmogorov-Smirnov Test			
		I2_MIN1	I2_MIN2
Veličina uzorka		100	100
Parametri raspodele	Sredina	22.6566	22.8980
	St.. devijacija	11.47773	11.55279
Kolmogorov-Smirnov test statistika		1.099	1.039
signifikantnost		.178	.231

Prethodna simulacija je pokazala slaganje sa normalnom raspodelom (signifikantnost =0.231) za pocetnih 6 varijabli koje su imale različite teoretske raspodele.

5. MODEL STRUKTURNE KORELACIONE ANALIZE ZASNOVAN NA VEKTORSKIM KOEFICIJENTIMA KORELACIJE

Istraživači u organizaciji ponekad su zainteresovani za testiranje zavisnih ili nezavisnih koeficijenata korelacije kada su oni jednaki. Olkin, Finn i Steiger predložili su nekoliko statističkih procedura za testiranje zavisnih koeficijenata korelacije u pojedinačnoj grupi i, gde meta-analitičke procedure mogu biti korišćene u testiranju nezavisnih koeficijenata korelacije u dve ili više grupa. Zbog česte uključenosti kompjuterskog programiranja, sprovođenje ovog istraživanja može biti otežano, posebno pri testiranju zavisnih koeficijenata korelacije.

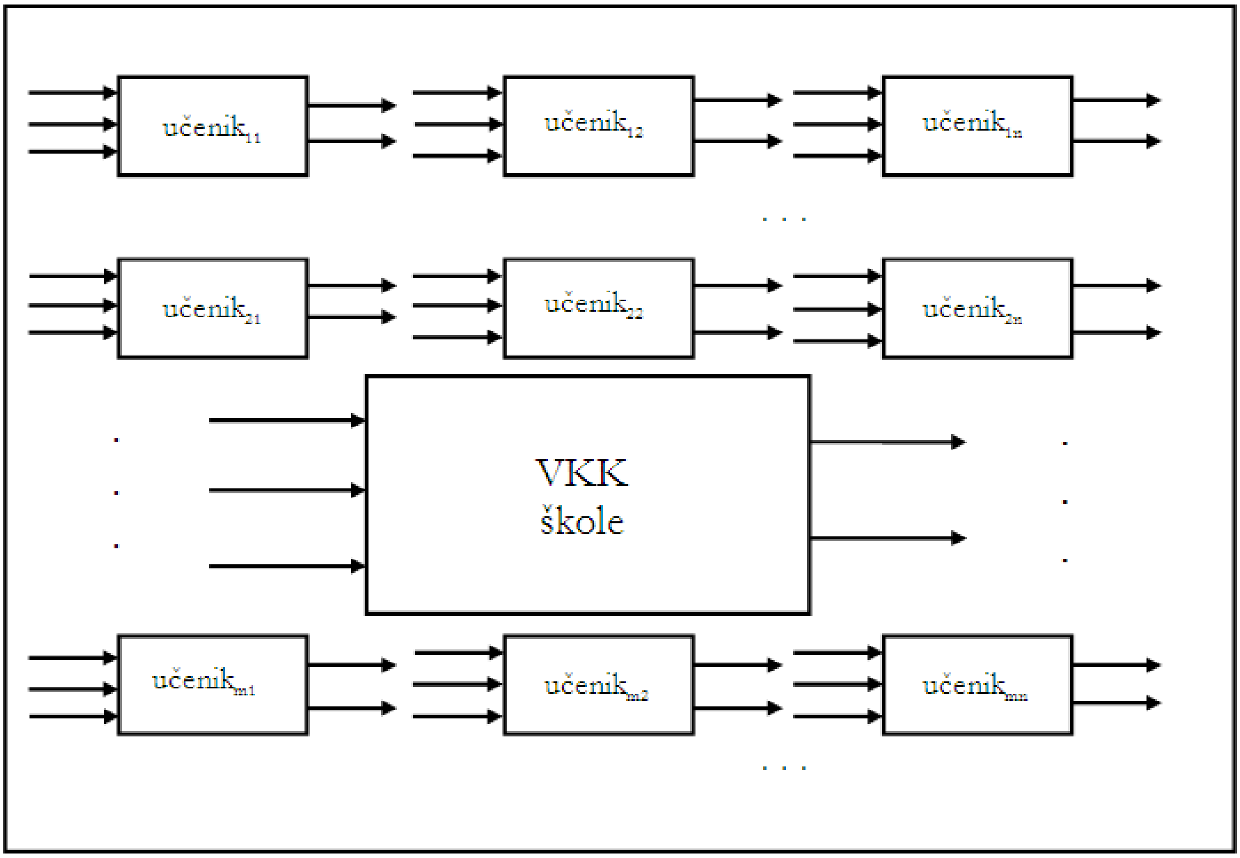
Generalno, sprovedena istraživanja su se uglavnom interesovala za testiranje hipoteza na standardnim metrikama (korelaciona matrica), pre nego metrike vrednosti reda (matrice kovarijanse). Na primer Turnley, Bolino, Laster i Bloodgood, (2003) koristeći Steigerovu formulu (1980) na testiranju zavisnih koeficijenata korelacije, pronašli su to da je korelacija između psihološke ispunjenosti i organizacionog ponašanja (organizacionih navika) ciljanih u organizaciji bila jača nego psihološke ispunjenosti i organizacionog ponašanja (navika) usmerenih individualno.

Zbog odstupanja između organizacionog ponašanja usmerenih ka organizaciji i individualno, razlikuju se (0.49 prema 0.35), jednakost u kovarijansama ne znači ujedno i snagu veze u isto vreme. Takva komparacija korelacionih koeficijenata je češće birana i više interpretirana u mnogim situacijama (Hunter & Hamilton, 2002, Hunter & Schmidt, 1990).

Nekoliko statističkih procedura je bilo predloženo da testiraju koeficijente korelacije (npr. Olkin & Finn, 1990, 1995, Steiger, 1980). Međutim istraživanja nije lako sprovesti, zato što je najčešće uključeno kompjutersko programiranje (npr. Graf & Alf, 1999). Na testiranju između grupa nezavisnih korelacionih koericijenata, meta-analitičke procedure kao Hadges i Olkin (1985) i Hunter i Schmidt (1990) mogu biti korišćene. Zavisni i nezavisni koeficijenti korelacije se često tretiraju različito.

U disertaciji je glavni akcenat dat na testiranju nezavisnosti vektorskog koeficijenta korelacije, odnosno međusobnom poređenju dva vektorska koeficijenta korelacije. Treba napomenuti da je analiza rađena za složenu korelacionu strukturu, obzirom da je korelaciona veza posmatrana između više ulaznih, odnosno izlaznih veličina. Samo izračunavanje vektorskog koeficijenta korelacije je jako složeno, pa je korišćena odgovarajuća aplikacija napisana u SPSS MATRIX jeziku. U opštem slučaju je to problem koji ima m ulaznih, odnosno n izlaznih varijabli, dok smo se u našem istraživanju zadržali na problemu od 3 ulazne i 2 izlazne veličine, što ne umanjuje opštost računanja i zaključaka koji iz toga mogu nastati. Na slici 5.1 data je jedna struktura na osnovu koje se može odrediti vektorski koeficijent korelacije.

Upravo ovakav model će biti osnova za rešavanje problema u obrazovanju, gde želimo sagledati sve dobre i loše strane škola koje su bile predmet našeg istraživanja, a sve u cilju podizanja kvaliteta obrazovanja na jedan viši nivo.



Slika 5.1 Prikaz složene korelacione strukture na primeru škole

Obrazovanje je u svakom društvu veoma bitan segment u koji neprestano treba ulagati. Posebno je bitno obavezno obrazovanje, dakle period od prvog do osmog razreda. Zato je neophodno uraditi presek postojećeg stanja u našem školstvu i jasno definisati kriterijume za evaluaciju stečenih znanja i kvaliteta nastavnog procesa. Pručavajući uspeh učenika na prijemnom ispitu u srednjim školama u poslednjih 7 godina, došlo se do zaključka da bi izrada lične karte za sve osnovne škole u Srbiji bilo od krucijelnog značaja za podizanje nivoa uspešnosti učenika, škola i društva u celini. Na taj način bi se po tačno utvrđenim kriterijumima znale sve bitne informacije za jednu školu, njene učenike i nastavno osoblje, i sve informacije bi bile javno dostupne. Da bi došlo do poboljšanja rada škola, mora postojati vera da je to zaista moguće. Proces rangiranja treba da skrene pažnju na to šta se zaista može postići i treba da da opipljive dokaze za to, i da jasno pokaže čak iako se uzmu faktori kao što je socijalni status učenika, obrazovanje njegovih roditelja, za koje mnogi misle da diktiraju način koliko će učenik imati uspeha u školi, neke škole su ipak uspešnije od drugih (Rutter et all, 1979).

Ovaj nalaz potvrđuju istraživanja sprovedena u različitim zemljama. Na taj način, ni roditelji ni nastavnici se neće iznenaditi činjenicom da to što se dešava u školi ima najveći uticaj na uspeh učenika u istoj. Izradom integralne lične karte osnovnih škola u Srbiji, možemo upoređivati sadašnje rezultate škole sa onima od prošle godine i na taj način videti da li se rad škole poboljšava. Upoređivanjem sa školama u okolini možemo prepoznati one koje su uspešnije i naučiti nešto od njih. Ukupne rezultate jedne škole treba postaviti na njeno mesto u odnosu na sve škole u sistemu. U okviru lične karte škole izveštaj rada škole bi trebalo da bude dokument koji sadrži sve bitne i objektivne pokazatelje rada škole, koji će biti lako dostupan javnosti, tako da svako može da analizira i upoređuje rad škola. Na ovaj način lična karta škole, odnosno integralna lična karta svih škola u zemlji bi pomogla roditeljima da odaberu pravu školu za svoje dete, ali i da ohrabri druge škole da poboljšaju svoj rad. Tako bi se stvorila zdrava konkurencija između škola, čime bi i nastavno osoblje i učenici imali jasnu sliku o svom mestu na hijerarhijskoj lestvici uspeha. Naravno, da bi se sve ovo sprovelo u praksi, jedan od prvih koraka mora biti učinjen od strane resornog ministarstva, koje bi zajedno sa ostalim zainteresovanim stranama, definisalo i ustanovilo kriterijume za rangiranje osnovnih škola.

U integralnoj ličnoj karti osnovnih škola u Srbiji centralno mesto bi zauzimala statistička analiza rezultata škola. Ona bi bila bazirana na rangiranju škola, efikasnosti škola i utvrđivanju funkcionalne pismenosti učenika. Sve ovo je moguće postići u jednom kompleksnom strukturnom modelu zasnovanom na vektorskom koeficijentu korelacije uz integraciju sa DBA (Jeremić, 2012) i DEA metodom.

5.1. Testiranje hipoteze o jednakosti dva vektorska koeficijenta korelacije

Istraživači u različitim oblastima vrlo često dolaze u situaciju da kada utvrde stepen korelacije između dve posmatrane veličine pokažu da li između njih postoji statistički značajna razlika. Ta informacija je vrlo često od presudnog značaja zbog interpretacije dobijenih rezultata, kao i donošenja pravih odluka koje su bitne za ostvarivanje ciljeva istraživanja. Osnova za definisanje funkcije za testiranje hipoteze je predstavljala hipoteza o testiranju nezavisnosti vektorskog koeficijenta korelacije. Sa druge strane, rezultati su primenjeni i evaluirani kroz praćenje rezultata učenika osnovnih škola u Srbiji koje su oni postigli na prijemnom ispitu prilikom upisa u srednje škole, a uzimajući broj poena koji su postignuti na testu iz matematike i srpskog jezika. Na ovaj način smo u mogućnosti da uporedimo dve škole, tj. koliko se razlikuju dve škole u tome u kojoj su meri učenici u sposobnosti da znanje koje su stekli kroz osmogodišnje školovanje pokažu na prijemnom ispitu. Izračunati vektorski koeficijent korelacije nam pokazuje stepen povezanosti postignutog uspeha na prijemnom ispitu sa uspehom koji su učenici postigli u školi. Kasnije će biti prikazan jedan način rangiranja škola uz primenu metode I-odstojanja na varijablama koje se zakonski regulisane prilikom polaganja prijemnog ispita za upis u srednje škole.

Posmatraćemo prost slučajan uzorak veličine N. Ako na elementima tog uzorka merimo obeležja **Y** i **X**, pri čemu je **Y** m-dimenzionalna aleatorna promenljiva, a **X** n-dimenzionalna slučajna promenljiva, onda se moguće vrednosti uzorka mogu dati u vidu jedne matrice reda $N \times (m+n)$

$$[\boldsymbol{Y}, \boldsymbol{X}] = \begin{bmatrix} \mathrm{Y}_{11} & \mathrm{Y}_{21} & \dots & \mathrm{Y}_{\mathrm{m}1} & \mathrm{X}_{11} & \mathrm{X}_{21} & \dots & \mathrm{X}_{\mathrm{n}1} \\ \mathrm{Y}_{12} & \mathrm{Y}_{22} & \dots & \mathrm{Y}_{\mathrm{m}2} & \mathrm{X}_{21} & \mathrm{X}_{22} & \dots & \mathrm{X}_{\mathrm{n}2} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \mathrm{Y}_{1\mathrm{N}} & \mathrm{Y}_{2\mathrm{N}} & \dots & \mathrm{Y}_{\mathrm{mN}} & \mathrm{X}_{1\mathrm{N}} & \mathrm{X}_{2\mathrm{N}} & \dots & \mathrm{X}_{\mathrm{nN}} \end{bmatrix}$$

Označimo sa **S** disperzionu matricu uzorka

$$S = \begin{bmatrix} S_{yy} & S_{yx} \\ S_{xy} & S_{xx} \end{bmatrix}$$

Generalizovana varijansa uzorka za m-dimenzionalnu promenljivu Y je determinanta $\left|\boldsymbol{S}_{yy}\right|$, za n-dimenzionalnu promenljivu X je determinanta $\left|\boldsymbol{S}_{xx}\right|$, a za (m+n)-dimenzionalnu slučajnu promenljivu je determinanta $\left|\boldsymbol{S}\right|$**.**

Analogno definiciji vektorskog koeficijenta korelacije populacije, možemo definisati vektorski koeficijent korelacije uzorka izrazom

$$R_v^2 = (1 - \frac{|S|}{|S_{yy}||S_{xx}|})$$

koji ce predstavljati ocenu vektorskog koeficijenta korelacije populacije.

Ako raspodela uzorka iz višedimenzionalne populacije zavisi od vektora parametra $\boldsymbol{\Theta}$ i ako su hipoteze:

H_0: $\boldsymbol{\Theta} \in \boldsymbol{\Omega}_0$ i alternativna H_1: $\boldsymbol{\Theta} \in \boldsymbol{\Omega}_1$,

tada je statistika količnik verodostojnosti (likelhood ratio – LR) za testiranje H_0 , definisana kao

$$\lambda = \frac{L_0^*}{L_1^*},$$

gde je L_i^* maksimalna vrednost funkcije verodostojnosti u regionu $\boldsymbol{\Omega}_i$, (i=0,1).

Test količnika verodostojnosti (The Likelihood Ratio Test– **LRT**) sa nivoom značajnosti α, za testiranje H_0 protiv alternativne H_1, određen je regionom

odbacivanja $\boldsymbol{R} = \{\boldsymbol{x} \mid \lambda(\boldsymbol{x}) < c\}$ gde je c određeno tako da je $\sup_{\Theta\in\Omega_0} P_\theta\{\boldsymbol{x} \in \boldsymbol{R}\} = \alpha$. Neka je uzorak izvučen iz p-dimenzionalne Normalne raspodele $N_p(\mu, W)$ i neka je $\hat{\boldsymbol{W}}$ ocena maksimalne verodostojnosti za **W** kada je H_0 tačna, **S** ocena maksimalne verodostojnosti za **W** kada je H_1 tačna, a $\overline{\boldsymbol{x}}$ je ocena sredine $\boldsymbol{\mu}$ u obe hipoteze.
Statistika za testiranje H_0 protiv H_1 definisana je izrazom:

$$-2\log\lambda = Np(A - \log G - 1)$$

pri čemu je N obim uzorka, dok su A i G aritmetička i geometrijska sredina karakterističnih vrednosti matrice $\hat{\boldsymbol{W}}^{-1}\boldsymbol{S}$.

Podelom p-dimenzionalnog vektora **Z** na podvektore **Y** (dimenzije m) i **X** (dimenzije n), p=m+n, iz disperzione matrice uzorka **S** podeljene na odgovarajuće podmatrice, odrediće se ocena vektorskog koeficijenta korelacije iz izraza

$$R_v^2 = \left(1 - \frac{|S|}{|S_{yy}||S_{xx}|}\right)$$

Želimo testirati hipotezu o nezavisnosti vektora Y i X, tj. hipotezu da je vektorski koeficijent korelacije jednak nuli:

$$H_0(R_v = 0)$$

Ova hipoteza je ekvivalentna hipotezi

$$H_0(W_{YX} = 0)$$

Kad je hipoteza H_0 tačna, ocena disperzione matrice data je sa

$$\hat{\boldsymbol{W}} = \begin{bmatrix} \boldsymbol{S}_{yy} & 0 \\ 0 & \boldsymbol{S}_{xx} \end{bmatrix},$$

pa je

$$\hat{\boldsymbol{W}}^{-1}\boldsymbol{S} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{S}_{yy}^{-1}\boldsymbol{S}_{yx} \\ \boldsymbol{S}_{xx}^{-1}\boldsymbol{S}_{xy} & \boldsymbol{I} \end{bmatrix},$$

tako da je

$$A = \frac{1}{2} tr(\hat{\boldsymbol{W}}^{-1}\boldsymbol{S}) = 1$$

$$G^P = \left|W^{-1}S\right| = \left|I - S_{yy}^{-1}S_{yx}S_{xx}^{-1}S_{xy}\right| = 1 - R_v^2$$

Statistika koja testira vektorski koeficijent korelacije data je sa

$$-2\log\lambda = -N\log(1 - R_v^2) = -N\log\prod_{i=1}^{m}\lambda_i$$

pri čemu su λ_i karakteristični koreni matrice

$$\left|I - S_{yy}^{-1}S_{yx}S_{xx}^{-1}S_{xy}\right| \text{ i } m \le n$$

Kad je hipoteza H_0 tačna, statistika $-2\log\lambda$ će imati malu vrednost, u protivnom vrednost statistike biće velika. Raspodela statistike $\lambda^{\frac{2}{N}} = (1 - R_v^2)$ data je tzv. Wilks-ovom raspodelom

$$\Lambda(n, N-1-m, m),$$

a aproksimativna raspodela je data preko Bartlett-ove aproksimacije:

$$-[N-\frac{1}{2}(m+n+3)]\log(1-R_v^2) \approx \chi^2_{mn}$$

tako da će se $\chi^2 - raspodela$ sa $m \cdot n$ stepeni slobode, koristiti za testiranje hipoteze H_0. Oblast prihvatanja ili odbacivanja hipoteze H_0 određena je na osnovu zaključka:

- kada je H_0 tačna, statistika ima malu vrednost,
- a kada je tačna alternativna hipoteza H_1, statistika će težiti većoj vrednosti.

Radi lakšeg razumevanja i pravljenja razlike u odnosu na kanoničku korelacionu analizu u daljem radu ćemo smatrati da je $R_v^2 = VKK$.

Ako dva nezavisna uzorka dolaze iz populacije sa $N(\mu, \Sigma)$, za izračunate vektorske koeficijente korelacije $VKK_1 \; i \; VKK_2$ treba testirati hipotezu

$$H_0(VKK_1 = VKK_2)$$

naspram alternativne

$$H_1(VKK_1 > VKK_2)$$

Uzimajući prethodno, statistika za poređenje dva vektorska koeficijenta korelacija predstavljaće količnik dve promenljive koje imaju $\chi^2 - raspodelu$, pa ćemo imati

$$\tau = \frac{\dfrac{-(N_1 - \frac{1}{2}(m_1 + n_1 + 3))\log(1 - \hat{VKK}_1^2)}{m_1 n_1 - 1}}{\dfrac{-(N_2 - \frac{1}{2}(m_2 + n_2 + 3))\log(1 - \hat{VKK}_2^2)}{m_2 n_2 - 1}},$$

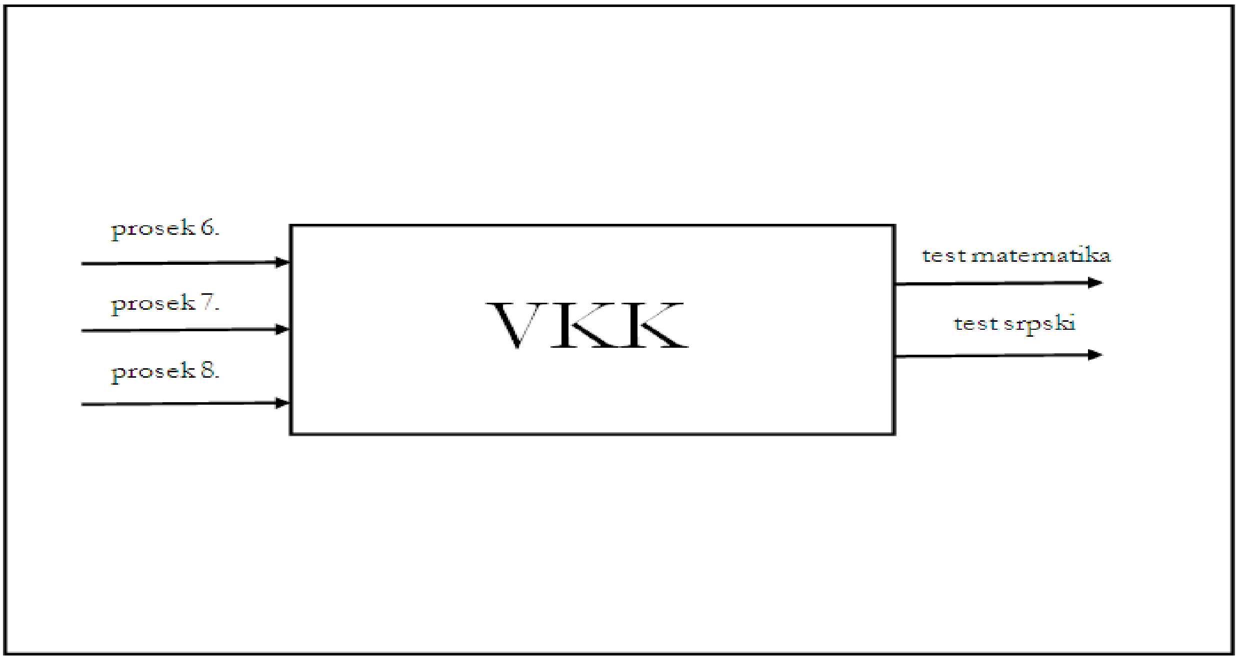
Ova statistika ima $F-raspodelu$ sa $(n_1 m_1 - 1)\ i\ (n_2 m_2 - 1)$ stepena slobode. Zato je verovatnoća

$$P(\{\tau > F_0 / VKK_1 = VKK_2\}$$

Za unapred definisan nivo značajnosti α, vrednost F_0 određujemo iz uslova $P(\tau > F_0) = \alpha$. Odluku o hipotezi H_0 donosimo na uobičajen način:

- ako je $\tau > F_0$, hipotezu H_0 odbacujemo;
- ako je $\tau < F_0$, hipotezu H_0 ne odbacujemo

Vrednost statistike τ izračunavamo iz uzoraka, dok vrednost za F_0 određujemo preko tablica za funkciju $F-raspodele$ sa $(n_1 m_1 - 1)\ i\ (n_2 m_2 - 1)$ stepena slobode.



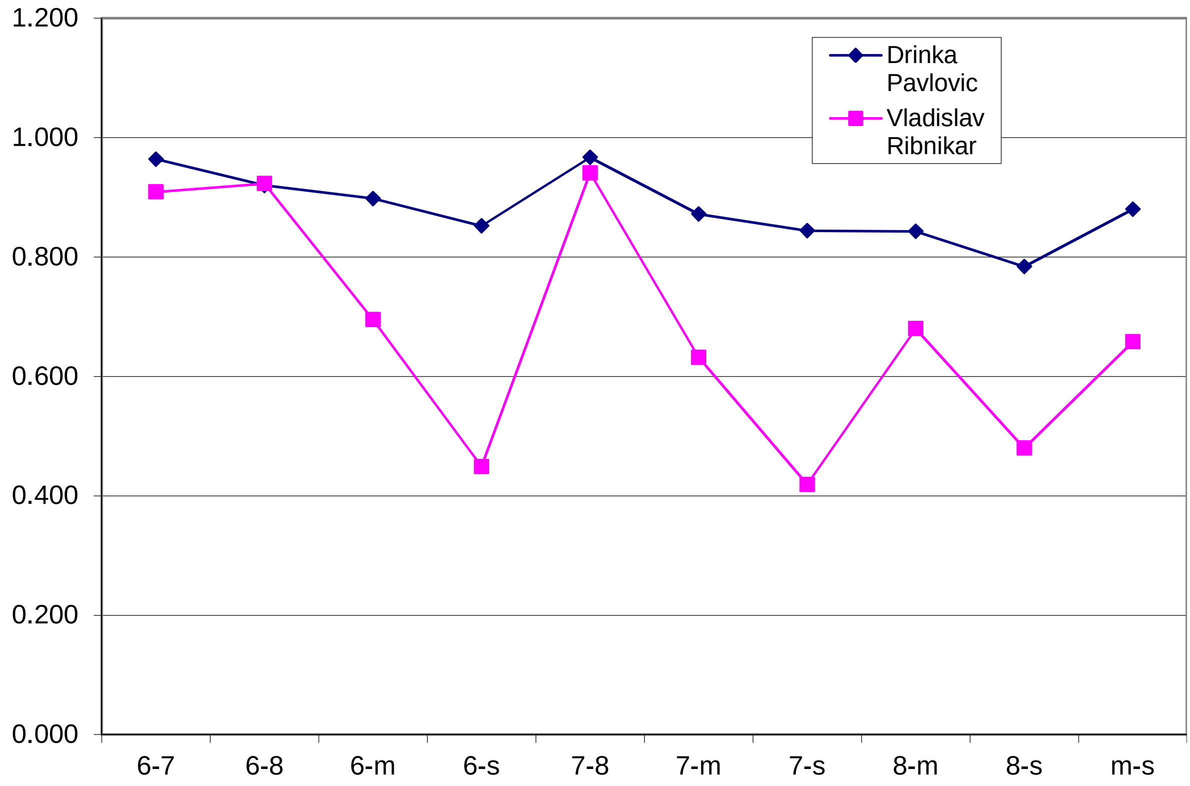
Slika 5.2 Ulazne i izlazne varijable za određivanje VKK

Na slici 5.2 je prikazana veza između ulaznih varijabli (prosečnih ocena u šestom, sedmom i osmom razredu) i izlaznih (broj poena na testu iz matematike i srpskog jezika) izražena kroz vektorski koeficijent korelacije. Na ovaj način su određeni vektorski koeficijenti korelacije za svaku školu i oni su dati u Tabeli 5.1 i biće korišćeni u daljoj analizi.

Tabela 5.1 Rezultati za vrednosti VKK

R.br.	ŠKOLA	Prosek 6	Prosek 7	Prosek 8	Test-mat	Test-srpski	VKK	Broj učenika
1	Drinka Pavlović	4.54	4.43	4.35	16.46	17.30	0.84	42
2	Josif Pančić	4.31	4.12	4.24	15.65	17.32	0.73	68
3	Stefan Nemanja	4.30	4.16	4.24	16.95	16.13	0.67	31
4	Vladislav Ribnikar	4.48	4.43	4.49	16.64	17.27	0.51	58
5	Sveti Sava	4.15	4.01	4.12	13.12	14.60	0.76	77
6	Veselin Masleša	4.19	4.10	4.28	19.23	18.36	0.55	40
7	Bora Stanković	4.15	4.02	4.08	13.35	15.70	0.64	68
8	Momčilo Živojinović	4.07	3.94	4.10	11.12	13.95	0.74	80
9	Gavrilo Princip	3.98	3.96	3.98	14.49	15.39	0.72	65
10	Vuk Karadžić	4.16	4.08	4.10	12.11	14.57	0.66	129
11	Vojvoda Stepa	4.12	3.94	4.00	15.98	15.83	0.53	51
12	Veljko Vlahović	4.20	4.26	4.33	13.01	14.76	0.52	78
13	Kosta Abrašević	3.99	3.91	3.88	10.35	12.49	0.73	71
14	Desanka Maksimović	4.11	3.92	4.02	13.12	14.35	0.56	101
15	Nikola Tesla	4.09	4.02	4.07	11.30	13.69	0.56	146
16	20. oktobar	3.94	3.85	3.86	15.57	16.30	0.51	76
17	Branko Radičević	3.85	3.69	3.75	16.66	16.42	0.48	79
18	Aleksa Šantić	3.76	3.61	3.76	11.90	12.96	0.64	92

Na osnovu tabele 5.1 kada smo uporedili vektorske koeficijente korelacije za škole Drinka Pavlović i Vladislav Ribnikar, na osnovu rezultata iz uzorka, dobijena vrednost statistike $\tau = 9.53$ je veća u odnosu na tabličnu vrednost $F_{5,5} = 5.05$ za nivo značajnosti $\alpha = 0.05$, pa pošto vrednost statistike upada u kritičnu oblast testa donosimo zaključak da odbacimo nultu hipotezu.



Slika 5.3 Uporedni prikaz korelacija ulaznih i izlaznih veličina

Dakle, u školi Drinka Pavlović postoji veći stepen slaganja izlaznih rezultata učenika i njihovog uspeha u školi, nego kod učenika škole Vladislav Ribnikar. Nameće se zaključak da su učenici škole Drinka Pavlović u većoj meri sposobni da svoje znanje pokažu na testu. Naravno, ovo je pre svega jako bitno, obzirom da se prosečne ocene učenika u ove dve škole ne razlikuju puno, kao i njihov uspeh na testu, ali vektorski koeficijent korelacije nam otkriva mnogo više stvari kad se uđe u analizu same složene strukture kovarijacione odnosno korelacione matrice. Zato se dobijeni rezultati naročito moraju uzeti u obzir kod rangiranja škola, a sve u cilju pravljenja lične karte jedne škole, gde bi bili određeni razni pokazatelji uspešnosti rada škole po raznim kriterijumima, a sve u cilju podizanja nivoa praktičnog obrazovanja u našoj zemlji.

5.2. Primena metode I-odstojanja i vektorskog koeficijenta korelacije u rangiranju osnovnih škola u Srbiji

Jedna od vrlo interesentnih tema u oblasti obrazovanja je kako i na koji način vrednovati rad i uspeh obrazovnih institucija. Naravno, postavlja se istovremeno i pitanje: Šta je to što predstavlja uspeh neke škole ? Da li je to uspeh njenih učenika ili uspeh njenih nastavnika? Kako i na koji način to utvrditi i izmeriti i uporediti? Odgovor na ova i još mnogo drugih pitanja nije lako dati. Između ostalog rangiranje osnovnih škola može da pomogne roditeljima da odabere pravu školu za svoju decu, ali i da se unapredi rad tih škola. Na taj način bi lako mogle da se upoređuju škole po različitim kriterijumima, a roditelji bi mogli u svakom momentu da provere kako neka škola napreduje u odnosu na druge. Sa druge strane, ovakvo rangiranje bi poboljšalo rad škola. To što je rejting škole javno dostupan privlači pažnju i može da bude motivacija za škole da rade bolje. Škole koje rade dobro će biti pohvaljene, rejting će im biti bolji, dok one koje rade loše će dobiti povratnu informaciju da njihova uspešnost opada. Ova vrsta pažnje obezbeđuje podsticaj za sve one vezane za rad u školi, da se fokusiraju na poboljšanje rada učenika.

Jedan od delova ove doktorske disertacije upravo će se baviti primenom multivarijantnih statističkih metoda kako bi se dala jedna bolja slika i pregled stanja u osnovnim školama u Srbiji.

Naime, u našoj zemlji ne postoje zvanični kriterijumi na osnovu kojih bi se izvršilo rangiranje osnovnih škola i na osnovu toga preduzele određene akcije u cilju poboljšanja uslova i načina rada u ovim obrazovnim institucijama. Rad sa decom, a pogotovo u periodu dok su u osnovnoj školi je od vitalnog značaja za njihov dalji rad, ne samo sto se tiče nastavka školovanja nego i za život uopšte. Zato obrazovanju na ovom nivou moramo posvetiti punu pažnju, ali to mora uraditi kako pojedinac, tako i država i to sa jasno i precizno definisanim programom. U periodu od 2004. do 2011. godine praćen je uspeh učenika osnovnih škola u Srbiji. Podaci koji su predmet analize odnose se na prosečnu ocenu učenika u šestom, sedmom i osmom razredu osnovne škole, kao i na uspeh koji su ti učenici postigli na prijemnom ispitu za srednju školu, i to na testu iz srpskog jezika i testu iz matematike. Analize su sprovedene i bazirane na najboljih 100 osnovnih škola u Beogradu, na kojima je primenjena metoda I-odstojanja, a sve u cilju kako bi se odredio rang škola u Beogradu.

Tabela 5.2 Rezultati kvadratnog I-odstojanja, rangovi i razlika rangova

R.br.	ŠKOLA	Prijemni	Ukupno	Rang	I2	I2 rang	Razlika	Broj učenika
1	Drinka Pavlović	32.819	87.7846	1	96.84	1	0	99
2	Kralj Petar I	31.93	86.7768	2	95.42	2	0	93
3	Kralj Aleksandar I Karađorđević	33.593	86.7642	3	87.81	3	0	113
4	Janko Veselinović	32.535	86.0038	5	84.44	4	1	126
5	Starina Novak	31.382	85.0696	7	82.02	5	2	97
6	Vladislav Ribnikar	33.144	86.3724	4	80.68	6	-2	129
7	20.oktobar	34.009	85.9638	6	78.74	7	-1	155
8	Laza Kostić	31.01	83.4784	13	71.79	8	5	102
9	Lazar Savatić	33	84.762	8	71.66	9	-1	108
10	Borislav Pekić	31.407	83.5966	11	71.18	10	1	197
11	Jelena Cvetković	32.03	83.58	12	69.13	11	1	119
12	Svetozar Marković	32.428	84.2332	10	68.27	12	-2	83
13	Jovan Miodragović	33.53	84.4244	9	67.79	13	-4	133
14	Jovan Sterija Popović	31.327	82.9382	18	67.23	14	4	142
15	Ljuba Nenadović	29.701	81.787	30	67.07	15	15	152
16	Ćirilo i Metodije	31.834	83.2512	15	66.62	16	-1	121
17	Nadežda Petrović	31.348	82.1664	27	66.37	17	10	92
18	Miloš Crnjanski	31.389	83.2918	14	66.27	19	-5	75
19	Mihailo Petrović Alas	31.88	83.1308	16	66.27	18	-2	95
20	Majka Jugovića	32.019	83.1286	17	64.78	20	-3	79
21	Ujedinjene nacije	31.174	82.304	25	64.01	21	4	178
22	Svetozar Miletić	30.4	82.0216	28	63.02	22	6	80
23	Josif Pančić	31.713	82.841	19	62.70	23	-4	179
24	Gornja Varoš	30.142	81.6732	33	62.40	24	9	77
25	Stefan Nemanja	29.46	81.1392	35	61.90	25	10	50
26	Ivan Gundulić	31.666	82.3684	24	61.51	26	-2	54
27	Banović Strahinja	31.112	82.5396	21	61.37	27	-6	80

R.br.	ŠKOLA	Prijemni	Ukupno	Rang	I2	I2 rang	Razlika	Broj učenika
28	Veselin Masleša	30.956	81.9636	29	60.77	28	1	101
29	Braća Baruh	29.958	81.0896	36	60.32	29	7	59
30	Rade Končar	32.248	82.3986	23	59.92	30	-7	95
31	Kneginja Milica	31.171	81.7734	31	59.57	31	0	111
32	Jovan Dučić	30.705	81.6282	34	59.52	32	2	107
33	Vlada Aksentijević	31.969	82.4298	22	59.39	33	-11	65
34	Ivo Andrić	30.362	81.6768	32	59.29	34	-2	152
35	Radoje Domanović	32.868	82.7996	20	58.93	35	-15	91
36	Stevan Sinđelić	29.127	80.4138	41	58.62	36	5	91
37	Mladost	31.654	82.1992	26	58.42	37	-11	110
38	Duško Radović	29.602	80.3424	42	58.40	38	4	186
39	Skadarlija	29.688	80.7621	38	57.21	39	-1	72
40	Zmaj Jova Jovanović	30.221	80.0886	43	54.64	40	3	115
41	1300 kaplara	28.997	79.5476	45	54.38	41	4	51
42	Đorđe Katić	31.337	80.7874	37	53.67	42	-5	144
43	Branko Ćopić	30.572	80.458	40	53.27	43	-3	160
44	Ivan Milutinović	28.202	78.9861	51	52.81	44	7	94
45	Dr Arčibald Rajs	28.076	78.0592	61	52.66	45	16	66
46	Petar Kočić	28.309	78.6822	55	52.57	46	9	68
47	Stevan Dukić	30.917	79.8635	44	52.27	47	-3	85
48	Bora Stanković	28.379	78.4078	58	51.41	48	10	169
49	Miroslav Antić	28.897	78.6944	54	51.25	49	5	136
50	France Prešern	29.005	79.2598	50	51.14	50	0	99
51	Filip Filipović	29.765	79.519	46	50.78	51	-5	81
52	NH Siniša Nikolajević	27.043	77.4258	68	49.96	52	16	58
53	Marija Bursać	31.087	80.5932	39	49.83	53	-14	81
54	Karađorđe	30.374	79.4952	47	49.61	54	-7	104
55	Sveti Sava	27.597	77.5714	67	48.68	55	12	260
56	Jovan Popović	29.792	79.3368	49	47.95	56	-7	96
57	Dositej Obradović	29.344	78.7396	53	47.69	57	-4	125
58	Ivan Goran Kovačić	30.327	78.7698	52	46.23	58	-6	75
59	Pavle Savić	30.987	79.4706	48	45.56	59	-11	195
60	Vojvoda Stepa	28.785	78.0262	62	45.54	60	2	128
61	Milan Rakić	28.628	77.8656	63	44.94	61	2	86
62	Vojvoda Radomir Putnik	27.058	76.4256	73	44.89	62	11	104

R.br.	ŠKOLA	Prijemni	Ukupno	Rang	I2	I2 rang	Razlika	Broj učenika
63	Oslobodioci Beograda	30.465	78.5762	57	44.31	63	-6	71
64	Stevan Sremac	27.437	76.4518	72	44.01	64	8	223
65	Filip Kljajić Fića	29.339	77.7712	64	43.58	65	-1	152
66	Ilija Birčanin	25.453	73.6174	94	43.05	66	28	172
67	Veljko Dugošević	30.012	78.5968	56	42.92	67	-11	98
68	Olga Petrov Radišić	24.542	73.512	96	42.74	68	28	131
69	Braća Jerković	28.626	77.5946	66	42.18	69	-3	115
70	Boško Palkovljević Pinki	24.556	73.5284	95	42.13	70	25	107
71	14. oktobar	26.281	75.0574	83	42.01	71	12	222
72	Jajinci	29.388	77.7436	65	41.84	72	-7	71
73	Vladimir Rolović	29.882	77.0104	70	41.24	73	-3	102
74	Branislav Nušić	30.536	78.2544	59	41.17	74	-15	110
75	Đura Jakšić	27.312	75.9466	75	41.08	75	0	65
76	Filip Višnjić	30.363	78.2514	60	41.02	76	-16	91
77	Vožd Karađorđe	26.725	75.7358	77	40.49	77	0	60
78	Mihajlo Pupin	27.987	76.8496	71	40.25	78	-7	116
79	Svetislav Golubović	28.239	77.0146	69	39.88	79	-10	109
80	Rade Drainac	25.794	74.2146	89	39.69	80	9	223
81	Vuk Kardžić	27.201	75.6718	79	39.43	81	-2	383
82	Nikola Tesla	26.493	74.5286	86	38.79	82	4	370
83	Sutjeska	26.769	75.4306	82	38.45	83	-1	52
84	Despot Stefan Lazarević	27.309	75.7562	76	39.29	84	-8	170
85	Desanka Maksimović	26.964	75.5524	81	37.92	85	-4	96
86	Vladimir Nazor	26.312	74.2112	88	37.78	86	2	108
87	Branko Radičević	29.063	76.2946	74	36.86	87	-13	361
88	Kosta Abrašević	26.032	74.1036	90	35.46	88	2	203
89	Vasa Čarapić	25.436	72.9746	100	35.37	89	11	94
90	Ilija Garašanin	25.524	73.0612	99	35.16	90	9	128
91	Đura Daničić	28.638	75.6964	78	34.84	91	-13	123
92	Jovan Ristić	26.748	74.0224	91	33.97	92	-1	119
93	Posavski partizani	28.081	74.0146	92	33.18	93	-1	68
94	Vojislav Voka Savić	28.196	,75.0264	84	32.58	94	-10	112
95	Zaga Malivuk	27.247	73.1478	98	32.25	95	3	85
96	Miloje Vasić	30.125	75.5546	80	32.14	96	-16	76
97	Dule Karaklajić	27.429	73.3794	97	31.82	97	0	134

R.br.	ŠKOLA	Prijemni	Ukupno	Rang	I2	I2 rang	Razlika	Broj učenika
98	Sonja Marinković	27.659	73.7826	93	31.47	98	-5	66
99	Gavrilo Princip	27.164	74.2908	87	30.54	99	-12	122
100	Milan Đ. Milićević	31.895	74.7918	85	17.34	100	-15	201

Primenom metode I-odstojanja je dobijen rang škola sa teritorije Beograda i to prvih 100 po ukupnom broju poena koje su učenici tih škola imali na prijemnom ispitu iz matematike i srpskog jezika, odnosno broju poena stečenih na osnovu opšteg uspeha u prethodnom školovanju kao prosečne ocene iz šestog, sedmog i osmog razreda. Uglavnom su škole imali isti rang kao i po ukupnom broju poena, ali obzirom da je I-odstojanje pokazalo visoku korelaciju sa uspehom učenika u šestom, sedmom i osmom razredu, škole koje su imale relativno visoke ocene su se popele na lestvici koje je dobijeno primenom I-odstojanja. Tako na primer, kroz brojne analize škole Drinka Pavlović, Vladislav Ribnikar, Kralj Petar I se jako dobro kotiraju po oba načina rangiranja - u samom su vrhu. Međutim, ono što je za nas interesantno biće ako u analizu uključimo vektorski koeficijent korelacije i njega iskoristimo kao kriterijum za primenu I-odstojanja. Analize su urađene na osnovu podataka za 18 škola iz Tabele 5.1, na osnovu kojih su dobijeni sledeći rezultati.

Tabela 5.3 Korelacije ulaznih I izlaznih varijabli sa I-odstojanjem

Varijable	**r**
1. Prosečna ocena u šestom razredu	0.874**
2. Prosečna ocena u sedmom razredu	0.795**
3. Prosečna ocena u osmom razredu	0.759**
4. VKK	0.615**
5. Broj poena na testu iz srpskog	0.567*
6. Broj poena na testu iz matematike	0.407

Tabela 5.4 Vrednosti I-odstojanja I rangovi

Škola	I-odstojanje	Rang
Drinka Pavlovic	9.886	1
Josif Pancic	7.518	2
Stefan Nemanja	6.451	3
Vladislav Ribnikar	6.286	4
Sveti Sava	5.830	5
Veselin Masleša	5.621	6
Bora Stankovic	5.058	7
Momcilo Živojinovic	4.891	8
Gavrilo Princip	4.886	9
Vuk Karadžic	4.847	10
Vojvoda Stepa	3.907	11
Veljko Vlahovic	3.818	12
Kosta Abraševic	3.755	13
Desanka Maksimovic	3.491	14
Nikola Tesla	3.104	15
20. oktobar	2.925	16
Branko Radicevic	2.198	17
Aleksa Šantic	1.837	18

Na osnovu dobijenih rezultata se vidi da je škola Drinka Pavlović i dalje na prvom mestu, ali je škola Vladislav Ribnikar pala na četvrtu poziciju i to prvenstveno zbog niske vrednosti koeficijenta korelacije. Iz isto razloga su škole Josif Pančić i Stefan Nemanja popravile svoj rejting, samo što je sada visoka vrednost vektorskog koeficijenta korelacije uticala na promenu ranga. Dakle, na ovaj način su one škole čiji su učenici u većoj meri sposobni da usvojeno znanje u školi materijalizuju osvojenim brojem poena na prijemnom, promenile rang ka boljem u odnosu na to kada vektorski koeficijent korelacije nije uzet u razmatranje. Iz gornje tabele se vidi da se sve varijable osim testa iz matematike imale statistički značajnu korelaciju sa I-odstojanjem. Naravno, na ovaj način nije zanemaren uticaj osvojenih broja poena na testu iz matematike, jer je ta informacija sada sadržana u vektorskom koeficijentu korelacije.

Tabela 5.5 Vrednisti za I-odstojanje, VKK, i rangovi

Ime	**Frek**	Sesti	Sedmi	Osmi	Mata	Srpski	VKK	I-distance	Idist*ln(1+VKK)	Rank
Matematicka gimnazija - ogled	**50**	4.940	4.849	4.767	19.120	17.980	0.472	71.81	27.763	1
Kralj Petar I	**93**	4.577	4.590	4.545	15.457	16.473	0.734	36.93	20.327	2
Drinka Pavlovic	**99**	4.628	4.547	4.567	15.960	16.859	0.669	36.88	18.891	3
Kralj Aleksandar I	**113**	4.378	4.381	4.534	16.549	17.044	0.702	29.58	15.731	5
Starina Novak	**97**	4.513	4.454	4.455	15.397	15.985	0.792	28.42	16.578	4
Janko Veselinovic	**126**	4.454	4.427	4.487	15.718	16.817	0.652	29.05	14.583	6
Vladislav Ribnikar	**129**	4.525	4.365	4.417	16.198	16.946	0.651	27.88	13.979	7
Vojvoda Radomir Putnik	**62**	4.421	4.399	4.464	16.137	16.129	0.651	27.62	13.848	8
20 oktobar	**155**	4.290	4.310	4.390	16.777	17.232	0.631	26.76	13.091	9
Lazar Savatic	**108**	4.341	4.259	4.341	16.269	16.731	0.672	22.62	11.627	15
Ivo Andric	**152**	4.360	4.367	4.402	15.010	15.352	0.767	22.05	12.553	11
Milos Crnjanski	**95**	4.408	4.323	4.245	15.668	15.721	0.692	21.88	11.507	16
Laza Kostic	**102**	4.384	4.336	4.397	14.892	16.118	0.727	21.60	11.802	12
Jelena Cetkovic	**119**	4.273	4.269	4.345	15.723	16.307	0.859	20.67	12.816	10
Borislav Pekic	**197**	4.359	4.306	4.382	15.470	15.937	0.650	21.33	10.682	19
Svetozar Markovic	**83**	4.406	4.227	4.318	16.006	16.422	0.769	20.55	11.722	14
Jovan Miodragovic	**133**	4.252	4.175	4.296	16.395	17.135	0.774	20.51	11.757	13
Majka Jugovica	**79**	4.276	4.243	4.258	15.987	16.032	0.701	19.78	10.507	21
Cirilo i Metodije	**121**	4.288	4.249	4.317	15.674	16.260	0.577	19.74	8.992	34
Nadezda Petrovic	**92**	4.313	4.285	4.356	14.810	15.538	0.818	18.34	10.962	17
Vlada Aksentijevic	**65**	4.228	4.213	4.174	15.738	16.231	0.798	18.05	10.589	20
Jovan Sterija Popovic	**142**	4.287	4.261	4.355	15.109	16.218	0.569	18.92	8.522	39
Josif Pancic	**179**	4.326	4.201	4.255	15.872	15.841	0.746	18.05	10.060	23
Banovic Strahinja	**80**	4.384	4.230	4.243	15.225	15.887	0.803	17.66	10.410	22
Ujedinjene nacije	**178**	4.229	4.250	4.303	15.126	16.048	0.713	17.97	9.672	27
Mladost	**110**	4.242	4.235	4.159	15.359	16.295	0.681	17.95	9.323	30
Ivan Gundulic	**54**	4.224	4.197	4.254	15.583	16.083	0.863	17.19	10.695	18
Svetozar Miletic	**80**	4.371	4.246	4.289	15.306	15.094	0.723	17.53	9.537	29
Radoje Domanovic	**91**	4.173	4.160	4.150	16.143	16.725	0.604	18.06	8.533	37
Mihailo Petrovic Alas	**75**	4.259	4.192	4.361	15.287	16.593	0.697	17.57	9.292	31
Ljuba Nenadovic	**152**	4.349	4.268	4.405	14.638	15.063	0.765	17.20	9.772	25
Stefan Nemanja	**50**	4.347	4.288	4.284	14.540	14.920	0.760	17.11	9.673	26
Veselin Maslesa	**101**	4.274	4.239	4.239	15.248	15.708	0.634	17.36	8.524	38

Jovan Ducic	**107**	4.267	4.232	4.231	15.051	15.654	0.779	16.58	9.551	28
Milan D Milicevic	**201**	4.293	4.171	4.260	15.221	16.674	0.699	16.68	8.841	36
Stevan Sindelic	**91**	4.307	4.273	4.241	14.341	14.786	0.860	15.80	9.805	24
Gornja Varos	**77**	4.331	4.223	4.329	14.584	15.558	0.790	15.78	9.187	32
Kneginja Milica	**111**	4.218	4.183	4.249	15.284	15.887	0.762	15.78	8.939	35
Rade Koncar	**95**	4.162	4.139	4.237	15.632	16.616	0.607	16.21	7.690	42
Braca Baruh	**59**	4.277	4.209	4.297	14.822	15.136	0.826	15.02	9.044	33
Skadarlija	**72**	4.327	4.218	4.223	14.771	14.917	0.685	14.91	7.780	41
Dusko Radovic	**186**	4.191	4.218	4.276	14.441	15.161	0.757	14.27	8.043	40
Dorde Krstic	**144**	4.114	4.072	4.177	15.476	15.861	0.677	12.84	6.638	45
1300 kaplara	**51**	4.225	4.193	4.222	14.235	14.755	0.722	12.65	6.875	43
Branko Copic	**160**	4.183	4.115	4.173	15.225	15.347	0.652	12.81	6.430	46
Marija Bursac	**81**	4.174	4.126	4.078	14.759	16.321	0.497	13.14	5.302	56
Filip Filipovic	**81**	4.137	4.153	4.148	14.259	15.506	0.769	11.92	6.799	44
Zmaj Jova Jovanovic	**115**	4.094	4.127	4.246	14.043	16.178	0.687	11.91	6.228	50
Ivan Milutinovic	**94**	4.297	4.182	4.217	13.697	14.505	0.703	11.46	6.101	52
Jovan Popovic	**96**	4.166	4.126	4.095	14.198	15.594	0.706	11.15	5.956	54
Stevan Dukic	**85**	4.002	4.036	4.198	15.076	15.841	0.718	11.01	5.958	53
dr Arcibald Rajs	**66**	4.072	4.189	4.235	13.394	14.682	0.759	10.89	6.150	51
France Presern	**99**	4.230	4.140	4.194	13.510	15.495	0.795	10.76	6.295	48
Karadorde	**104**	4.085	4.051	4.144	14.995	15.375	0.807	10.57	6.254	49
Dositej Obradovic	**125**	4.120	4.108	4.121	14.128	15.216	0.867	10.09	6.300	47
Pavle Savic	**195**	4.093	3.984	4.044	15.233	15.754	0.637	10.08	4.968	58
Petar Kocic	**68**	4.222	4.121	4.250	13.985	14.324	0.677	9.93	5.134	57
Ivan Goran Kovacic	**75**	3.984	4.022	4.104	14.120	16.207	0.788	9.16	5.323	55
Oslobodioci Beograda	**71**	4.032	3.953	4.042	15.338	15.127	0.731	8.98	4.927	60
Veljko Dugosevic	**98**	4.126	4.008	4.018	14.694	15.316	0.748	8.88	4.959	59
Sveti Sava	**260**	4.169	4.143	4.181	13.110	14.487	0.640	9.11	4.507	66
Bora Stankovic	**169**	4.154	4.101	4.251	13.172	15.207	0.680	8.98	4.659	63
Braca Jerkovic	**115**	4.167	4.031	4.044	14.100	15.526	0.753	8.63	4.844	61
Miroslav Antic	**136**	4.133	4.067	4.250	13.191	15.699	0.702	8.73	4.643	64
Filip Visnjic	**91**	4.060	3.943	3.968	15.198	15.165	0.695	8.50	4.485	67
Milan Rakic	**86**	4.133	4.073	4.104	13.657	14.971	0.778	8.27	4.759	62
Vojvoda Stepa	**128**	4.124	4.070	4.115	13.621	15.164	0.585	8.35	3.846	73
NH Sinisa Nikolajevic	**58**	4.254	4.097	4.245	12.612	14.431	0.794	7.81	4.565	65
Jajinci	**71**	4.128	3.947	4.014	15.134	14.254	0.655	7.76	3.909	71
Branislav Nusic	**110**	4.054	3.874	4.002	14.818	15.718	0.727	7.53	4.114	69
Dura Jaksic	**65**	4.058	4.074	4.029	13.938	13.362	0.803	7.20	4.244	68
Svetislav Golubovic Mitraljeta	**109**	4.139	4.050	4.004	12.959	15.280	0.745	7.08	3.942	70

Filip Kljajic Fica	**152**	4.048	3.962	4.098	14.135	15.204	0.751	6.95	3.893	72
Vladimir Rolovic	**102**	3.832	3.921	4.029	14.632	15.250	0.673	7.01	3.607	75
Stevan Sremac	**223**	4.054	4.075	4.124	12.785	14.652	0.737	6.86	3.788	74
Desanka Maksimovic	**96**	4.106	4.058	3.984	13.042	13.922	0.610	6.02	2.867	79
Mihajlo Pupin	**116**	4.194	3.967	4.055	13.297	14.690	0.764	5.53	3.139	77
Branko Radicevic	**361**	3.934	3.918	3.955	14.055	15.008	0.660	5.64	2.858	80
Bosko Palkovljevic Pinki	**107**	4.036	4.097	4.109	11.542	13.014	0.787	5.44	3.158	76
Vozd Karadorde	**60**	4.159	4.021	4.072	12.833	13.892	0.784	5.27	3.051	78
Despot Stefan Lazarevic	**170**	4.101	3.994	4.017	13.241	14.068	0.737	5.12	2.827	81
14 oktobar	**222**	4.038	4.042	4.118	12.385	13.896	0.693	5.16	2.717	82
Vuk Karadzic	**383**	4.073	3.993	4.052	13.162	14.039	0.640	4.99	2.469	84
Dura Danicic	**123**	3.965	3.872	3.928	13.894	14.744	0.729	4.60	2.519	83
Olga Petrov	**131**	4.047	4.042	4.153	11.947	12.595	0.706	4.36	2.329	86
Sutjeska	**52**	4.149	3.973	4.043	12.788	13.981	0.643	4.36	2.165	89
Vojislav Voka Savic	**112**	3.928	3.899	3.880	13.598	14.598	0.714	4.27	2.301	87
Kosta Abrasevic	**203**	4.041	4.003	3.973	12.502	13.530	0.793	4.03	2.353	85
Vladimir Nazor	**108**	3.951	3.990	4.037	12.819	13.481	0.727	4.03	2.202	88
Gavrilo Princip	**122**	4.005	3.938	3.840	13.168	13.996	0.744	3.80	2.113	90
Nikola Tesla	**370**	3.979	3.955	4.075	12.651	13.842	0.734	3.62	1.993	91
Rade Drainac	**223**	4.050	3.944	4.110	12.305	13.489	0.797	3.12	1.829	92
Sonja Marinkovic	**66**	3.882	3.774	3.875	13.909	13.750	0.775	3.07	1.762	93
Vasa Carapic	**94**	3.942	3.949	3.993	12.957	12.479	0.764	3.06	1.737	94
Jovan Ristic	**119**	3.947	3.913	3.958	12.874	13.874	0.722	3.07	1.669	95
Jovan Jovanovic Zmaj	**201**	3.956	3.976	3.980	10.925	11.764	0.821	2.33	1.397	96
Ilija Garasanin	**128**	3.961	3.917	4.006	11.434	14.090	0.524	2.42	1.020	98
Momcilo Zivojinovic	**171**	3.896	3.905	3.955	12.594	12.944	0.732	2.19	1.203	97
Vasa Pelagic	**193**	3.937	3.825	3.825	12.176	13.959	0.722	1.61	0.875	99
Prva obrenovacka osnovna skola	**103**	3.894	3.712	3.834	12.374	13.534	0.797	1.06	0.621	100

U prethodnoj tabeli su prikazani rezultati rangiranja najboljih 100 osnovnih škola u Beogradu, gde su posle primene metode I-odstojanja dobijeni skorovi za svaku škole po formuli $skor = Idist * \ln(1+VKK)$. Na ovaj način je uzet u obzir uticaj vektorskog koeficijenta korelacije koji je izračunat za svaku školu. Najbolje škole koje su dobijene ovim načinom rangiranja su i škole koje su na dobrom glasu u javnosti. Takođe, ovi rezultati se u većoj meri slažu i sa istraživanjem Zavoda za vrednovanje kvaliteta obrazovanja i vaspitanja, koji je za potrebe Ministarstva prosvete izvršio evaluaciju rada učenika i osnovnih škola u Srbiji za 2011. godinu. Rangovi dobijeni na ovaj način su u korelaciji sa DEA metodom r=0.655 i ova korelacija je statistički visoko značajna.

Tabela 5.6 Vrednosti za I-odstojanje, VKK i rang skorovi

Name	Sesti	Sedmi	Osmi	Mata	Srpski	Grad	Poeni	I-distance	VKK	rang-skor
Matematicka gimnazija - ogled	4.940	4.849	4.767	19.120	17.980	1	95.324	48.25	0.472	19.716
Sveti Sava	4.672	4.531	4.737	17.034	17.406	3	90.200	34.06	0.784	18.654
Cele kula	4.532	4.482	4.585	16.245	19.264	3	89.905	30.27	0.64	14.974
Car Konstantin	4.568	4.594	4.722	16.073	17.528	3	89.137	30.6	0.647	15.268
Dorde Natosevic	4.548	4.427	4.489	17.658	17.453	2	88.967	26.53	0.619	12.782
Ucitelj Tasa	4.531	4.535	4.703	15.826	17.250	3	88.152	28.11	0.554	12.392
Drinka Pavlovic	4.628	4.547	4.567	15.960	16.859	1	87.787	22.64	0.669	11.597
Vasa Pelagic	4.470	4.414	4.546	16.271	17.734	5	87.725	24.07	0.496	9.695
Jovan Popovic	4.534	4.483	4.472	16.454	17.206	2	87.616	21.36	0.846	13.094
Kralj Petar I	4.577	4.590	4.545	15.457	16.473	1	86.778	19.8	0.734	10.899
Kralj Aleksandar I	4.378	4.381	4.534	16.549	17.044	1	86.765	22.5	0.702	11.966
Dusan Radovic	4.487	4.451	4.582	15.655	16.979	3	86.714	21.93	0.718	11.868
Petefi Sandor	4.475	4.353	4.391	16.802	16.802	2	86.480	18.87	0.58	8.632
Vladislav Ribnikar	4.525	4.365	4.417	16.198	16.946	1	86.372	18.14	0.651	9.095
Janko Veselinovic	4.454	4.427	4.487	15.718	16.817	1	86.007	18.48	0.652	9.277
20 oktobar	4.290	4.310	4.390	16.777	17.232	1	85.969	19.29	0.631	9.437
Dura Danicic	4.458	4.501	4.569	15.596	15.985	2	85.693	19.34	0.756	10.889
Dositej Obradovic	4.507	4.457	4.485	15.787	15.910	3	85.493	16.97	0.651	8.508
Vojvoda Radomir Putnik	4.421	4.399	4.464	16.137	16.129	1	85.402	17.34	0.651	8.694
Josif Kostic	4.210	4.197	4.346	17.082	17.189	5	85.283	18.91	0.424	6.684
Prva vojvodanska brigada	4.205	4.225	4.318	16.590	17.657	2	85.239	17.92	0.535	7.679
Radoje Domanovic	4.233	4.358	4.400	16.052	17.165	4	85.181	17.22	0.747	9.607
Kosta Trifkovic	4.590	4.324	4.323	15.424	16.714	2	85.086	13.59	0.712	7.307
Starina Novak	4.513	4.454	4.455	15.397	15.985	1	85.070	15.31	0.792	8.931
Ratko Vukicevic	4.560	4.501	4.483	14.781	15.878	3	84.835	14.94	0.814	8.897
Lazar Savatic	4.341	4.259	4.341	16.269	16.731	1	84.764	15.45	0.672	7.942
Vozd Karadorde	4.423	4.438	4.425	14.868	16.434	3	84.446	14.02	0.838	8.534
Sonja Marinkovic	4.413	4.314	4.333	16.049	16.061	2	84.350	13.55	0.532	5.780
Cegar	4.203	4.239	4.391	15.444	17.214	3	83.990	15.3	0.678	7.919

Svetozar Markovic Toza	4.463	4.411	4.428	14.623	15.596	2	83.427	12.1	0.682	6.292
21 oktobar	4.298	4.222	4.258	15.319	16.686	4	83.117	11.25	0.643	5.586
Vuk Karadzic	4.315	4.231	4.403	15.318	15.973	5	83.087	12.79	0.493	5.126
Zarko Zrenjanin	4.327	4.286	4.392	15.286	15.742	2	83.048	12.16	0.741	6.742
Trajko Stamenkovic	4.293	4.282	4.438	14.923	16.048	5	83.023	13.08	0.559	5.808
Radoje Domanovic	4.466	4.458	4.506	13.473	15.204	3	82.397	12.36	0.725	6.739
Bora Stankovic	4.152	4.167	4.214	17.115	15.009	5	82.256	12.49	0.517	5.205
Svetozar Markovic	4.304	4.241	4.379	14.583	15.583	5	81.862	10.1	0.818	6.037
Sveti Sava	4.077	4.147	4.184	15.250	16.966	4	81.848	10.38	0.707	5.551
Ivo Lola Ribar	4.330	4.306	4.381	14.820	14.676	2	81.564	9.4	0.793	5.489
Stanislav Sremcevic	4.204	4.115	4.199	15.273	15.899	4	81.244	8.47	0.776	4.865
Vozd Karadorde	4.104	4.098	4.157	14.667	15.850	5	79.953	6.62	0.71	3.552
Treci kragujevacki bataljon	4.076	4.000	4.098	14.516	16.180	4	79.392	6.42	0.742	3.563
Desanka Maksimovic	4.195	4.225	4.282	13.945	14.570	5	79.323	5.56	0.797	3.259
Kosta Stamenkovic	4.163	4.150	4.252	12.958	15.892	5	79.110	6.09	0.649	3.046
Svetozar Markovic	4.158	4.149	4.240	14.051	14.781	4	79.020	5.14	0.729	2.814
Moma Stanojlovic	4.003	3.927	4.040	15.100	15.969	4	78.949	6.8	0.806	4.020
Mirko Jovanovic	3.995	4.030	4.156	12.965	14.819	4	76.508	3	0.799	1.762
Milutin i Draginja Todorovic	4.089	4.040	4.061	12.811	14.055	4	75.626	1.4	0.703	0.745
Radoje Domanovic	4.064	3.933	3.992	13.698	13.750	5	75.404	1.57	0.692	0.826
Jovan Popovic	4.094	4.042	4.070	12.272	11.846	4	72.942	0.25	0.807	0.148

Da bismo uporedili rezultate škola u različitim gradovima u Srbiji, za potrebe istraživanja smo odredili po 10 najboljih škola u Beogradu(1), Novom Sadu(2), Nišu(3), Kragujevcu(4) i Leskovcu(5). Posle toga je primenjena metoda I-odstojanja, određen je vektorski koeficijent korelacije za svaku školu, pa je na osnovu ovih vrednosti određen skor uspešnosti za svaku školu, odnosno njen rang. Pokazuje se da je Matematička gimanazija- ogledna odeljenja, najuspešnija škola i kada se poredi sa školama u drugim gradovima. Zanimljivo je istaći, da su tri škole iz Niša, Sveti Sava, Ćele Kula i Car Konstantin visoko kotirane, odmah iza Matematičke gimnazije, baš kao i škola Đorđe Natošević iz Novog Sada. Rezultati pokazuju da nije tačno uvreženo mišljenje koje vlada u javnosti da su škole iz Beograda najbolje.

5.3. Izgradnja integralne lične karte osnovnih škola u Srbiji

Na internetu postoji monogo sajtova, na kojima je moguće pogledati rang određene škole, ali i onih na kojima je moguće rangirati škole po različitim kriterijumima na području SAD, Kanade, Australije i Evrope. Osnovna stvar koju treba uzeti u obzir prilikom rangiranja je ispunjavanje „ Izveštaja o radu škole". Ovaj izveštaj predstavlja primarni dokument koji sadrži pregršt bitnih, objektivnih pokazivača rada jedne škole, u jednoj celini, koji će biti lako dostupan javnosti, tako da svako može da analizira i upoređuje rad škola. Na ovaj način izveštaj rada škole pomaže roditeljima da odaberu pravu školu za svoje dete, ali i da ohrabri druge škole da poboljšaju svoj rad. Roditelji u svakom momentu mogu da provere kako neka škola napreduje u odnosu na druge, jer su zbog dodtupnosti izveštaja o radu škola u mogućnosti da lako upoređuju škole po različitim kriterijumima. Sa druge strane, cilj rangiranja škola je poboljšanje njihovog rada. Naime, to što je rejting škola javno dostupan privlači pažnju i može da bude motivacija za škole da rade bolje. Škole koje rade dobro bivaju pohvaljene, dok one koje rade loše, bivaju upozorene da njihova uspešnost opada. Ova vrsta pažnje obezbeđuje podsticaj za sve one, koji su vezani za rad u školi, da se fokusiraju na poboljšanje rada učenika. Upoređivanje sadašnjih rezultata škole sa onima od prošlih godina, možemo videti da li se rad škole poboljšava. Upoređivanjem sa školama u okolini možemo prepoznati one koje su uspešnije i naučiti nešto od njih. Ukupni rezultati jedne škole je postavljaju na njeno mesto u odnosu na sve škole u sistemu.

5.3.1.Kriterijumi rangiranja škola u Velikoj Britaniji

Većina škola u svetu, a naravno i u Velikoj Britaniji se razlikuje po tipovima. Neke od njih su: državne škole, religijski zasnovane škole(zbog finansiranja koje potiče iz religijskih struktura), akademije (škola koje je direktno povezana i finansirana od strane ministrastva), nezavisne škole (imaju svoju nezavisnost od uticaja drugih, osim od ministarstva) i zadužbine (škola koja je zadužbina neke značajne ličnosti). Kriterijumi koji se primenjuju u Velikoj Britaniji na osnovu istraživanja Nacionalnog zavoda za obrazovanje:

- Procenat đaka koji ostvare nivo4, ili više na engleskom i matematici
- Procenat đaka koji ostvare nivo5 na engleskom i matematici
- Procenat đaka koji ostvare očekivani napredak na engleskom
- Procenat đaka uključen u merenju ostvarivanja napretka na engleskom
- Procenat đaka koji ostvari očekivani napredak u matematici
- Procenat đaka uključen u merenju ostvarivanja napretka iz matematike
- Prosečna ocena đaka ostvarena na testovima
- Ukupan broj đaka koji je uključen u istraživanje
- Procenat ukupnog broja đaka koji je ostvario nivo4+ iz oba predmeta
- Procenat od ukupnog broja đaka koji imaju neki hendikep
- Procenat hendikepiranih đaka koji ostavre nivo4+ na oba predmeta
- Procenat đaka određenog profesora koji je ostvario nivo3 ili niži nivo znanja
- Procenat đaka određenog profesora koji je ostvario nivo4 ili viši nivo znanja
- Procenat đaka određenog profesora koji je ostvario nivo5 ili viši nivo znanja
- Procenat đaka određenog profesora koji su neopravdano odsutni sa časa
- Procenat đaka koji su testirani i pripadaju tekućoj generaciji
- Procenat đaka koji su podobni za ovu fazu testiranja
- Procenat đaka kojima engleski nije maternji jezik
- Procenat časova na kojima su đaci odsutni
- Procenat časova na kojima su đaci neopravdano odsutni
- Procenat đaka koji više od 15% časova nisu pohađali konstatno
- Procenat đaka koji više od 20% časova nisu pohađali konstatno
- Broj đaka koji pohađa školu
- Procenat đaka koji ima pravo na besplatan obrok u školi
- Ukupan prihod škole po đaku

- Ukupan trošak škole po đaku
- Broj učitelja u školi
- Broj asistenata učiteljima u školi
- Broj osoblja koje čiči podršku nastavi
- Broj učitelja koji su na stalnom zaposlenju
- Broj asistenata učitelja koji su na stalnom zaposlenju
- Odnos đaka i učitelja
- Prosečna plata učitelja
- Broj đaka koji su stalni
- Broj đaka koji su na spisku za testiranje
- Prosečan broj godina koje ima učitelj
- Procenat učitelja koji imaju preko 50 godina
- Procenat učitelja sa manje od 3 godine radnog iskustva kao učitelj
- Procenat učitelja koji ima platu veću od nacionalnog proseka plate učitelja
- Da li je škola prošla inspekciju (obrazovna, komunalna i sl.)

Osim ovako definisanih kriterijuma i urađenog rangiranja na osnovu njih, u Velikoj Britaniji je u žiži javnosti uvek rangiranje po The Telegraph magazinu. Naime, tabele rangiranja po The Telegraph magazinu se zasnivaju na učinku 11-godišnjaka iz matematike i maternjeg jezika. Magazin rangira škole po procentu učenika koji su stekli nivo4 – standard koji se očekuje za njihov uzrast, na testovima iz engleskog i matematike. Prema Vladi, najmanje 60% učenika trebalo bi da dostigne ovaj cilj u većini škola. Prosečan skor po poenima je mera kojoj megazin pridaje veliki značaj. Naime, testovima učenika se daju određeni bodovi za nivo kiji postižu. Nivo2 ili ispod se vrednuje 15 poena, nivo3 vredi 21 poen, nivo vredi 27 poena, nivo5 vredi 33 poena i nivo6 vredi 39 poena. Bodovi svakog učenika za engleski i matematiku se sabiraju i dele sa brojem testova koji su rađeni, da bi se dobio prosečni skor škole na osnovu rezuktata učenika. Osim ove mere, posebna pažnja je usmerena ka meri dodatne vrednosti kojom se procenjuje iznos napretka učenika koji se može uočiti između 7 i 11 godina. Takođe se uzima u obzir i niz drugih faktora, kao što su učenici koji govore engleski jezik kao drugi jezik, učenici sa posebnim potrebama, a takođe i deca koja imaju pravo na besplatne obroke u školi. Na ovaj način se može videti i „mera napretka učenika“ za datu školu – drugi ključni Vladin pokazatelj. Od učenika starosti između 7 i 11 godina se očekuje da načine napredak od „dva nivoa“. Za prosečnog učenika ovo znači da postigne Nivo2 na proceni koja se radi u sedmoj godini i Nivo4 na proceni koju učenik radi u jedanaestoj godini.

5.3.2.Kriterijumi rangiranja škola u Americi

U većini slučajeva se škole rangiraju na osnovu prijavljenih rezultata testova. Sistem rangiranja za većinu država je sledeći: uzimaju se u obzir sve škole koje imaju testove za matematiku i engleski jezik. Sledeći parametri su od posebne važnosti: prosečna ocena iz matematike u svim razredima, prosečna ocena iz engleskog jezika u svim razredima, a potom se na osnovu ovih ocena dobijaju kombinovani rezultati i na kraju im se dodeljuje odgovarajući rang.

Tabela 5.7 Vrednosti za rang i rang procenat

Škola	Rang	Rang Procenat
Lincoln Elementary	15-ta od 100 osnovnih škola	.85
Jefferson Elementary	25-ta od 100 osnovnih škola	.75
Jackson High School	5-a od 50 osnovnih škola	.90
	Srednji rang procenat (Skor ranga za okrug):	.8333

Ovakav proračun se napravi za sve oblasti, a potom se odredi spisak okruga po skor rangu za okrug.

Prestižni Chicago-Sun-Times magazin, dobitnik Pilitzer-ove nagrade za 2011. godinu zasniva svoje ekskluzivno rangiranje državnih škola na osnovu prosečnih rezultata postignutih na državnim testovima uspešnosti. Analizirani su samo rezultati koji su ostvareni u oblasti čitanja i matematike na standardnom testu uspešnosti. Rangiranje osnovnih škola se zasniva na školama koje testiraju najmanje dva razreda. Prilikom rangiranja koristi se standardizacija podataka radi analiziranja „skale rezultata“ svakog državnog testa čitanja i matematike. Ovaj metod poredi svaki rezultat testa sa državnim prosekom i izračunava pravi prosek škole koji se onda poredi sa ostalim školama. Ovakav sistem omogućava veću idefinisanost među najboljim školama, zato što izračunava prosečnu vrednost za svaki rezultat, umesto da broji samo učenika koji postižu ili premašuju rezultat koji je potreban za prolaz (Thurston, 1926). Rangiranje uključuje pokazatelj koji izražava procenat učenika koji su postigli isti ili lošiji rezultat u odnosu na prosečnog studenta u svakoj od rangiranih škola. U središtu grupe proseci škola su mnogo bliže jedni drugima tako da razlike u mestu na rang listi između škola mogu da odražavaju male razlike u prosečnim rezultatima. Takođe, škole u sredini imaju sklonost da imaju više veza na rang listi, tako da veće razlike u rang mestima mogu da odražavaju manje razlike nego što je to slučaj sa školama koje se nalaze na vrhu ili dnu liste. Ovaj indikator pokazuje u kojoj je meri određena škola postigla bolje ili lošije rezultate u odnosu na škole koje su neposredno iznad, odnosno ispod nje.

Tabela 5.8 Vrednosti za rang i relativni skor

Čikago rang	**Škola**	**Procenat**	**Državni rang**
1	Decatur *	89.71	1
2	Keller*	81.94	2
3	Lenart*	80.65	3
4	Edison*	79.05	4
5	Skinner*	73.50	6
6	McDade*	71.97	7
7	Poe*	69.04	9
8	Lincoln	66.82	14
9	Bell	62.82	29
10	Oriole Park	62.67	32
11	A. Jackson	62.06	37
12	Edgebrook	61.29	48

13	Hawthorne	60.91	52
14	LaSalle*	58.63	79
15	Burley	56.67	110
16	Blaine	55.09	134
17	South Loop	53.90	154
18	Wildwood	52.39	193
19	Orozco	51.71	212
20	Norwood Park	51.08	228
21	Franklin	50.80	235
22	Alcott	49.20	281
23	Ogden	47.13	338
24	Stone	46.49	361
25	Ebinger	46.30	372
26	Edison Park	46.10	380
27	Murray Language	45.54	393
28	Thorp	44.83	417
29	Canty	43.72	461
30	Disney	42.78	502
31	Ward J	42.66	506
32	Nettelhorst	42.58	510
33	Mount Greenwood	42.15	530
34	Beaubien	41.76	549
35	Solomon	41.76	549
36	Chicago	41.37	565
37	Coonley	40.98	585
38	Audubon	40.79	593
39	Healy	40.71	596
40	Sheridan	40.25	615
41	Sutherland	39.67	645
42	Owen	39.36	660
43	Pershing West	39.24	667
44	Drummond	38.40	706
45	Newberry	38.13	724
46	Garvy J	38.09	727

47	Courtenay	37.75	745
48	Locke A Elem	37.56	750
49	Sauganash	37.15	771
50	Ariel	37.07	772

Jedno od najobuhvatnijih istraživanja u pogledu rangiranja osnovnih škola je obavila radna grupa sa Fraser instituta, a ona se tiču osnovnih škola u Vašingtonu. Osnova izveštaja o radu škole je sveobuhvatno ocenjivanje akademskog učinka svake škole. Ocena akademskog učinka se zasniva na osnovu sedam pokazatelja. Svi pokazatelji su izvedeni na osnovu rezultata standardnih testova (Washington Assessment of Student Learning – WASL) u oblasti čitanja, pisanja, znanja iz matematike i nauke:

- Prosečan nivo uspeha na WASL proceni čitanja u 3, 4, 5 i 6 razredu
- Prosečan nivo uspeha na WASL proceni pisanja u 4 razredu
- Prosečan nivo uspeha na WASL proceni znanja matematike u 3, 4, 5 i 6 razredu.
- Prosečan nivo uspeha na WASL proceni znanja nauke u 5 razredu
- Procenat neuspešnih WASL procena
- Razlike između učenika čije porodice imaju mali godišnji prihod i učenika čije porodice nemaju mali godišnji prihod u prosečnom nivou uspešnosti na WASL proceni čitanja u petom razredu
- Razlike između učenika čije porodice imaju mali godišnji prihod i učenika čije porodice nemaju mali godišnji prihod u prosečnom nivou uspešnosti na WASL proceni znanja iz matematike u petom razredu

Izabran je ovaj skup pokazatelja zato što pružaju sistematičan uvid u učinak škola. Pošto su pokazatelji zasnovani na podacima koji se prikupljaju na godišnjem nivou, mi možemo proceniti ne samo učinak škole, već njen napredak ili nazadovanje tokom vremena. Najvažniji zadatak osnovnih škola jeste podučavanje dece osnovnim veštinama u oblasti čitanja, pisanja i matematike. Osnovno znanje čitanja, pisanja i računanja predstavlja suštinsku podlogu za celoživotno učenje. Istraživanje je bazirano na rezultatima testova koji procenjuju učenika u okviru ovih dimenzija. Razlike između učenika u pogledu sposobnosti, motivacije i radnih navika neizbežno imaju određeni uticaj na konačne rezultate. Ipak, postoji vidljiva razlika u prosečnim rezultatima na WASL testovima između škola u istom okrugu. Takođe, postoji i razlika u okviru iste škole između rezultata učenika u različitim predmetima i različitim razredima. Takve razlike ne mogu biti objašnjene pozivanjem na individualne i porodične osobenosti učenika. Iz tih razloga čini se opravdano uključiti prosečne ocene na testovima iz ova četiri predmeta kao pokazatelje uspešnog podučavanja. Posebno je zanimljiv pokazatelj stope neuspeha na WASL testovima. On se dobija deljenjem ukupnog broja svih testova koji su pružili dovoljno informacija za izračunavanje rezultata, ali nisu ispunili definisan državni standard, sa ukupnim brojem takvih testova koji su učenici u datoj školi uradili. Pošto su čitanje, pisanje i znanje iz matematike i prirodnih nauka važni za dalji intelektualni i lični razvoj, učenici bi trebalo da pokažu da ispunjavaju standard predviđen za njihov razred u datim predmetima. Sa druge strane, škole imaju obavezu da osiguraju da njeni učenici budu u stanju da to i urade.

Iako je svaki pokazatelj bitan, skoro u svim slučajevima svaka škola u nekim pokazateljima postiže bolje rezultate, a u nekim lošije (Bukvić, 2002). Kao što predavač mora da donese odluku o učenikovom opštem učinku, tako i nama treba opšti pokazatelj učinka škole. Kao što predavači kombinuju rezultate testova, domaće zadatke i aktivnost na času da bi ocenili učenike, tako i mi kombinujemo sve indikatore da bismo došli do opšte ocene. Opšta ocena učinka škole pruža odgovor na pitanje: „Uopšteno, kakav je akademski učinak ove škole u poređenju sa ostalim školama ?"

Da bi se dobila ova ocena, rezulatati su prvo za svaki od sedam pokazatelja pojedinačno za svaku školsku godinu bili standardizovani. Standardizovane vrednosti mogu biti kombinovane i upoređivane. Standardizovanim podacima su potom dodeljeni težinski koeficijenti i kombinovani su kako bi proizveli opšti standardizovani rezultat. Na kraju, ovaj rezultat je pretvoren u rang(od 1 do 10). Na osnovu ovog ranga(od 1 do 10) određeno je mesto škole. Treba primetititi da je rang(od 1 do 10) relativno rangiranje, tj ono meri učinak svake škole pojedinačno u poređenju sa svim ostalim školama u državi (Welsh, 2001). Stoga, čak iako škola postigne opštu ocenu 10, veoma je verovatno da se ona može i popraviti. Opšta ocena 0 znači da je škola imala najlošiji učinak u zemlji. Ipak to ne znači da ta škola nije ništa uradila za svoje učenike. Na slici 5.4 prikazan je izveštaj o radu škole sa parametrima koji uzeti za evaluaciju rada učenika.

A – **GEOGRAPHICAL AREA**

B – **School name [Public] City** — **Grades: K-6** — **Enrollment: 328**

C – Low income: 70.9% — Ethnicity: Wh: 89.6% Hi: 4.0%

D – Tests not written not exempt: 0.0% — 2007-08 Rank: 989/1130 — Last 5 yrs 907/1037

	Academic Performance		**2004**	**2005**	**2006**	**2007**	**2008**	**Trends**	
E –	Avg level:	Reading	2.6	2.7	2.7	2.6	2.6	▲	
F –		Writing	n/a	2.2	2.3	2.5	2.4	n/a	
G –		Math	2.0	2.2	2.2	2.2	2.3	—	
H –		Science	2.0	2.0	2.0	1.9	2.1	▼	– M
I –	Tests below standard (%)		64.8	55.9	55.6	51.4	51.9	—	
J –	Low-income gap:	Reading	n/a	n/a	n/a	n/a	n/a	n/a	
K –		Math	n/a	n/a	n/a	n/a	n/a	n/a	
L –	**Overall rating out of 10**		**4.2**	**4.0**	**3.9**	**3.9**	**3.9**	▼	

Slika 5.4 Izveštaj o radu škole – Report Card (Chicago-Sun-Times magazin,2011)

Isto tako pošto se radi o realtivnom merenju da bi škola pokazala napredak u svom rangu(od 1 do 10) ona mora napredovati brže nego prosek. Ako se popravi, ali za stopu koja je niža od proseka, pokazaće nazadovanje u svojoj oceni. Da li se škola popravlja u akademskom smislu? Uglavnom se prilikom ovakvih istraživanja uzimaju podaci za poslednjih pet godina. Za razliku od podataka za jednu godinu, istorijski izveštaji pružaju dokaze o promeni (ili izostanku promene) tokom određenog vremenskog perioda. Upravo za ovakvu svrhu je određen trend pokazatelj, koji treba da identifikuje one dimenzije učinka škole u kojima je promena najverovatnija, a ne da identifikuje fluktuaciju u rezultatima koja je prouzrokovana slučajnim događajima.

5.3.3.Kriterijumi rangiranja škola u Srbiji

U Srbiji ne postoje zvanični kriterijumi propisani od nadležnih institucija na osnovu kojih bi se moglo izvršiti rangiranje škola. Tokom 2012. godine od strane Zavoda za vrednovanje kvaliteta obrazovanja i vaspitanja je objavljeno istraživanje po kome su objavljena imena 50 najboljih škola u Srbiji. Prilikom evaluacije rada škola kao parametri su uzeti prosečne ocene učenika u šestom, sedmom i osmom razredu, kao i prosečan uspeh na testu iz matematike i srpskog jezika. Objavljivanje ovih rezultata je imalo veoma veliki odjek u javnosti i puno kritika na način kako je rangiranje urađeno. Spisak najboljih osnovnih škola se nije slagao sa mišljenjem roditelja i njihovim viđenjem šta je to „dobra“ škola. U prethodnom poglavlju su prikazani neki od rezultata koji su dobijeni u istraživanju tokom izrade ove disertacije i koji se generalno slažu sa rezultatima dobijenim od strane Zavoda za vrednovanje kvaliteta obrazovanja i vaspitanja. Međutim, očigledno je da postojeći kriterijumi koji su uzeti u obzir za evaluaciju kvaliteta obrazovanja nisu dovoljni da bismo imali kompletnu i što verniju sliku rada i uspeha učenika, odnosno osnovnih škola. Imajući u vidu napore koje ulaže Ministarstvo prosvete u ovoj oblasti, kao i određena rešenja koja već postoje u drugim državama, u okviru disertacije je predložen spisak kriterijuma koje bi trebalo uzeti u obzir prilikom rangiranja škola. Pri tom je veoma važno da država, odnosno resorno ministarstvo, zvanično podrži napor ka definisanju jedinstvenih kriterijuma koje bi onda svi učesnici u ovom lancu morali poštovati.

Očekivano je da se u ocenjivanju škola uključe rezultati koje učenici ostvare tokom školovanja. Ti rezultati se mogu pratiti na različite načine. Na početku je neophodna adekvatna podela kriterijuma, odnosno karakteristika, da bi se omogučilo dodeljivanje određenih nivoa značajnosti pojedinim kriterijumima u svrhu istraživanja, odnosno rangiranja. Pre svega imamo podelu parametara po tome nad kim se oni ocenjuju. Ta podela se svodi na sledeće:

- Karakteristike škole
- Karakteristike osoblja škole
- Karakteristike učenika
- Opšti kriterijumi
- Kriterijumi istraživanja

Karakteristike škole mogu biti određene na više načina. Kao prvo se izdvaja tip škole i osnovna podela po tipu škole je na privatne i državne. Sledeća karakteristika škole se odnosi na lokaciju škole, odnosno regija, grad i opština kojoj škola pripada. Tako je lakše definisati posebne liste škola, što bi roditeljima omogućilo lakši izbor, u slučaju da je lokacija bitna. Neophodno je definisati troškove i prihode koje ima škola. Ukoliko su u pitanju državne škole, unapred se zna budžet škole, a kod privatnih to uglavnom zavisi od cene školarine, kao i od donacija, privatnih sponzora škole i slično. Posebno treba naglasiti koliko se troši po đaku koji pohađa tu školu. Osim toga, može se definisati trošak škole na renoviranje prostorija, objekata, inventara, kao i opreme koju koriste đaci. Stanje škole se može odrediti tako što se pogleda spisak inspekcija koje je škola prošla. Ključne su sanitarna, komunalna i obrazovna inspekcija. U okviru politike upisa treba navesti kriterijume upisa u škole. Radna snaga škole predstavlja veoma bitnu karakteristiku škole. To su učitelji, asistenti učitelja, tetkice, direktor, sekretar, psiholog, psihijatar. Gleda se broj zaposlenih na određenoj poziciji, odnos broja đaka i učitelja, prosečna plata zaposlenog, veličina svakog odeljenja, odnosno broj đaka po odeljenju. Sveukupno odsustvo đaka sa časova, kao i procenat časova na kojima su đaci odsutni je veoma bitna karakteristika, kao i procenat neopravdanog odsustva. Treba uzeti u obzir procenat učenika koji više od 15% časova, odnosno više od 25% časova nisu pohađali nastavu u kontinuitetu.

Karakteristike osoblja škole se odnose na zaposlene. Trebalo bi uzeti u obzir sledeće kriterijume:

- Prosečan broj godina koje ima učitelj

- Procenat učitelja koji imaju preko 50 godina
- Procenat učitelja sa manje od 3 godine radnog iskustva kao učitelj
- Prosečna plata učitelja
- Procenat učitelja koji ima platu veću od državnog proseka plate učitelja
- Procenat đaka koji ostvare očekivan napredak kod učitelja
- Procenat đaka koji ostvare ocene 4 ili više iz srpskog jezika kod učitelja
- Procenat đaka koji ostvare ocene 4 ili više iz matematike kod učitelja
- Procenat đaka tog učitelja koji je ostvario ocenu 3 ili niži nivo znanja
- Procenat đaka tog učitelja koji su opravdano odsutni sa časova
- Procenat đaka tog učitelja koji su neopravdano odsutni sa časova

Karakteristike đaka koje treba uzeti u obzir su:

- Da li je đak trenutne generacije ili ne
- Da li đak može biti testiran radi dobijanja statističkih podataka
- Da li đak pripada grupi koja ima pravo da koristi besplatne obroke u školi
- Veroispovest, nacionalnost
- Socijalni status
- Posebne potrebe

Opšti kriterijumi se mogu primeniti direktno na školu, ali i pojedinačno na đake, odnosno učitelje:

- Broj đaka koji su redovni/vanredni
- Broj đaka koji su na spisku za testiranje
- Broj đaka koji govore srpski kao drugi jezik (veoma značajan kriterijum u oblastima blizu granica, gde dolazi do mešanja structure stanovništva)
- Procenat đaka koji nisu srpskog porekla
- Procenat đaka romske nacionalnosti
- Procenat đaka nacionalnih manjina
- Broj đaka sa posebnim potrebama
- Da li su đaci deca roditelja koji imaju loš socijalni status

- Prosečna ocena iz matematike u datoj godini na osnovu testova
- Prosečna ocean iz srpskog jezika u datoj godini na osnovu testova
- Procenat đaka koji ostvare ocenu 4 ili vise na predmetima srpski i matematika
- Procenat đaka koji ostvare ocenu 3 ili manje na predmetima srpski i matematika
- Procenat đaka koji ostvare očekivani napredak na srpskom jeziku
- Procenat đaka koji ostvare očekivani napredak na matematici

Primena predloženih kriterijuma za vrednovanje kvaliteta obrazovanja uz izgradnju lične karte svih škola u Srbiji, pružilo bi osnovu za javno praćenje rada svih škola. Na ovaj način bi svi učesnici u sistemu obrazovanja bili aktivirani da daju maksimalni doprinos u cilju kvalitetnijeg i efikasnijeg rada škola i njihovih učenika.

6. ZAKLJUČAK

Analiza složene korelacione strukture je tema koja se vrlo intenzivno razvija i privlači veliku pažnju istraživača u mnogim oblastima. Različite metode su korišćene, ali su kanonička korelaciona analiza i modelovanje strukturne jednacine (SEM) najčešće zastupljene. U različitim oblastima npr. psihologiji, medicini, geodeziji su korišćeni još i neki drugi modeli za analizu korelacione strukture kao i testiranje hipoteza. Upravo kroz doktorsku disertaciju je predstavljen potpuno novi model strukturne korelacione analize zasnovan na vektorskim koeficijentima korelacije. Sve navedeno nas vodi do zaključka da materija koja će biti izložena u ovoj disertaciji ima posebnu vrednost i predstavlja vredan naučni doprinos.

Glavna hipoteza koja je razvijana u okviru doktorske disertacije je da je moguće odrediti statistiku za testiranje hipoteze o jednakosti dva vektorska koeficijenta korelacije. Na osnovu toga predložen je model strukturne korelacione analiza u kome će se izvršiti poređenje dve nezavisne korelacione strukture, koji se može primeniti u raznim organizacionim sistemima. Model strukturne korelacione analize zasnovan na vektorskim koeficijentima korelacije se ističe svojom primenljivošću i mogućnošću uključivanja velikog broja varijabli (ulaza i izlaza).

Pored glavne hipoteze, treba naglasiti da je u radu predstavljen jedan način rangiranja zasnovan na vektorskim koeficijentima korelacije jedne korelacione strukture uz primenu metode Ivanovićevog odstojanja. Na taj način, dobijena je jedna verna slika posmatranih objekata koji su radi postizanja preferenci rangirani tj. postavljeni u relacioni odnos. Takođe, treba naglasiti da je u radu prikazan jedan algoritam za rešavanje problema grupisanja sa unapred definisanim ograničenjima, kao modifikacija postojećeg K-mean algoritma.

U doktorskoj disertaciji predstavljena je detaljna analiza problema sa krajnjim ciljem da se novorazvijeni model korelacione strukturne analize integriše sa metodom DBA i DEA i na taj način dobije jednu sasvim drugačiju dimenziju u cilju rešavanja problema merenja efikasnosti i otkrivanja zakonitosti u složenim korelacionim strukturama.

U doktorskoj disertaciji je razmatran problem empirijskog i matematičkog dokaza o slaganju I-odstojanja sa normalnom raspodelom. Imajući u vidu dobijene rezultate o slaganju I-odstojanja sa normalnom raspodelom dalja istraživanja se mogu usmeriti za unapređenje metode I-odstojanja u postupku rangiranja.

Rezultati dosadašnjih i budućih istraživanja na temu strukturne korelacione analiza zasnovane na vektorskim koeficijentima korelacije kao i metodi I-odstojanja su i biće objavljeni u naučnim časopisima međunarodnog značaja, kao i saopšteni na skupovima u zemlji i inostranstvu.

6.1. Doprinosi doktorske disertacije

U okviru doktorske disertacije je dato nekoliko osnovnih doprinosa.

Glavni doprinos se ogleda u definisanju i primeni test statistike za poređenje dva vektorska koeficijenta korelacije. Na ovaj način je moguće utvrditi vezu između izlaznih i ulaznih veličina jednog organizacionog sistema, ali i utvrditi i izmeriti razlike između dva organizaciona sistema, tako da se može uvideti stepen značajnosti sličnosti ili razlike između posmatranih organizacionih sistema. Kandidat je dao teorijski prikaz postojećih test statistika za poređenje dve korelacione strukture, kao i niz eksperimentalnih rezultata koji verifikuju usvojeni koncept. Za izračunavanje vektorskog koeficijenta korelacije, test statistike i kritične oblasti testa napisan je program u Matrix programskom jeziku za SPSS (v.21) programski paket, sa ciljem da bude pristupačan širem krugu korisnika.

Drugi doprinos se odnosi na problem određivanja raspodele I-odstojanja. Određena je raspodela kvadratne forme normalno raspoređenih vektora, pa je njenom primenom na kvadratno I-odstojanje pokazano da ima normalnu raspodelu. Takođe, u opštem slučaju, a za šta je korišćena *Bootstrap* metoda, koja kao svoj sastavni deo podrazumeva primenu Monte-Karlo simulacije, pokazano (osim u izuzetnim slučajevima) je slaganje I-odstojanja sa teoretskom normalnom raspodelom. Time je Ivanovićevo odstojanje dobilo novu dimenziju posmatranja i značajno unapređen kvalitet za njegovu primenu.

Treći doprinos se odnosi na definisanje jednog algoritma za probleme rangiranja. Rangiranje se zasniva na I-odstojanju, ali je uzet u obzir odnos izlaznih i ulaznih veličina izražen kroz vektorski koeficijent korelacije. Ovime je pokazano da se vektorski koeficijent korelacije može koristiti kao težinski faktor u procesu rangiranja i time omogućiti bolje tj.“realnije“ proces rangiranja i same rezultate.

Četvrti doprinos je dat kroz kreiranje jednog a priori načina grupisanja sa unapred definisanim ograničenjima, kao jedna modifikacija McQeen-ovog *K-mean* algoritma nehijerarhijskog grupisanja.

7. LITERATURA

1. Agresti, A. (1996): An Introduction to □ategorical Data Analysis, John Wiley & Sons Inc. New York.

2. Agresti, A. (1984): Analysis of Ordinal Categorical Data, John Wiley & Sons Inc. New York.
3. Agresti, A. (1981) : A Hierarchical System of Interaction Measures for Multidimensional Contingency Tables. J. Roy. Statist. Soc. B 43:293-301.
4. Agresti, A. & Agresti B. (1979): Statistical Methods for the Social Sciences, San Francisco: Dellen.
5. Anderberg, M. R. (1973): Cluster Analysis for Applications, Academic Press, London.
6. Anderson, T. W. (1966): An Introduction to Multivariate Statistical Analysis, 7th ed., John Wiley and Sons, London.
7. Bartlet, M. S. (1941): The statistical significance of canonical correlation. Biometrika,32,29-38.
8. Bartholomew, D. J. (1980): Factor Analysis for Categorical Data (with discussion). J. Roy.Statist. Soc. B 42:293-321.
9. Batagelj, V., Hermann, H.B., Ferligoj, A. & Žiberna, A. (2006): Data science and classification. Springer, Berlin.
10. Bogosavljević, S. & Kovačević, M. (1996): Analiza grupisanja II, Majski skup 96., Savezni zavod za statistiku, Beograd.
11. Bogosavljević, S. (1985): Apriorne metode klasifikacije ekonomskih pojava, Doktorska disertacija, Beograd.
12. Bogosavljević, S. (1988): Evaluacija klasifikacione strukture, Zbornik radova, Majski skup '87, Sekcije za klasifikacije Saveza statističkih društava Jugoslavije, Beograd, SZS.
13. Bogosavljević, S. (1997): O statističkim metodama u rangiranju, Seminar katedre za matematiku i informatiku, FON, Beograd.
14. Bogosavljević, S. (1996): Formalno definisanje i uređenje hijerarhijske klasifikacije, u Bogosavljević, S. & Kovačević, M. (red.): Analiza grupisanja II, Savezni zavod za statistiku, Beograd, 43-48.
15. Breckling, J. (1989): The Analysis of Directional Time Series: Applications to Wind Speed and Direction, Springer-Verlag, 238 pp.

16. Breslow, N. (1982): Covariance Adjustment of Relative-Risk Estimates in Matched Studies. Biometrics 38: 661-672.

17. Brown, M. B. & Benedetti, J. K. (1977): Sampling Behavior of Tests of Correlation inTwo-Way Contingency Tables. J. Amer. Statist. Assoc. 72: 309-315.

18. Bukvić, A. (2002) :Merenje intelektualnih sposobnosti, preuzeto iz zbornika: Psihološka istraživanja 2, Instituta za psihologiju F. Fak, Beograd.

19. Bulajić, M. (2002): Geodemografski model tržišnog prostora Srbije, Doktorska disertacija, Fakultet organizacionih nauka, Univerzitet u Beogradu, Beograd.

20. Charnes, A., Cooper W. W. & Rhodes E. L. (1978): Measuring the Efficiency of Decision Making Units, European Journal of Operational Research, 2(6), 429-444.

21. Charles, B. N. (1959) : Empirical models of interlevel correlation of winds, J. Meteor.,16,581-585.

22. Clrosby, D. S., Breaker, L. C. & Gemmill, W. H. (1990): A definition for vector correlation and its application to marine surface winds, National Meteorogical Center Office Note No. 365, 50pp.

23. Cramer, H. (1946): Mathematical Methods of Statistics. Princeton: Princeton University Press.

24. De Leeuw, J. (1973): Canonical analysis of categorical data, Doctoral dissertation, University of Leiden.

25. Deming, W. E. & Stephan, F. F. (1940): On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals Are Known, Ann. Math. Statist. 11:427-444.

26. Detzius, R. (1916): Extension of correlation methods and method of least squares to vectors, Sitzungsber., Akad. Wiss. Wien, 125(lia), 3-20.

27. Djoković, A., Radojicić, Z. & Vuković, N. (2007): Vector correlation coefifcient as an evaluation measure, Balcor ’07, 381-389, Zlatibor.

28. Djokovic, A., Jeremic, V. & Radojicic, Z. (2012): Towards efficient elementary school education: a Serbian perspective. Actual problems of economics, 137, 294-300.
29. Djordević, Z. (1969): Učenicki dosije i praćenje razvoja učenika, Jugoslovenski zavod za proučavanje školskih i prosvetnih pitanja, Beograd.
30. Dixon, W. J. & Massey, F. J. Jr. (1983): Introduction to statistical analysis (Tokyo: McGraw Hill), Hansell.
31. Dobrota, M., Jeremic, V., Jovanovic-Milenkovic, M. & Đokovic, A. (2012): Students' Satisfaction with Information System of Faculty of Organizational Sciences, IISES and University of Economics in Prague, Lisbon.
32. Embretson, S. E. (1996): The New Rules of Measurement, Psychological Assessment 8:4:341-349.
33. Everitt, B. S. (1997): The Analysis of Contingency Tables, London, Chapman and Hall.
34. Farewell, V. T. (1982) : A Note on Regression Analysis of Ordinal Data with Variability of Classification. Biometrika 69: 533-538.
35. Ferligoj, A. (1989): Razvrščavanje v skupine. Metodološki zvezki, 4, JUS,Ljubljana.
36. Flanders, W. D. (1985): A new variance estimator for the Mantel-Haenszel odds ratio, Biometrics 41, 637 – 642.
37. Gans, L. P. & Robertson, C. A. (1981): Distributions of Goodman and Kruskal's Gamma and Spearman's Rho in 2 x 2 Tables for Small and Moderate Sample Sizes. J. Amer.Statist. Assoc. 76:942-946.
38. Goodman, L. A. (1972): Some Multiplicative Models for the Analysis of Cross-Classified Data. Proc. 6th Berkeley Symposium on Mathematical Statistics and Probability. 1, 649-696.
39. Goodman, L. A. (1981): Association Models and Canonical Correlation in the Analysis of Cross-Classifications Having Ordered Categories. J. Amer. Statist. Assoc. 76:320-334.

40. Goodman, L. A. & Kruskal W. H. (1954) : Measures of Association for Cross Classifications. J. Amer. Statist. Assoc. 49: 732-764.
41. Guilford, J. P. (1968): Osnove psihološke i pedagoške statistike, Savremena administracija Beograd.
42. Guttman, L. (1954): Some necessary conditions for common factor analysis, Psychometrika 19:149-161.
43. Guttman, L. (1953): Image theory for the structure of quantitive variates, Psychometrika, 18, 277-296.
44. Haberman, S. J. (1981): Tests for Independence in Two-Way Contingency Tables Based on Canonical Correlation and on Linear-by-Linear Interaction. Ann. Statist. 9: 1178-1186.
45. Hooper, J. W. (1959): Simultaneous equations and cabonical correlation theory, Econometrica, 27, 245-256.
46. Hošek, A. & Radovanović, D. (1994): Klasifikacija primarnih faktora agresivnosti, Majski skup 1994, Beograd,
47. Hošek, A. (1993): Komparativna klasifikacija nekih indikatora socijalnog statusa. Zbornik radova 6 i 7 sekcije za klasifikacije Saveza statističkih društava Jugoslavije, Beograd, 237-252.
48. Hošek, A. & Momirović, K. (1994): Optimalna eksploatacija informacija koje sadrže sociometrijski podaci, Majski skup 1994, Beograd
49. Hoteling, H. (1935): The most predictable criterion. Journal of Educational Psychology, 26:139-142
50. Hotelling, H. (1959): Relation between two sets of variates, Biometrika, 28, 321-377.
51. Hotelling, H. (1933): Analysis of a complex of statistical variables into principal components, Journal of Educational Psychology, 24:417-41,498-520.
52. Ivanović, B. (1977): Teorija klasifkacije, Institut za ekonomiku industrije, Beograd.

53. Ivić, I., Milinković, M., Rosandić, R. & Smiljanić, V. (1978): Razvoj i merenje inteligencije - Tom I, Inteligencija, njen razvoj i merenje, drugo izdanje, Zavod za udžbenike i nastavna sredstva, Beograd.

54. Jeremić, V., Đoković, A., Mladenović, N. & Radojičić, Z. (2011): New method for ranking chess Olympics teams. 10th Balkan Conference on Operational Research-BALCOR 2011.

55. Jeremić, V., Vukmirović, D., Radojičić, Z. & Đoković, A. (2011): Towards a framework for evaluating ICT infrastructure of countries: a Serbian perspective. Metalurgia International, 16(9), 15-18.

56. Jeremić, V., Bulajić, M., Marković, A. & Đoković, A. (2011): Indeks razvijenosti e-Uprave kao ključni indikator razvijenosti IKT infrastrukture. SPIN 2011, Beograd, 563-569.

57. Kaiser, H. F. (1958): The varimax criterion for the analytic rotation in factor analysis. Psychometrika, 23:187-200.

58. Kaiser, H. F. & Michael, W. B. (1975): Domain validity and generalizability, Educational and Psychological Measurement, 35, 1, 31-35.

59. Kaiser, H. F. & Caffrey, Y. (1965): Alpha factor analysis, Psychometrika, 30, 1-44.

60. Kaufman, L. & Rousseeuw, P. J. (1990): Finding groups in data: An introduction to cluster analysis, John Wiley, New York..

61. Kendall, G. K. & Stuart, A. (1967) : The Advanced Theory of Statistics Vol. 2.2d ed. Hafner Publishing Company, 690 pp.

62. Kendall, M. G. (1938): A New Measure of Rank Correlation. Biometrika 30:81-93.

63. Kendall, M. G. (1945): The Treatment of Ties in Rank Problems. Biometrika 33:239-251.

64. Kendall, M. G. (1970): Rank Correlation Methods. 4th ed. London: Griffin.

65. Knezević, G. & Momirović, K. (1996): RTT9G, program za analizu metrijskih karakteristika kompozitnih mernih instrumenata. U P. Kostid, Problemi

merenja u psihologiji, 2,37-56. Institut za kriminoIoška i sociološka istraživanja, Beograd.

66. Knezević, G. & Momirović, K. (1996): Algoritam i program (QCCR) za analizu relacija kanoničke korelacijske analize i kanoničke analize kovarijansi. U P. Kostić, Problemi merenja u psihologiji, 2, 51-1 A, Institut za kriminološkaa i sociološka istraživanja, Beograd.

67. Kovačevič, P., Wolf, B., Momirović, K. & Hosek, A. (2001): Distribucija količnika inteligencije nakon eliminacije unikne varijanse testova, Psihologija, XXXIV.

68. Kovačić, Z. (1992): Multivarijaciona analiza, Ekonomski fakultet, Beograd.

69. Kvaščev, R. (1980): Sposobnosti za učenje i ličnost, Zavod za udžbenike i nastavna sredstva, Beograd.

70. Lancaster, H. O. & M. A. Hamdan. (1964): �stimation of the Correlation Coefficient in Contingency Tables with Possibly Nonmetrical Characters." Psychomttrika 29:383-391.

71. Lachenbruch, P. A. (1975): Discriminant Analysis, Hafner, New York.

72. Maletić, P., Kreća, M., Jeremić, V. & Đoković, A. (2011): Ranking of municipalities in Vojvodina through development level of SM� in agribusiness. SYM-OP-IS 2011, Zlatibor, 543-546.

73. Maletic, P., Kreca, M., Jeremic, V., Bulajic, M. & Đokovic, A. (2012): The ranking of municipaities in Serbia through the development level of SM� in agribusiness. Int. J. Agricult. Stat. Sci., 8(1), 7-13.

74. Mantel, N. (1963): Chi-Squared Tests with One Degree of Freedom; �xtensions of the Mantel-Haenszel Procedure. J. Amer. Statist. Assoc. 58:690-700.

75. Martić, M. (1999): Analiza obavijenih podataka sa primerima, Doktorska disertacija, Fakultet organizacionih nauka, Univerzitet u Beogradu, Beograd.

76. Mardia K. V., Kent J. T. & Bibby J .M. (1979): Multivariate Analysis, Academic Press, New York.

77. Maxwell, A. E. (1977): Multivariate Analysis in Behavioural Research, Chapman and Hall, London.

78. Milenković, N., Jeremić, V., Đoković, A. & Dobrota, M. (2011): Statistički pristup merenju socio-ekonomske razvijenosti MENA zemalja. SPIN 2011, Beograd, 554-559.

79. Momirović, K. & Fajgelj, S. (1994): Faktorska analiza nominalnih varijabli, Sociološki pregled, 21:1, 369-384.

80. Momirović, K. (1988): Komparativna analiza nekih mera asocijacije izmedu dva skupa kvantitativnih varijabli, Tehnički izveštaj, Institut za kriminološka i sociološka istraživanja, Beograd.

81. Momirović, K. (1988): Uvod u analizu nominalnih varijabli, Metodološke sveske, Jugoslovensko udruženje za sociologiju, Ljubljana.

82. Momirović, K. & Dobrić, V. (1984): O nekim odnosima između kanoničke i kvazikanoničke diskriminativne analize. Biokibernetika, 5:17-22.

83. Momirović, K. & Hošek, A. (1994): Jedna primitivna mera sličnosti između dve otvorene razlivene klasifikacije. Majski skup 1994, Beograd.

84. Nikodijevic, A., Anđelkovic-Labrovic, J. & Đokovic, A. (2012): Sindrom sagorevanja među studentima Fakulteta organizacionih nauka,, Management, 64, 47-53.

85. Novick, M. R. (1966): The axioms and principal results of classical test theory, Journal of Mathematical Psychology, 3:1-18.

86. Olsen, J. B. (1990): Applying computerized adaptive testing in schools. Measurement & Evaluation in Counseling & Development, Vol. 23:1.

87. Paskota, M. (2002): Nominalne promenljive u diskriminacionoj analizi, Doktorska disertacija, Ekonomski fakultet, Univerzitet u Beogradu.

88. Petrovic-Đorđevic, D., Đokovic, A. & Savic, G. (2010): Merenje tehnicke efikasnosti fudbalske reprezentacije Srbije u utakmicama kvalifikacija za SP 2010. SymOrg2010, Zlatibor.

89. Press, S .J. (1949): Linear combinations of non-central chi-square variates, Ann. Math. Stat. 37: 480-487.

90. Quade, D. (1974): Nonparametric Partial Correlation. Chapter 13 in Measurement in the Social Sciences. Ed. by H. M. Blalock. Chicago: Aldine.

91. Radojičić, Z., Vuković, N. & Vukmirović, D. (2001): Određivanje "Zone osetljivosti", SymOpIs '01, Beograd.

92. Radojičić, Z., Janić, B. & Vukmirović, D. (1995): Statistical Approach to Define Activity Index of Disease , 3rd Balkan Conference of Operational Research, Thessaloniki, Greece.

93. Radojičić, Z., Stefanović, T. & Vukmirović, D. (1988): Rangiranje preduzeća metodom Ivanovićevog odstojanja, Analiza grupisanja IV, SZS, Kosmaj.

94. Radojičić, Z., Vuković, N. & Vukmirović, D. (2003): Applying Coefficients of Preference in Ranking (CPR), YUJOR Vol 13, No2. Belgrade.

95. Radojičić, Z. (1994): Primena metode nehijerarhijskog klasifikovanja u izboru računarske opreme, Diplomski rad, Beograd.

96. Radojičić, Z. (2001): Statističko merenje intenziteta pojava, Magistarski rad, Fakultet organizacionih nauka, Univerzitet u Beogradu, Beograd.

97. Radojičić, Z. (2007): Statistički model ocenjivanja na subjektivno procenjenim karakteristikama, Doktorska disertacija, Fakultet organizacionih nauka, Univerzitet u Beogradu, Beograd.

98. Radojičić, Z. (1997): Srećni, manje srećni i oni koji to nisu, Analiza grupisanja III, Sirogojno, Savezni zavod za statistiku.

99. Raju, N. S. & Drasgow, F. (1993): An Empirical Comparison of the Area Methods, Chi-square Test, and the Mantel-Haenszel Technique for Assessing Differential Functioning, Educational & Psychological Measurement, 53(2), 301—321.

100. Rao, C. R. (1965): The Use and Interpretation of Principal Component Analysis in Aplied Research, Sakhya.

101. Rao, C. R. (1955): Estimation and tests of significance in factor analysis, Psychometryc.

102. Reynolds, H. T. (1985): Analysis of nominal data, 2. Printing, Sage publications, Beverly Hills.

103. Rohatgi, V. K. (1976): An Introduction to Probability Theory and Mathematical Statistics, John Wiley and Sons, 684 pp.
104. Ruben, H. (1963): A new result on the distribution of quadratic forms, Ann. Math. Stat., 34:1582-1584.
105. Shrout, P. E. & Fleiss, J. L. (1979): Intraclass correlations: Uses in assessing rater reliability, Psychological Bulletin, 86, 420-428.
106. Sidick, J. T., Barrett, G. V. (1994): Three-alternative multiple choice tests: An attractive option, Personnel Psychology, 47(4), 829-836.
107. Somers, R. H. (1974): Analysis of Partial Rank Correlation Measures Based on the Product-Moment Model: Part One. Social Forces 53:229-246.
108. Spearman, C. E. (1904): The proof and measurement of association between two things American Journal of Psychology, 15:72-101.
109. Spearman, C. E. (1904): General intelligence, objectively determined and measured, American Journal of Psychology, 15:201-293.
110. Stephens, M. A. (1979): Vector correlation, Biometrika, 66, 41-48.
111. Steiger, H. J. (1980): Testing Pattern Hypotheses on Correlation Matrices:Alternative statistics and some empirical results, University of British Columbia.
112. Stevens, S. S. (1951): Mathematics, measurement, and psychophysics. U S.S. Stevens (Ur.), Handbook of experimental psychology, 1-49, Wiley, New York.
113. Stuart, A. (1963): Calculation of Spearman's Rho for Ordered Two-Way Classifications. Amer. Statist. 17:23-24.
114. Suknović, M., Čupić, M. & Radojičić, Z. (2002): The Application Of The Group Decision Making Model, 6th Balkan Conference of Operational Research, Thessaloniki, Greece.
115. Tadin, I. (1969): Baterija varijabli za prognozu uspjeha u školama II stupnja, Republički zavod za zapošljavanje, Zagreb.
116. Thurstone L. L. (1926): The scoring of individual performance, Journal of Educational Psychology, 17, 446-457.
117. Thurstone, L. L. (1931): The reliability and validity of tests, Edwards, Ann Arbor.

118. Thurstone, L. L. (1934): The Vectors of Mind, Psychological Review, 41, 1-32.
119. Trišić, B. & Delibašić B, (2010): Generički algoritam za klasterovanje, SymOpIs '10, Tara.
120. Vukmirović, D., Vuković, N., Marković, A. & Radojičić, Z. (1994): Skraćeni metod hijerarhijskog klasifikovanja, Zbornik radova SymOpIs '94, Kotor, Fakultet organizacionih nauka, Beograd.
121. Vukmirović, D. (1992): Model hijerarhijskog klasifikovanja, Ekonomski fakultet, Beograd.
122. Vuković, N. (1977): Generalizacija višestrukog i kolektivnog koeficijenta korelacije, Doktorska disertacija, PMF, Novi Sad.
123. Vuković, N. (2001): nswer Tree and VCC, Euro 2001, Roterdam, Holland.
124. Vuković, N. (2000): PC verovatnoća i statistika, Fakultet organizacionih nauka, Beograd.
125. Vuković, N. (2001): Vektorski koeficijent korelacije uzorka - primena i implementacija, Statistička revija, Beograd.
126. Vuković, N. (1987): Statistička analiza, Naučna knjiga, Beograd.
127. Wainer, H. & Kiely, G. L. (1987): Item clusters and computerized adaptive testing: case testlets. Journal of Educational Measurement, 24, 185-201.
128. Watson, G. S. (1960): More significanse tests on the sphere, Biometrica, 47, 87-91.
129. Welsh, W. B. & Betz, N. E. (2001): Tests and ssessment, 4th ed., Prentice-Hall International London.
130. Wolf, B. & Momirović, K. (1994): Neke varijacije na Cramer-Hotellingovu temu, Majski skup 1994, Beograd
131. Wylie, D. P., Hinton, B. B., Howland M. H. & Lord, R. J. (1985): utocorrelation of wind observations, Mon. Wea. Rev., 113, 849-857.

PRILOG

U delu disertacije razmatrana je problematika slaganja I-odstojanja sa normalnom raspodelom. Za rešavanje ovog problema korišćen je Bootstrap metoda koja kao svoj sastavni deo podrazumeva primenu Monte-Karlo simulacije. Rezultati dobijeni ovim elsperimentima dati su u ovom prilogu. Takođe za kreiranja algoritma nehijerarhijske klasifikacije sa unapred definisanim ograničenjima, koji je opisan u trećem poglavlju disertacije, napisano je programsko rešenje i ono je dato u ovom prilogu.

Prilog 1.

Tabela 8.1 Rezultat generisanja slučajnih promenljivih sa normalnom raspodelom

n1	n2	n3	n4	n5
-0.18829	0.27268	-1.64371	-0.043	-0.21261
-0.20511	1.738496	-0.35964	-1.01647	-0.66241
-1.11599	-1.51504	0.717622	0.895226	1.008637
0.289207	-0.64003	0.813179	1.058872	0.915164
0.382071	0.979485	0.46075	-1.71699	-0.69608
0.436786	-0.79509	-0.78627	-0.28607	-0.33325
2.027344	-0.70473	-0.59347	1.344297	0.634564
-1.60311	-0.19463	-2.16689	-1.93507	0.164311
0.429958	0.271231	1.388232	-0.17788	1.234287
0.232387	0.917507	0.094661	0.164589	-0.72312
-0.38495	1.495321	-1.68583	0.152748	0.737577
-1.64418	0.803908	1.219592	-0.25156	-1.27627
-1.35995	1.159547	-0.82683	1.06888	-0.66865
0.027254	1.903868	-0.79186	0.509306	2.009235
1.171467	-0.39852	0.550635	-1.05589	0.543964

0.550871	-1.60132	-2.16623	0.195	1.166417
0.131854	-0.18219	0.7165	-0.12817	-0.90996
-1.9922	0.481941	0.771551	-1.03396	-0.24464
0.506145	0.283005	1.005337	-0.6858	-0.76295
0.679451	-0.31664	1.633918	-0.06622	0.036601
0.144679	0.537208	-1.21781	0.583158	-0.11721
0.287627	-0.30062	1.179327	0.778012	0.484095
-0.85725	0.81091	1.113308	-0.37328	-0.37984
-0.26606	-0.20809	0.723656	0.250103	1.138898
-1.27402	-0.14289	-1.00633	2.186187	0.447952
-2.00944	1.309502	-1.37289	0.957022	0.653915
-0.46877	1.079678	0.525024	0.978662	-1.88921
0.331002	-0.56644	-1.46853	-0.3052	0.8986
0.886734	0.046945	0.890319	0.657713	0.769639
-0.41407	-0.62376	0.709287	1.132815	-0.09813
-0.95383	-0.55077	0.156851	0.616054	0.022404
1.000621	0.419478	-0.17244	-1.67396	-1.16776
2.586554	0.436213	0.421992	0.834269	0.137135
2.079089	0.023642	0.131689	0.235328	-0.09917
0.022024	0.308012	-1.02711	-0.38175	-0.75546
1.173436	1.38818	-1.47345	-0.88556	-0.31772
-0.25687	-0.62582	-1.81293	0.735561	0.465313
1.146488	1.124682	-0.48999	0.08566	0.344599
-0.74249	0.594202	-1.12062	0.766015	-0.70264
0.594915	3.111612	-0.43968	0.029602	0.199428
-1.46026	-0.61791	0.582484	-0.46596	1.186153
-0.77625	-2.36591	1.326916	0.513789	0.248911
-0.90026	0.755502	0.315064	-2.05152	0.585126
0.070464	0.105252	1.035113	0.876297	0.625582
-0.67777	0.644401	-0.90751	0.468976	1.678491
-0.68304	-1.77837	0.561016	0.525693	0.828143
-0.02642	-1.02992	-0.9781	-0.68576	0.59872

1.087824	-2.38661	0.758848	0.759852	-0.60575
0.313079	0.481008	-0.19715	-0.61312	0.733692
1.109502	-0.63444	-0.08457	-0.82012	-0.09757
-0.63129	-1.2718	-0.51503	-2.18042	-0.04155
-1.65178	1.202452	1.697981	0.672768	0.442428
-1.28442	0.468543	0.037959	1.439153	0.13349
-0.39469	-1.61232	0.541547	0.568409	0.210679
0.312064	-0.32656	-0.68347	-0.44245	0.252641
1.212479	1.30546	0.22962	-0.70304	1.856797
0.653622	0.280924	0.195125	0.081057	-1.68093
1.034456	-0.20613	0.377673	0.757909	-0.84516
0.620889	-0.01982	1.516296	-1.54512	-0.43442
-0.70893	-1.90358	0.212693	-0.82974	-0.27053
0.212994	0.007905	-0.16026	0.135019	0.866447
-1.28258	0.171669	-1.09435	-0.40541	-0.61591
0.51311	0.646863	0.523656	-0.34523	-2.04632
0.186914	0.456818	-0.34962	-0.13407	-0.84692
0.23548	-0.4311	-1.94944	1.458571	0.770853
-0.17731	0.484569	1.062724	0.584689	-0.20166
-0.41777	-0.14435	0.197162	-2.02245	-1.23886
0.134179	0.254003	-0.03012	-0.88059	-0.11847
-0.91009	-0.07449	0.263471	-0.80312	-1.52562
1.506474	1.226785	-1.36957	1.529556	0.139233
0.568888	-0.81313	0.519474	-1.0435	-0.6548
-0.88913	-0.2176	0.183985	1.338265	-1.0372
-1.98052	0.587587	-0.16569	-0.01129	-0.19671
-1.52334	-0.43766	-0.76108	2.34262	0.852083
0.378277	-0.03683	1.094764	0.663965	-0.86253
1.453574	-1.18271	0.511349	-0.05979	-1.20297
-0.18134	1.244519	1.236304	0.172928	-1.03461
0.533314	-1.23476	0.05234	0.998349	-0.70544
3.647896	-0.38557	0.144631	1.987757	1.58188

0.691927	0.304595	-2.01822	1.169665	1.531342
0.409814	0.211409	-1.70632	0.65712	0.937702
-1.14778	-0.55431	1.128208	-0.90996	0.308611
-1.26954	-0.63428	0.956356	-0.6944	-0.78486
-0.7799	-0.49256	0.938077	0.021282	0.537355
-1.57817	-0.62796	1.365064	0.484062	0.07263
-2.09572	-0.56034	1.056258	-0.48443	1.148667
-2.4442	-0.12523	-0.22543	0.72142	-0.17527
-1.78085	0.365346	-0.00391	-0.49417	-2.38405
-0.24233	-0.6212	-0.20508	2.51263	-0.2416
1.432344	0.548351	-1.20186	-0.80315	-0.52584
-0.76817	1.368059	0.079671	1.102633	-0.11169
-1.47571	0.99121	-0.95425	-0.6391	0.985352
-1.21652	0.122157	1.199297	0.891405	-0.45822
-0.64973	0.090569	-0.77675	-0.00501	-0.29182
-1.69708	1.808027	-0.13807	0.760613	-1.46685
-0.9275	-0.51884	0.212624	1.41662	-0.92006
-0.20859	1.078626	-0.44576	-0.07591	-0.03161
-1.19559	0.081313	-0.09647	-0.32922	1.240154
-1.11137	-0.77864	1.499247	-1.30149	-1.82018
-0.46223	-0.32284	0.834936	0.56748	0.209717

Tabela 8.2 Test slaganja I-odstojanja (za varijable po normalnoj raspodeli)

Kolmogorov-Smirnov test			
		I2_MIN1234	I2_MIN3142
Veličina uzorka		100	100
Parametri raspodele	Sredina	25.2894	25.4233
	St. devijacija	8.82672	8.87536
Vrednost za Kolmogorov-Smirnov test		.736	.798
signifikantnost		.650	.547

Tabela 8.3 Rezultat generisanog 4-dim vektora sa uniformnom raspodelom

u1	u2	u3	u4
0.750928	0.856975	0.579418	0.657854
0.263556	0.872306	0.289486	0.573945
0.658842	0.545747	0.29853	0.863767
0.060528	0.204571	0.836798	0.246708
0.65532	0.715526	0.758816	0.805896
0.925769	0.496676	0.346443	0.510693
0.661765	0.253133	0.901172	0.019549
0.981662	0.06212	0.090517	0.87547
0.660873	0.459048	0.509996	0.891765
0.37722	0.864146	0.552928	0.466797
0.350849	0.89313	0.995619	0.020274
0.779145	0.378753	0.103749	0.743317
0.418158	0.771398	0.198195	0.250843
0.665188	0.373603	0.888455	0.694881
0.117393	0.516353	0.603234	0.985759
0.277987	0.659362	0.589629	0.939048
0.911849	0.309559	0.275837	0.45409
0.924408	0.093649	0.246346	0.998585
0.592595	0.810174	0.317176	0.409851
0.294615	0.215629	0.051015	0.204627
0.381979	0.626302	0.48245	0.117559
0.126293	0.240365	0.669931	0.543282

0.270448	0.168391	0.871085	0.817426
0.186965	0.614711	0.29984	0.36199
0.157254	0.877708	0.900253	0.252246
0.965481	0.389205	0.17742	0.216945
0.693268	0.981566	0.235196	0.525585
0.373443	0.419732	0.825789	0.567056
0.076515	0.949519	0.24057	0.301658
0.844967	0.414385	0.031396	0.071434
0.150059	0.931784	0.004545	0.132584
0.967446	0.401619	0.087546	0.782046
0.982856	0.795998	0.5799	0.984993
0.160799	0.89957	0.864268	0.75465
0.14597	0.620967	0.619628	0.094362
0.157113	0.607247	0.305711	0.684598
0.894059	0.873981	0.8149	0.114388
0.068486	0.184141	0.102192	0.360123
0.462052	0.273098	0.411555	0.225109
0.650269	0.099343	0.132569	0.231305
0.906857	0.847843	0.206021	0.856048
0.42796	0.446917	0.977839	0.735486
0.091517	0.320116	0.965487	0.693655
0.951737	0.61591	0.071176	0.349866
0.777944	0.400329	0.812734	0.717876
0.610545	0.198519	0.408772	0.428824
0.382151	0.879823	0.555587	0.297969
0.515907	0.317453	0.575574	0.334336
0.071689	0.56463	0.729747	0.164079
0.88551	0.415533	0.805655	0.529357
0.11287	0.072316	0.156708	0.365851
0.132122	0.265853	0.821942	0.024029
0.088182	0.55315	0.301538	0.211734
0.700583	0.135273	0.63173	0.291577

0.722664	0.6511	0.265794	0.035468
0.10426	0.977832	0.927735	0.777742
0.225764	0.181142	0.535588	0.591009
0.455378	0.065383	0.852404	0.18211
0.548762	0.525922	0.344466	0.793969
0.99519	0.896183	0.016979	0.015173
0.434145	0.464935	0.099199	0.263563
0.631099	0.025999	0.287992	0.555257
0.122163	0.710015	0.725808	0.493054
0.066305	0.782335	0.212039	0.799204
0.82453	0.5405	0.240129	0.809513
0.160047	0.62794	0.769938	0.279324
0.000593	0.210746	0.524915	0.870981
0.867866	0.399263	0.333661	0.48562
0.143869	0.589612	0.869651	0.359312
0.395288	0.120812	0.209327	0.975788
0.056432	0.57176	0.2472	0.617296
0.632669	0.908435	0.37902	0.345633
0.20745	0.732904	0.285266	0.37012
0.367812	0.774991	0.933224	0.994876
0.192921	0.999139	0.020291	0.763753
0.767315	0.901683	0.224674	0.858414
0.05616	0.194315	0.449349	0.264418
0.828706	0.30592	0.799405	0.639298
0.686221	0.843715	0.978309	0.770736
0.723812	0.639052	0.248492	0.522157
0.635878	0.699542	0.406525	0.198891
0.603177	0.833281	0.96804	0.025352
0.544362	0.034478	0.172001	0.761618
0.524566	0.815286	0.134557	0.181836
0.346101	0.886214	0.477318	0.587693
0.700461	0.694308	0.433875	0.535461

0.533265	0.494282	0.649592	0.736879
0.578722	0.877164	0.264497	0.479182
0.978452	0.80682	0.498963	0.323116
0.509587	0.639236	0.782765	0.634204
0.305137	0.039136	0.725873	0.863315
0.876863	0.342545	0.241924	0.830125
0.016282	0.955174	0.865497	0.734994
0.036689	0.248069	0.582994	0.714881
0.028902	0.747577	0.915389	0.652609
0.819474	0.403355	0.506822	0.531873
0.439156	0.949925	0.291225	0.045705
0.243042	0.992852	0.732055	0.342728
0.80857	0.235266	0.429226	0.918869
0.128288	0.8336	0.492681	0.137916

Tabela 8.4 Test slaganja I-odstojanja (za uniformnu raspodelu)

Kolmogorov-Smirnov Test			
		I2_MIN1234	I2_MIN4231
N		100	100
Normal Parameters[a,b]	Mean	15.1326	15.1224
	Std. Deviation	7.02710	7.06252
Kolmogorov-Smirnov Z		.880	.705
Asymp. Sig. (2-tailed)		.420	.703

Tabela 8.5 Rezultat generisanog 5-dim vektora sa uniformnom raspodelom

u1	u2	u3	u4	u5
0.750928	0.856975	0.579418	0.657854	0.50573
0.263556	0.872306	0.289486	0.573945	0.999156
0.658842	0.545747	0.29853	0.863767	0.747955
0.060528	0.204571	0.836798	0.246708	0.930314
0.65532	0.715526	0.758816	0.805896	0.101333
0.925769	0.496676	0.346443	0.510693	0.866384
0.661765	0.253133	0.901172	0.019549	0.558406
0.981662	0.06212	0.090517	0.87547	0.139485
0.660873	0.459048	0.509996	0.891765	0.571129
0.37722	0.864146	0.552928	0.466797	0.266191
0.350849	0.89313	0.995619	0.020274	0.276648
0.779145	0.378753	0.103749	0.743317	0.807703
0.418158	0.771398	0.198195	0.250843	0.926271
0.665188	0.373603	0.888455	0.694881	0.962351
0.117393	0.516353	0.603234	0.985759	0.957111
0.277987	0.659362	0.589629	0.939048	0.455485
0.911849	0.309559	0.275837	0.45409	0.747496
0.924408	0.093649	0.246346	0.998585	0.043083
0.592595	0.810174	0.317176	0.409851	0.35072
0.294615	0.215629	0.051015	0.204627	0.128592
0.381979	0.626302	0.48245	0.117559	0.870084
0.126293	0.240365	0.669931	0.543282	0.501868
0.270448	0.168391	0.871085	0.817426	0.78029
0.186965	0.614711	0.29984	0.36199	0.531594
0.157254	0.877708	0.900253	0.252246	0.82833
0.965481	0.389205	0.17742	0.216945	0.582121
0.693268	0.981566	0.235196	0.525585	0.342931
0.373443	0.419732	0.825789	0.567056	0.800458
0.076515	0.949519	0.24057	0.301658	0.019798

0.844967	0.414385	0.031396	0.071434	0.586346
0.150059	0.931784	0.004545	0.132584	0.598034
0.967446	0.401619	0.087546	0.782046	0.412151
0.982856	0.795998	0.5799	0.984993	0.747349
0.160799	0.89957	0.864268	0.75465	0.873824
0.14597	0.620967	0.619628	0.094362	0.940546
0.157113	0.607247	0.305711	0.684598	0.254071
0.894059	0.873981	0.8149	0.114388	0.486805
0.068486	0.184141	0.102192	0.360123	0.699098
0.462052	0.273098	0.411555	0.225109	0.253114
0.650269	0.099343	0.132569	0.231305	0.732161
0.906857	0.847843	0.206021	0.856048	0.321557
0.42796	0.446917	0.977839	0.735486	0.889563
0.091517	0.320116	0.965487	0.693655	0.886446
0.951737	0.61591	0.071176	0.349866	0.350207
0.777944	0.400329	0.812734	0.717876	0.920855
0.610545	0.198519	0.408772	0.428824	0.523284
0.382151	0.879823	0.555587	0.297969	0.850881
0.515907	0.317453	0.575574	0.334336	0.966886
0.071689	0.56463	0.729747	0.164079	0.05708
0.88551	0.415533	0.805655	0.529357	0.730547
0.11287	0.072316	0.156708	0.365851	0.798935
0.132122	0.265853	0.821942	0.024029	0.169374
0.088182	0.55315	0.301538	0.211734	0.444919
0.700583	0.135273	0.63173	0.291577	0.102553
0.722664	0.6511	0.265794	0.035468	0.870347
0.10426	0.977832	0.927735	0.777742	0.239708
0.225764	0.181142	0.535588	0.591009	0.399511
0.455378	0.065383	0.852404	0.18211	0.213328
0.548762	0.525922	0.344466	0.793969	0.986623
0.99519	0.896183	0.016979	0.015173	0.012274
0.434145	0.464935	0.099199	0.263563	0.940092

0.631099	0.025999	0.287992	0.555257	0.428255
0.122163	0.710015	0.725808	0.493054	0.335277
0.066305	0.782335	0.212039	0.799204	0.700629
0.82453	0.5405	0.240129	0.809513	0.376497
0.160047	0.62794	0.769938	0.279324	0.608378
0.000593	0.210746	0.524915	0.870981	0.967356
0.867866	0.399263	0.333661	0.48562	0.317071
0.143869	0.589612	0.869651	0.359312	0.041553
0.395288	0.120812	0.209327	0.975788	0.904981
0.056432	0.57176	0.2472	0.617296	0.708068
0.632669	0.908435	0.37902	0.345633	0.575231
0.20745	0.732904	0.285266	0.37012	0.821219
0.367812	0.774991	0.933224	0.994876	0.698728
0.192921	0.999139	0.020291	0.763753	0.091421
0.767315	0.901683	0.224674	0.858414	0.013914
0.05616	0.194315	0.449349	0.264418	0.818889
0.828706	0.30592	0.799405	0.639298	0.279613
0.686221	0.843715	0.978309	0.770736	0.507461
0.723812	0.639052	0.248492	0.522157	0.540578
0.635878	0.699542	0.406525	0.198891	0.461198
0.603177	0.833281	0.96804	0.025352	0.383324
0.544362	0.034478	0.172001	0.761618	0.275328
0.524566	0.815286	0.134557	0.181836	0.230409
0.346101	0.886214	0.477318	0.587693	0.519244
0.700461	0.694308	0.433875	0.535461	0.735426
0.533265	0.494282	0.649592	0.736879	0.576673
0.578722	0.877164	0.264497	0.479182	0.602894
0.978452	0.80682	0.498963	0.323116	0.904987
0.509587	0.639236	0.782765	0.634204	0.536602
0.305137	0.039136	0.725873	0.863315	0.957674
0.876863	0.342545	0.241924	0.830125	0.190969
0.016282	0.955174	0.865497	0.734994	0.712389

0.036689	0.248069	0.582994	0.714881	0.713175
0.028902	0.747577	0.915389	0.652609	0.040115
0.819474	0.403355	0.506822	0.531873	0.740812
0.439156	0.949925	0.291225	0.045705	0.39455
0.243042	0.992852	0.732055	0.342728	0.922289
0.80857	0.235266	0.429226	0.918869	0.09928
0.128288	0.8336	0.492681	0.137916	0.677347

Tabela 8.6 Test slaganja I-odstojanja (za uniformnu raspodelu)

Kolmogorov-Smirnov Test				
		I2_MIN12345	I2_MIN24315	I2_MIN23415
Veličina uzorka		100	100	100
Parametri raspodele	sredina	19.0698	19.0598	19.0577
	St. devijacija	7.22488	7.24483	7.24400
Kolmogorov-Smirnov test statistika		.848	.694	.702
signifikantnost		.469	.721	.709

Tabela 8.7 Rezultat generisanog 4-dim vektora sa eksponencijalnom raspodelom

e1	**e2**	**e3**	**e4**
0.112478	0.359618	2.196645	0.336446
0.533361	1.068087	0.386037	1.244321
0.287232	1.427853	0.154303	1.389536
1.567823	0.25596	0.693378	0.838251
1.637152	1.3595	0.686049	1.157358
0.382284	0.230743	0.344393	1.295957
0.400155	0.591147	0.063327	0.484177
0.140096	0.106103	2.649426	1.256369
0.059389	0.313796	0.951002	3.306978
0.360698	0.664775	0.418902	1.006028
1.725106	0.059345	0.161898	0.279855
2.123822	0.904846	0.57834	1.757696
0.074966	1.1209	2.561834	0.202663
0.61177	0.218133	0.039537	0.679408
1.160347	0.476521	2.082487	1.074326
2.508776	0.170652	0.881036	2.597386
0.462286	2.523516	0.086385	2.694184
3.416034	1.452127	1.38766	2.906853
0.151939	0.907111	1.76672	0.881939
2.219438	0.07195	0.526251	0.621322
3.074658	0.579229	0.184102	1.279868
0.588508	2.788128	0.163977	0.720288
1.36696	0.749009	2.107655	0.021043
0.400322	0.597199	0.199617	0.326224
1.046992	1.147977	0.811818	2.653621
1.149165	0.011386	0.313108	0.353726
0.12428	0.2782	2.530624	0.353287
2.074256	0.007937	0.4112	1.59742

0.392099	1.107804	2.247781	0.307224
1.701219	0.325875	0.539727	0.379753
0.224434	1.553748	0.718113	1.19199
0.048064	1.824698	3.065749	0.667236
1.022045	0.137668	0.178684	0.064878
0.604435	0.425377	0.987098	0.436582
3.576425	0.308109	0.49484	0.1259
2.692184	0.771681	0.371021	0.122222
1.614054	1.338874	1.620623	0.401148
1.429995	0.096377	0.345196	0.162366
1.070883	2.077612	0.649609	0.04518
0.021371	0.453998	4.369118	1.901445
0.400872	1.842166	0.355855	0.285736
0.897106	1.186114	0.169439	1.714384
0.528333	0.202413	1.287711	0.125179
0.920513	0.79368	0.76663	0.973867
1.288973	0.060237	3.000044	0.032175
0.109605	1.654016	0.982619	0.56275
1.155187	0.722817	0.051396	0.664271
0.578155	0.539419	0.534118	2.057596
2.264051	0.555491	2.476347	0.239363
2.07185	0.416767	0.882515	0.538518
4.291943	0.428841	1.288611	0.754633
1.359087	0.681121	0.234289	1.484256
0.235063	0.165603	0.021356	0.551388
0.144301	1.064413	1.175353	0.754111
0.059802	1.28368	1.715012	0.989011
0.258795	1.173518	1.137141	0.496917
2.9542	0.88857	1.177929	0.130595
0.204523	2.619797	0.080113	0.517011
3.374453	3.112975	3.25491	1.282775
3.924585	0.371989	0.008921	0.732742

0.421207	2.129955	2.954681	0.949918
0.309583	0.185046	1.265735	0.38211
0.535462	1.613503	0.188712	0.424233
0.416505	0.258731	1.658877	0.004347
1.184293	0.020952	0.919574	1.056874
0.751082	3.753407	0.582184	0.932324
2.94842	0.791042	2.370507	0.37296
0.900014	0.855607	0.661562	0.039265
3.553019	0.996863	3.502837	1.771868
0.664897	0.131541	0.593874	0.458191
1.061394	1.139625	0.588278	0.00616
0.00355	0.266202	0.988082	1.083858
0.235712	1.914072	3.47427	0.969162
0.279309	2.299093	0.05283	1.229073
2.38438	1.069331	3.208228	0.25529
0.071176	0.126251	0.8306	0.276486
1.007534	0.136438	3.279504	0.180006
0.029366	2.662716	0.793603	0.135242
0.090734	1.359176	1.716074	1.177372
3.657841	2.116441	0.058023	0.072209
0.397388	0.257351	0.671444	0.53594
0.658519	0.234494	5.097753	3.0982
1.070542	0.048351	0.611334	0.180838
2.020703	0.077147	0.499556	2.440088
0.396532	0.068443	0.162576	2.567871
0.691625	0.860069	0.211184	0.372005
1.18067	1.564699	0.68928	3.528308
1.920802	1.552614	2.418828	0.445807
0.227212	0.93523	1.453732	0.012001
1.619449	1.608159	2.324176	1.002299
0.080805	0.937196	2.912455	2.639572
0.211857	0.147708	0.190302	1.029105

1.372726	0.209296	0.155431	0.080266
1.389949	1.210217	0.008005	1.296866
0.257229	0.056888	1.226955	0.424233
1.561806	0.752028	0.459675	0.154586
0.659582	0.520779	0.946269	0.525156
0.328723	3.562712	0.229409	1.520866
1.371698	0.675922	0.031669	0.497509
2.263215	2.943534	0.971473	0.527706

Tabela 8.8 Test slaganja I-odstojanja (za eksponencijalnu raspodelu)

Kolmogorov-Smirnov Test			
		I2_MIN1234	I2_MIN3142
Veličina uzorka		100	100
Parametri raspodele	Sredina	7.2881	7.2812
	St. devijacija	8.50378	8.50284
Kolmogorov-Smirnov test statistika		2.031	2.032
signifikantnost		.001	.001

Tabela 8.9 Rezultat generisanog 5-dim vektora sa eksponencijalnom raspodelom

e1	e2	e3	e4	e5
0.112478	0.359618	2.196645	0.336446	2.591104
0.533361	1.068087	0.386037	1.244321	1.981975
0.287232	1.427853	0.154303	1.389536	1.295581
1.567823	0.25596	0.693378	0.838251	1.065569
1.637152	1.3595	0.686049	1.157358	1.641384
0.382284	0.230743	0.344393	1.295957	0.361904
0.400155	0.591147	0.063327	0.484177	0.852133
0.140096	0.106103	2.649426	1.256369	0.516019
0.059389	0.313796	0.951002	3.306978	0.143503
0.360698	0.664775	0.418902	1.006028	0.461126
1.725106	0.059345	0.161898	0.279855	0.539037
2.123822	0.904846	0.57834	1.757696	0.105489
0.074966	1.1209	2.561834	0.202663	0.278889
0.61177	0.218133	0.039537	0.679408	1.347625
1.160347	0.476521	2.082487	1.074326	0.306599
2.508776	0.170652	0.881036	2.597386	2.099029
0.462286	2.523516	0.086385	2.694184	3.979743
3.416034	1.452127	1.38766	2.906853	1.459977
0.151939	0.907111	1.76672	0.881939	2.99299
2.219438	0.07195	0.526251	0.621322	0.189778
3.074658	0.579229	0.184102	1.279868	0.320022
0.588508	2.788128	0.163977	0.720288	0.046479
1.36696	0.749009	2.107655	0.021043	0.130582
0.400322	0.597199	0.199617	0.326224	3.057497
1.046992	1.147977	0.811818	2.653621	0.572306
1.149165	0.011386	0.313108	0.353726	0.551813
0.12428	0.2782	2.530624	0.353287	0.696851
2.074256	0.007937	0.4112	1.59742	0.139829

0.392099	1.107804	2.247781	0.307224	1.486593
1.701219	0.325875	0.539727	0.379753	0.849806
0.224434	1.553748	0.718113	1.19199	1.342881
0.048064	1.824698	3.065749	0.667236	1.441571
1.022045	0.137668	0.178684	0.064878	2.802092
0.604435	0.425377	0.987098	0.436582	0.505152
3.576425	0.308109	0.49484	0.1259	0.191531
2.692184	0.771681	0.371021	0.122222	1.880483
1.614054	1.338874	1.620623	0.401148	2.134135
1.429995	0.096377	0.345196	0.162366	1.406259
1.070883	2.077612	0.649609	0.04518	1.952897
0.021371	0.453998	4.369118	1.901445	0.327831
0.400872	1.842166	0.355855	0.285736	1.224235
0.897106	1.186114	0.169439	1.714384	0.568011
0.528333	0.202413	1.287711	0.125179	0.945367
0.920513	0.79368	0.76663	0.973867	0.820415
1.288973	0.060237	3.000044	0.032175	0.481702
0.109605	1.654016	0.982619	0.56275	1.700466
1.155187	0.722817	0.051396	0.664271	0.630232
0.578155	0.539419	0.534118	2.057596	1.525887
2.264051	0.555491	2.476347	0.239363	0.51711
2.07185	0.416767	0.882515	0.538518	2.604869
4.291943	0.428841	1.288611	0.754633	0.237472
1.359087	0.681121	0.234289	1.484256	0.384901
0.235063	0.165603	0.021356	0.551388	0.29025
0.144301	1.064413	1.175353	0.754111	0.027004
0.059802	1.28368	1.715012	0.989011	0.678097
0.258795	1.173518	1.137141	0.496917	0.956372
2.9542	0.88857	1.177929	0.130595	0.110161
0.204523	2.619797	0.080113	0.517011	1.459288
3.374453	3.112975	3.25491	1.282775	0.225325
3.924585	0.371989	0.008921	0.732742	2.761384

0.421207	2.129955	2.954681	0.949918	0.846363
0.309583	0.185046	1.265735	0.38211	0.62504
0.535462	1.613503	0.188712	0.424233	0.557613
0.416505	0.258731	1.658877	0.004347	0.594232
1.184293	0.020952	0.919574	1.056874	0.915278
0.751082	3.753407	0.582184	0.932324	1.40184
2.94842	0.791042	2.370507	0.37296	0.531475
0.900014	0.855607	0.661562	0.039265	0.976084
3.553019	0.996863	3.502837	1.771868	2.021649
0.664897	0.131541	0.593874	0.458191	0.519399
1.061394	1.139625	0.588278	0.00616	0.720182
0.00355	0.266202	0.988082	1.083858	0.446251
0.235712	1.914072	3.47427	0.969162	0.744546
0.279309	2.299093	0.05283	1.229073	0.239323
2.38438	1.069331	3.208228	0.25529	0.535516
0.071176	0.126251	0.8306	0.276486	3.490072
1.007534	0.136438	3.279504	0.180006	0.272997
0.029366	2.662716	0.793603	0.135242	1.132632
0.090734	1.359176	1.716074	1.177372	2.75445
3.657841	2.116441	0.058023	0.072209	0.444014
0.397388	0.257351	0.671444	0.53594	2.986329
0.658519	0.234494	5.097753	3.0982	0.150558
1.070542	0.048351	0.611334	0.180838	0.263908
2.020703	0.077147	0.499556	2.440088	0.353021
0.396532	0.068443	0.162576	2.567871	0.021787
0.691625	0.860069	0.211184	0.372005	0.459214
1.18067	1.564699	0.68928	3.528308	0.499647
1.920802	1.552614	2.418828	0.445807	0.234249
0.227212	0.93523	1.453732	0.012001	2.91199
1.619449	1.608159	2.324176	1.002299	1.095312
0.080805	0.937196	2.912455	2.639572	1.011743
0.211857	0.147708	0.190302	1.029105	0.718035

1.372726	0.209296	0.155431	0.080266	0.384136
1.389949	1.210217	0.008005	1.296866	1.481934
0.257229	0.056888	1.226955	0.424233	0.16537
1.561806	0.752028	0.459675	0.154586	0.040737
0.659582	0.520779	0.946269	0.525156	0.043926
0.328723	3.562712	0.229409	1.520866	0.269526
1.371698	0.675922	0.031669	0.497509	0.831811
2.263215	2.943534	0.971473	0.527706	0.277151

Tabela 8.10 Test slaganja I-odstojanja (za eksponencijalnu raspodelu)

Kolmogorov-Smirnov Test			
		I2_MIN12345	I2_MIN53142
Veličina uzorka		100	100
Parametri raspodele	Sredina	9.2471	9.2433
	St. devijacija	9.64595	9.63386
Kolmogorov-Smirnov test statistika		1.837	1.836
signifikantnost		.002	.002

Tabela 8.11 Rezultat generisanog 4-dim vektora sa vejbulovom raspodelom

w1	w2	w3	w4
1.635326	0.08101	0.789287	0.752
0.90995	4.349374	1.410824	2.039307
1.6908	1.374513	0.534429	0.894048
0.873388	0.461183	0.376242	0.692301
1.673969	2.565082	0.260993	0.415518
0.015881	1.146221	0.458098	1.212595
3.757633	0.662618	1.299342	0.185493
0.629499	0.943091	0.42006	3.020043
1.688921	1.100545	0.100442	2.982729
1.670728	0.5269	2.288017	1.289713
2.200557	2.191544	0.082398	1.098192
0.390956	0.185885	0.610404	1.754158
0.109382	0.874719	0.46829	0.442436
0.699049	0.47465	3.309192	0.762574
0.826406	0.706912	1.220657	0.513388
0.165975	0.93844	1.622836	0.953026
0.077326	1.376858	0.19662	3.642863
0.411665	1.157407	0.500651	0.834518
0.003781	4.194297	0.894234	1.238114
0.445455	0.642635	1.064503	4.066254
0.071221	0.223451	2.093852	2.635067
0.072146	1.615786	1.011213	0.332103
0.792049	1.010075	2.119089	0.412339
2.71603	0.244229	0.213681	1.200665
1.993959	0.567731	0.855086	1.543192
0.133235	4.196681	0.374934	0.160618
3.447196	0.329928	2.554404	1.602079
0.143318	0.532887	0.607026	0.784936
0.085124	0.192362	0.033763	0.970624

1.153476	0.613297	0.15472	0.286656
0.820438	0.707184	5.408997	1.075734
0.792755	0.869605	0.327141	0.340955
4.265307	0.090624	3.552126	0.398779
0.193121	0.208712	0.068165	0.379499
0.707846	0.08232	0.685958	1.706839
1.346796	0.141928	0.247919	1.825497
0.742376	0.570101	0.440065	0.25654
0.196646	0.585439	0.045383	0.847367
0.113747	1.253292	2.560284	1.345736
1.109695	1.51672	0.459664	0.360865
1.269498	0.562067	1.047085	0.043082
1.087534	0.156258	1.833549	0.022983
0.7919	0.345137	0.269014	0.633765
0.831778	2.827693	0.485896	0.586306
0.433658	0.12286	2.562869	3.434165
0.486099	2.926915	1.172781	0.693357
0.233921	1.067174	1.790195	0.379547
0.784429	0.060645	1.367413	0.546871
1.747926	0.300183	0.866608	0.284399
4.109177	4.375735	1.5496	0.798278
1.63718	1.064467	2.861389	0.492864
3.022601	0.588337	0.599961	3.729081
1.270294	0.117541	0.329886	0.167947
1.428689	1.755156	0.2587	0.028191
2.278189	0.346812	0.143745	0.530656
0.227162	1.865721	0.377725	0.178341
0.980205	1.220644	1.067328	0.592211
0.444043	0.647656	3.181339	0.377599
0.655503	2.257728	0.136689	0.551855
1.000656	0.238262	2.163175	1.534425
2.830576	1.029799	0.000631	0.025592

0.255214	2.162063	0.869209	0.391937
1.877888	0.594193	0.613626	0.635918
1.586089	0.155062	0.504509	2.970179
4.54129	0.024315	0.526646	0.937538
0.47388	0.772972	0.085323	0.221111
2.315136	0.047619	0.44246	0.273154
2.751225	2.867846	1.954323	0.928995
0.110994	0.843128	0.780425	1.146096
0.147632	0.344609	0.170083	0.95503
1.108386	0.45511	1.534177	1.28919
1.408799	0.323879	5.035596	1.522715
0.240051	0.666975	0.528038	0.974632
0.590636	0.135562	2.350236	3.899539
2.278733	0.822064	0.288055	0.99676
1.628709	0.133401	0.616455	0.052408
1.295119	0.738442	0.238156	0.429133
0.005076	0.391931	0.408868	0.792155
0.977479	0.601031	1.157296	0.011439
0.698461	1.435399	1.084582	0.198245
0.032414	1.462342	1.546454	0.882863
0.015038	0.599097	1.163582	0.777341
0.101899	0.367828	0.123765	2.095102
1.136346	4.207735	0.123289	0.029691
1.18436	0.155401	1.222784	0.134763
0.07581	1.259589	1.073179	0.262713
0.835173	0.618354	0.755106	1.053059
0.821685	2.445029	0.938969	0.231703
4.354292	0.208238	0.75493	0.46002
0.180303	1.452834	1.680246	0.025728
0.013297	3.10741	0.177361	2.39921
0.255568	2.841371	0.728252	0.907243
0.280472	0.331846	0.252766	1.5703

0.32666	0.038102	0.935246	0.848668
1.009222	0.375471	1.810003	2.032416
3.134053	0.791337	0.181517	0.411919
0.066412	1.810437	0.615635	0.496979
0.63101	0.771386	2.710144	0.036849
0.305055	3.103408	0.105243	0.36512
0.415249	3.506528	0.581401	0.259087

Tabela 8.12 Test slaganja I-odstojanja (za vejbulovu raspodelu)

Kolmogorov-Smirnov Test		
		I2_MIN1234
Veličina uzorka		100
Parametri raspodele	Sredina	7.7166
	St. devijacija	7.55359
Kolmogorov-Smirnov test statistika		1.983
signifikantnost		.001

Tabela 8.13 Rezultat generisanog 5-dim vektora sa vejbulovom raspodelom

w1	w2	w3	w4	w5
1.635326	0.08101	0.789287	0.752	4.490906
0.90995	4.349374	1.410824	2.039307	0.193738
1.6908	1.374513	0.534429	0.894048	3.416848
0.873388	0.461183	0.376242	0.692301	2.372409
1.673969	2.565082	0.260993	0.415518	0.117636
0.015881	1.146221	0.458098	1.212595	0.662742
3.757633	0.662618	1.299342	0.185493	0.038489
0.629499	0.943091	0.42006	3.020043	0.980841
1.688921	1.100545	0.100442	2.982729	1.261489
1.670728	0.5269	2.288017	1.289713	1.194962
2.200557	2.191544	0.082398	1.098192	1.200384
0.390956	0.185885	0.610404	1.754158	3.052836
0.109382	0.874719	0.46829	0.442436	0.01675
0.699049	0.47465	3.309192	0.762574	0.332797
0.826406	0.706912	1.220657	0.513388	0.121945
0.165975	0.93844	1.622836	0.953026	0.712026
0.077326	1.376858	0.19662	3.642863	0.992384
0.411665	1.157407	0.500651	0.834518	0.368542
0.003781	4.194297	0.894234	1.238114	2.970847
0.445455	0.642635	1.064503	4.066254	0.958587
0.071221	0.223451	2.093852	2.635067	0.565509
0.072146	1.615786	1.011213	0.332103	1.958186
0.792049	1.010075	2.119089	0.412339	1.065108
2.71603	0.244229	0.213681	1.200665	1.268663
1.993959	0.567731	0.855086	1.543192	1.377246
0.133235	4.196681	0.374934	0.160618	0.307396
3.447196	0.329928	2.554404	1.602079	0.600243
0.143318	0.532887	0.607026	0.784936	0.870474
0.085124	0.192362	0.033763	0.970624	1.038926
1.153476	0.613297	0.15472	0.286656	0.550123

0.820438	0.707184	5.408997	1.075734	1.171746
0.792755	0.869605	0.327141	0.340955	0.155153
4.265307	0.090624	3.552126	0.398779	2.057037
0.193121	0.208712	0.068165	0.379499	1.878918
0.707846	0.08232	0.685958	1.706839	0.976963
1.346796	0.141928	0.247919	1.825497	0.028199
0.742376	0.570101	0.440065	0.25654	0.033008
0.196646	0.585439	0.045383	0.847367	0.179632
0.113747	1.253292	2.560284	1.345736	0.581997
1.109695	1.51672	0.459664	0.360865	0.084584
1.269498	0.562067	1.047085	0.043082	0.419984
1.087534	0.156258	1.833549	0.022983	1.190697
0.7919	0.345137	0.269014	0.633765	0.734241
0.831778	2.827693	0.485896	0.586306	0.851039
0.433658	0.12286	2.562869	3.434165	2.732897
0.486099	2.926915	1.172781	0.693357	0.34849
0.233921	1.067174	1.790195	0.379547	0.926678
0.784429	0.060645	1.367413	0.546871	1.869497
1.747926	0.300183	0.866608	0.284399	0.702118
4.109177	4.375735	1.5496	0.798278	0.088514
1.63718	1.064467	2.861389	0.492864	0.43335
3.022601	0.588337	0.599961	3.729081	0.713934
1.270294	0.117541	0.329886	0.167947	0.401918
1.428689	1.755156	0.2587	0.028191	0.714077
2.278189	0.346812	0.143745	0.530656	0.369355
0.227162	1.865721	0.377725	0.178341	1.012042
0.980205	1.220644	1.067328	0.592211	0.3925
0.444043	0.647656	3.181339	0.377599	0.84897
0.655503	2.257728	0.136689	0.551855	0.381732
1.000656	0.238262	2.163175	1.534425	0.255024
2.830576	1.029799	0.000631	0.025592	0.494191
0.255214	2.162063	0.869209	0.391937	1.935286

1.877888	0.594193	0.613626	0.635918	0.607764
1.586089	0.155062	0.504509	2.970179	0.026319
4.54129	0.024315	0.526646	0.937538	0.078352
0.47388	0.772972	0.085323	0.221111	1.673193
2.315136	0.047619	0.44246	0.273154	1.169652
2.751225	2.867846	1.954323	0.928995	3.064826
0.110994	0.843128	0.780425	1.146096	3.098917
0.147632	0.344609	0.170083	0.95503	0.160505
1.108386	0.45511	1.534177	1.28919	2.134713
1.408799	0.323879	5.035596	1.522715	0.354362
0.240051	0.666975	0.528038	0.974632	2.214021
0.590636	0.135562	2.350236	3.899539	2.569718
2.278733	0.822064	0.288055	0.99676	0.309524
1.628709	0.133401	0.616455	0.052408	0.653488
1.295119	0.738442	0.238156	0.429133	0.549164
0.005076	0.391931	0.408868	0.792155	3.454142
0.977479	0.601031	1.157296	0.011439	1.150949
0.698461	1.435399	1.084582	0.198245	2.906401
0.032414	1.462342	1.546454	0.882863	0.514744
0.015038	0.599097	1.163582	0.777341	2.727176
0.101899	0.367828	0.123765	2.095102	0.329051
1.136346	4.207735	0.123289	0.029691	0.673646
1.18436	0.155401	1.222784	0.134763	3.494987
0.07581	1.259589	1.073179	0.262713	0.446913
0.835173	0.618354	0.755106	1.053059	0.127025
0.821685	2.445029	0.938969	0.231703	1.579458
4.354292	0.208238	0.75493	0.46002	0.520217
0.180303	1.452834	1.680246	0.025728	0.075584
0.013297	3.10741	0.177361	2.39921	0.960402
0.255568	2.841371	0.728252	0.907243	0.246976
0.280472	0.331846	0.252766	1.5703	0.434181
0.32666	0.038102	0.935246	0.848668	2.176048

1.009222	0.375471	1.810003	2.032416	0.15391
3.134053	0.791337	0.181517	0.411919	0.794436
0.066412	1.810437	0.615635	0.496979	2.885806
0.63101	0.771386	2.710144	0.036849	0.533696
0.305055	3.103408	0.105243	0.36512	0.163684
0.415249	3.506528	0.581401	0.259087	0.017256

Tabela 8.14 Test slaganja I-odstojanja (za vejbulovu raspodelu)

One-Sample Kolmogorov-Smirnov Test		
		I2_MIN12345
Veličina uzorka		100
Parametri raspodele	Sredina	9.8774
	St. devijacija	7.82704
Kolmogorov-Smirnov test statistika		1.259
signifikantnost		.084

Tabela 8.15 Rezultat generisanog 4-dim vektora sa binomnom raspodelom

b1	**b2**	**b3**	**b4**
0.545	0.58	0.475	0.54
0.48	0.495	0.435	0.515
0.47	0.505	0.56	0.485
0.46	0.53	0.54	0.55
0.52	0.43	0.5	0.495
0.505	0.485	0.48	0.465
0.47	0.415	0.425	0.53
0.505	0.505	0.555	0.465
0.49	0.57	0.515	0.5
0.515	0.525	0.455	0.495
0.49	0.5	0.515	0.52
0.49	0.54	0.595	0.45
0.465	0.5	0.49	0.475
0.49	0.46	0.505	0.575
0.565	0.54	0.56	0.48
0.53	0.58	0.545	0.49
0.48	0.495	0.53	0.485
0.46	0.48	0.435	0.475
0.46	0.51	0.515	0.47
0.57	0.465	0.545	0.515
0.54	0.43	0.54	0.48
0.49	0.525	0.485	0.535
0.49	0.535	0.54	0.495
0.5	0.49	0.515	0.485
0.51	0.465	0.44	0.485
0.485	0.51	0.505	0.51
0.49	0.485	0.45	0.48
0.505	0.515	0.505	0.55
0.54	0.59	0.475	0.455
0.52	0.405	0.585	0.535

0.52	0.415	0.51	0.495
0.48	0.495	0.555	0.485
0.495	0.495	0.51	0.45
0.495	0.455	0.5	0.525
0.51	0.485	0.54	0.495
0.47	0.515	0.485	0.515
0.485	0.54	0.475	0.53
0.52	0.455	0.445	0.505
0.505	0.52	0.48	0.51
0.515	0.485	0.51	0.525
0.475	0.55	0.415	0.47
0.57	0.46	0.505	0.47
0.465	0.48	0.555	0.485
0.47	0.505	0.505	0.53
0.53	0.47	0.495	0.5
0.495	0.49	0.56	0.5
0.535	0.505	0.51	0.43
0.545	0.535	0.505	0.505
0.58	0.52	0.475	0.495
0.505	0.515	0.505	0.505
0.475	0.465	0.535	0.495
0.545	0.58	0.445	0.51
0.47	0.455	0.485	0.5
0.505	0.495	0.51	0.535
0.52	0.51	0.48	0.53
0.46	0.47	0.51	0.52
0.5	0.455	0.545	0.53
0.525	0.535	0.485	0.515
0.5	0.56	0.55	0.5
0.495	0.485	0.545	0.485
0.47	0.515	0.525	0.465
0.47	0.52	0.525	0.52

0.465	0.47	0.505	0.52
0.51	0.485	0.515	0.5
0.575	0.495	0.535	0.48
0.49	0.515	0.55	0.535
0.475	0.56	0.535	0.465
0.48	0.525	0.535	0.49
0.485	0.51	0.49	0.49
0.505	0.555	0.505	0.505
0.455	0.515	0.555	0.5
0.525	0.51	0.425	0.465
0.46	0.53	0.505	0.515
0.535	0.56	0.5	0.445
0.48	0.5	0.47	0.53
0.475	0.515	0.445	0.525
0.56	0.48	0.51	0.51
0.495	0.46	0.495	0.505
0.52	0.5	0.48	0.52
0.52	0.465	0.54	0.485
0.5	0.505	0.465	0.51
0.545	0.585	0.48	0.53
0.565	0.485	0.505	0.515
0.52	0.515	0.55	0.515
0.525	0.445	0.455	0.52
0.525	0.5	0.53	0.515
0.455	0.425	0.525	0.475
0.475	0.46	0.475	0.485
0.46	0.45	0.52	0.48
0.48	0.495	0.465	0.45
0.5	0.515	0.535	0.46
0.42	0.455	0.555	0.525
0.5	0.475	0.48	0.6
0.54	0.51	0.5	0.53

0.48	0.505	0.51	0.48
0.51	0.525	0.54	0.535
0.48	0.525	0.51	0.545
0.54	0.48	0.52	0.5
0.5	0.52	0.53	0.455
0.485	0.52	0.495	0.555

Tabela 8.16 Test slaganja I-odstojanja (za binomnu raspodelu)

One-Sample Kolmogorov-Smirnov Test		
		I2_MIN1234
Veličina uzorka		100
Parametri raspodele	Sredina	28.8668
	St. devijacija	10.06584
Kolmogorov-Smirnov test statistika		.539
signifikantnost		.933

Tabela 8.17 Rezultat generisanog 5-dim vektora sa binomnom raspodelom

b1	b2	b3	b4	b5
0.545	0.58	0.475	0.54	0.465
0.48	0.495	0.435	0.515	0.475
0.47	0.505	0.56	0.485	0.475
0.46	0.53	0.54	0.55	0.56
0.52	0.43	0.5	0.495	0.47
0.505	0.485	0.48	0.465	0.525
0.47	0.415	0.425	0.53	0.455
0.505	0.505	0.555	0.465	0.555
0.49	0.57	0.515	0.5	0.48
0.515	0.525	0.455	0.495	0.495
0.49	0.5	0.515	0.52	0.465
0.49	0.54	0.595	0.45	0.485
0.465	0.5	0.49	0.475	0.535
0.49	0.46	0.505	0.575	0.52
0.565	0.54	0.56	0.48	0.535
0.53	0.58	0.545	0.49	0.555
0.48	0.495	0.53	0.485	0.46
0.46	0.48	0.435	0.475	0.52
0.46	0.51	0.515	0.47	0.48
0.57	0.465	0.545	0.515	0.47
0.54	0.43	0.54	0.48	0.455
0.49	0.525	0.485	0.535	0.495
0.49	0.535	0.54	0.495	0.53
0.5	0.49	0.515	0.485	0.56
0.51	0.465	0.44	0.485	0.525
0.485	0.51	0.505	0.51	0.48
0.49	0.485	0.45	0.48	0.475
0.505	0.515	0.505	0.55	0.5
0.54	0.59	0.475	0.455	0.485
0.52	0.405	0.585	0.535	0.55

0.52	0.415	0.51	0.495	0.45
0.48	0.495	0.555	0.485	0.48
0.495	0.495	0.51	0.45	0.435
0.495	0.455	0.5	0.525	0.475
0.51	0.485	0.54	0.495	0.43
0.47	0.515	0.485	0.515	0.525
0.485	0.54	0.475	0.53	0.545
0.52	0.455	0.445	0.505	0.47
0.505	0.52	0.48	0.51	0.5
0.515	0.485	0.51	0.525	0.555
0.475	0.55	0.415	0.47	0.555
0.57	0.46	0.505	0.47	0.475
0.465	0.48	0.555	0.485	0.52
0.47	0.505	0.505	0.53	0.48
0.53	0.47	0.495	0.5	0.455
0.495	0.49	0.56	0.5	0.5
0.535	0.505	0.51	0.43	0.505
0.545	0.535	0.505	0.505	0.535
0.58	0.52	0.475	0.495	0.53
0.505	0.515	0.505	0.505	0.515
0.475	0.465	0.535	0.495	0.53
0.545	0.58	0.445	0.51	0.54
0.47	0.455	0.485	0.5	0.525
0.505	0.495	0.51	0.535	0.515
0.52	0.51	0.48	0.53	0.525
0.46	0.47	0.51	0.52	0.55
0.5	0.455	0.545	0.53	0.435
0.525	0.535	0.485	0.515	0.515
0.5	0.56	0.55	0.5	0.5
0.495	0.485	0.545	0.485	0.54
0.47	0.515	0.525	0.465	0.545
0.47	0.52	0.525	0.52	0.485

0.465	0.47	0.505	0.52	0.535
0.51	0.485	0.515	0.5	0.535
0.575	0.495	0.535	0.48	0.47
0.49	0.515	0.55	0.535	0.48
0.475	0.56	0.535	0.465	0.465
0.48	0.525	0.535	0.49	0.445
0.485	0.51	0.49	0.49	0.505
0.505	0.555	0.505	0.505	0.48
0.455	0.515	0.555	0.5	0.515
0.525	0.51	0.425	0.465	0.46
0.46	0.53	0.505	0.515	0.5
0.535	0.56	0.5	0.445	0.555
0.48	0.5	0.47	0.53	0.585
0.475	0.515	0.445	0.525	0.515
0.56	0.48	0.51	0.51	0.495
0.495	0.46	0.495	0.505	0.49
0.52	0.5	0.48	0.52	0.535
0.52	0.465	0.54	0.485	0.51
0.5	0.505	0.465	0.51	0.495
0.545	0.585	0.48	0.53	0.47
0.565	0.485	0.505	0.515	0.46
0.52	0.515	0.55	0.515	0.515
0.525	0.445	0.455	0.52	0.475
0.525	0.5	0.53	0.515	0.515
0.455	0.425	0.525	0.475	0.475
0.475	0.46	0.475	0.485	0.52
0.46	0.45	0.52	0.48	0.485
0.48	0.495	0.465	0.45	0.47
0.5	0.515	0.535	0.46	0.51
0.42	0.455	0.555	0.525	0.525
0.5	0.475	0.48	0.6	0.5
0.54	0.51	0.5	0.53	0.45

0.48	0.505	0.51	0.48	0.49
0.51	0.525	0.54	0.535	0.525
0.48	0.525	0.51	0.545	0.495
0.54	0.48	0.52	0.5	0.475
0.5	0.52	0.53	0.455	0.475
0.485	0.52	0.495	0.555	0.485

Tabela 8.18 Test slaganja I-odstojanja (za binomnu raspodelu)

Kolmogorov-Smirnov Test		
		I2_MIN21534
Veličina uzorka		100
Parametri raspodele	Sredina	33.9773
	St. devijacija	11.08650
Kolmogorov-Smirnov test statistika		.580
signifikantnost		.889

Tabela 8.19 Rezultat generisanog 4-dim vektora sa Puasonovom raspodelom

p1	p2	p3	p4
1.6	1.3	1.1	0.5
1	0.9	1.2	1.2
0.9	1.3	0.7	1.1
0.7	1.2	1.1	1
0.7	0.7	0.8	1.2
1.6	1	1	1
1.1	1.1	1.1	0.7
1.1	0.6	0.7	1.1
1	1	1.4	1
0.5	1.6	1.4	0.9
0.9	0.6	0.9	1
1.1	1.6	0.9	1.2
0.6	0.7	0.8	1.7
1.2	1.3	0.7	0.6
1.1	1	1.1	1.9
0.7	0.9	0.8	1.1
1	0.5	1.3	1.2
1.5	1.3	1	0.9
1.2	0.6	0.4	1.3
1.2	0.5	0.5	1.1
0.9	0.9	1.4	0.9
1.4	1.2	1.1	0.9
1.4	1.3	0.9	1
1.3	1.3	0.9	0.4
0.9	0.7	0.6	0.9
1	1.2	1	1.1
0.5	0.8	1	0.8
1.6	0.4	0.9	0.6
1.3	0.8	0.7	1
0.8	0.4	1	0.7

0.6	0.7	1.1	0.7
1.5	1.6	0.7	0.8
0.9	1.2	0.3	1.1
1	1.3	0.7	1
1	1.3	0.7	1
0.7	0.7	1.2	1.1
0.6	0.5	1.1	0.6
0.5	0.6	0.6	0.8
1.3	0.4	1.2	1.1
0.8	1.2	1.5	1.5
0.7	0.8	0.8	0.8
0.4	1.7	1.4	0.5
1.2	1	0.7	0.7
0.7	1.7	0.7	0.9
0.7	0.9	1.2	1.2
0.6	1.1	0.6	0.8
0.5	1.6	1.3	1.1
1.2	1.4	1.5	0.7
1.1	1.4	0.6	1.2
0.6	1.2	0.9	0.6
0.4	1.1	1	1.4
0.6	0.4	1	0.9
0.9	1.2	0.9	0.7
1.2	0.7	0.9	0.7
1.2	0.9	1.3	0.8
1.3	1.1	0.7	1.1
1.5	1.5	2	0.7
0.8	0.5	1.5	1.2
0.8	0.6	0.5	0.7
0.9	1	0.5	1.7
1	0.6	0.8	0.7
0.9	0.8	1.2	1

1.3	0.8	0.9	0.8
1.5	0.9	1.1	0.7
0.5	1.2	1.5	1
1	1	0.7	0.5
1.4	0.6	0.6	1.4
1.3	0.8	1.2	1.1
1.2	0.7	0.6	1.7
0.5	0.8	1.3	0.8
0.7	1.3	1.2	1.3
1.2	0.8	0.8	1
1.3	1.2	0.5	1.2
0.8	1	0.6	0.9
0.9	1.4	1.2	0.9
0.9	0.7	0.8	1.1
1	1.4	0.5	1
1	1.5	0.4	0.9
1.3	1.3	1.4	0.7
1.7	0.7	0.9	0.7
0.9	1.2	1.5	1.3
0.5	1.2	0.7	1.2
0.6	0.5	1.2	0.9
0.7	1.7	1	0.7
0.6	0.7	1.4	0.8
0.8	1.1	1.1	1.3
1.4	0.7	0.9	0.7
0.8	0.9	1	0.5
0.8	1	1	0.9
0.9	1.3	0.3	0.8
1.3	0.7	1.1	0.7
1	1.7	1.3	1.3
1.1	1	0.8	0.9
0.8	1.1	0.9	1

0.6	0.7	0.6	1.3
1.9	1	1.2	1.3
1.5	1.1	0.9	1.3
1.5	0.8	0.9	1.1
0.9	1	1	0.6
0.5	1	1	1.2

Tabela 8.20 Test slaganja I-odstojanja (za Puasonovu raspodelu)

Kolmogorov-Smirnov Test			
		I2_MIN1234	I2_MIN3241
Veličina uzorka		100	100
Parametri raspodele	Sredina	18.2944	18.3171
	St. devijacija	8.49862	8.51519
Kolmogorov-Smirnov test statistika		.883	.922
signifikantnost		.417	.363

Tabela 8.21 Rezultat generisanog 5-dim vektora sa Puasonovom raspodelom

p1	p2	p3	p4	p5
1.6	1.3	1.1	0.5	1.1
1	0.9	1.2	1.2	1.4
0.9	1.3	0.7	1.1	0.7
0.7	1.2	1.1	1	1.1
0.7	0.7	0.8	1.2	0.7
1.6	1	1	1	0.8
1.1	1.1	1.1	0.7	0.6
1.1	0.6	0.7	1.1	0.8
1	1	1.4	1	0.7
0.5	1.6	1.4	0.9	1.6
0.9	0.6	0.9	1	0.6
1.1	1.6	0.9	1.2	0.9
0.6	0.7	0.8	1.7	0.8
1.2	1.3	0.7	0.6	0.7
1.1	1	1.1	1.9	1.3
0.7	0.9	0.8	1.1	1
1	0.5	1.3	1.2	1.3
1.5	1.3	1	0.9	0.9
1.2	0.6	0.4	1.3	0.8
1.2	0.5	0.5	1.1	0.6
0.9	0.9	1.4	0.9	0.6
1.4	1.2	1.1	0.9	1.5
1.4	1.3	0.9	1	0.5
1.3	1.3	0.9	0.4	0.9
0.9	0.7	0.6	0.9	0.9
1	1.2	1	1.1	1.3
0.5	0.8	1	0.8	0.6
1.6	0.4	0.9	0.6	1.1
1.3	0.8	0.7	1	1.1
0.8	0.4	1	0.7	1.1

0.6	0.7	1.1	0.7	0.9
1.5	1.6	0.7	0.8	0.7
0.9	1.2	0.3	1.1	1
1	1.3	0.7	1	0.7
1	1.3	0.7	1	1.7
0.7	0.7	1.2	1.1	0.7
0.6	0.5	1.1	0.6	0.8
0.5	0.6	0.6	0.8	1.3
1.3	0.4	1.2	1.1	0.9
0.8	1.2	1.5	1.5	0.8
0.7	0.8	0.8	0.8	1.8
0.4	1.7	1.4	0.5	1
1.2	1	0.7	0.7	0.6
0.7	1.7	0.7	0.9	0.8
0.7	0.9	1.2	1.2	1.3
0.6	1.1	0.6	0.8	0.9
0.5	1.6	1.3	1.1	1
1.2	1.4	1.5	0.7	0.7
1.1	1.4	0.6	1.2	1.6
0.6	1.2	0.9	0.6	0.8
0.4	1.1	1	1.4	1.6
0.6	0.4	1	0.9	0.7
0.9	1.2	0.9	0.7	0.7
1.2	0.7	0.9	0.7	1
1.2	0.9	1.3	0.8	0.9
1.3	1.1	0.7	1.1	1.2
1.5	1.5	2	0.7	1.2
0.8	0.5	1.5	1.2	0.6
0.8	0.6	0.5	0.7	1.1
0.9	1	0.5	1.7	0.7
1	0.6	0.8	0.7	0.4
0.9	0.8	1.2	1	1

1.3	0.8	0.9	0.8	1.2
1.5	0.9	1.1	0.7	0.6
0.5	1.2	1.5	1	0.6
1	1	0.7	0.5	0.8
1.4	0.6	0.6	1.4	0.7
1.3	0.8	1.2	1.1	0.8
1.2	0.7	0.6	1.7	0.8
0.5	0.8	1.3	0.8	1.4
0.7	1.3	1.2	1.3	1.4
1.2	0.8	0.8	1	1.3
1.3	1.2	0.5	1.2	1.1
0.8	1	0.6	0.9	0.8
0.9	1.4	1.2	0.9	1.3
0.9	0.7	0.8	1.1	1.2
1	1.4	0.5	1	1
1	1.5	0.4	0.9	1.7
1.3	1.3	1.4	0.7	1.1
1.7	0.7	0.9	0.7	0.7
0.9	1.2	1.5	1.3	0.9
0.5	1.2	0.7	1.2	0.7
0.6	0.5	1.2	0.9	0.8
0.7	1.7	1	0.7	0.7
0.6	0.7	1.4	0.8	1
0.8	1.1	1.1	1.3	1.2
1.4	0.7	0.9	0.7	0.9
0.8	0.9	1	0.5	0.9
0.8	1	1	0.9	1.2
0.9	1.3	0.3	0.8	0.6
1.3	0.7	1.1	0.7	0.8
1	1.7	1.3	1.3	1.1
1.1	1	0.8	0.9	1.7
0.8	1.1	0.9	1	1

0.6	0.7	0.6	1.3	1.5
1.9	1	1.2	1.3	1.8
1.5	1.1	0.9	1.3	0.9
1.5	0.8	0.9	1.1	1
0.9	1	1	0.6	0.6
0.5	1	1	1.2	0.8

Tabela 8.22 Test slaganja I-odstojanja (za Puasonovu raspodelu)

One-Sample Kolmogorov-Smirnov Test			
		I2_MIN12345	I2_MIN52341
Veličina uzorka		100	100
Parametri raspodele	Sredina	22.3326	22.3231
	St. devijacija	10.01265	10.04293
Kolmogorov-Smirnov test statistika		.663	.667
signifikantnost		.772	.766

Prilog 2.

Program KLASTERING napisan u Javi

Main.java class

```
package clustering;

public class Main {

    public static void main(String[] args) {
        frmStart fs = new frmStart();
        fs.setVisible(true);
    }
}
```

frmStart.java class

```
package clustering;

import java.awt.event.ActionEvent;
import java.awt.event.ActionListener;
import java.util.ArrayList;
import javax.swing.GroupLayout;
import javax.swing.JButton;
import javax.swing.JFrame;
import javax.swing.JLabel;
import javax.swing.JOptionPane;
import javax.swing.JScrollPane;
import javax.swing.JTextArea;
import javax.swing.JTextField;
import javax.swing.LayoutStyle;

public class frmStart extends JFrame {
    private JButton jbStart;
    private JLabel jLabel1;
    private JScrollPane jScrollPane1;
    private JTextArea jtOutput;
    private JTextField jtInputFile;

    public frmStart() {
        initComponents();
    }

    private void initComponents() {
        this.jtInputFile = new JTextField();
        this.jLabel1 = new JLabel();
        this.jbStart = new JButton();
        this.jScrollPane1 = new JScrollPane();
        this.jtOutput = new JTextArea();

        setDefaultCloseOperation(3);
        setTitle("Clustering by Nebojsa Pavicic");

        this.jLabel1.setText("Unesi naziv fajla:");

        this.jbStart.setText("Pokreni");
        this.jbStart.addActionListener(new ActionListener() {
```

```
            public void actionPerformed(ActionEvent evt) {
                frmStart.this.jButton1ActionPerformed(evt);
            }
        });
        this.jtOutput.setColumns(20);
        this.jtOutput.setRows(5);
        this.jScrollPane1.setViewportView(this.jtOutput);

        GroupLayout layout = new GroupLayout(getContentPane());
        getContentPane().setLayout(layout);

layout.setHorizontalGroup(layout.createParallelGroup(GroupLayout.Alignment.LEADING).a
ddGroup(layout.createSequentialGroup().addContainerGap().addGroup(layout.createParall
elGroup(GroupLayout.Alignment.LEADING).addComponent(this.jScrollPane1, -1, 440,
32767).addGroup(layout.createSequentialGroup().addComponent(this.jLabel1, -2, 96, -
2).addPreferredGap(LayoutStyle.ComponentPlacement.RELATED).addComponent(this.jtInputF
ile, -2, 143, -2)).addComponent(this.jbStart)).addContainerGap()));


layout.setVerticalGroup(layout.createParallelGroup(GroupLayout.Alignment.LEADING).add
Group(layout.createSequentialGroup().addGap(18, 18,
18).addGroup(layout.createParallelGroup(GroupLayout.Alignment.BASELINE).addComponent(
this.jLabel1).addComponent(this.jtInputFile, -2, -1, -2)).addGap(18, 18,
18).addComponent(this.jbStart).addGap(36, 36, 36).addComponent(this.jScrollPane1, -1,
292, 32767).addContainerGap()));

        pack();
    }

    private void jButton1ActionPerformed(ActionEvent evt) {
        if (this.jtInputFile.getText().trim().equals("")) {
            JOptionPane.showMessageDialog(this, "Morate da unesete fajl iz koga se
ucitavaju koordinate (fajl mora da bude u istom direktorijumu!)");
            return;
        }

        jtOutput.setText("");

        ArrayList lista = null;
        try {
            lista = Util.loadFile(this.jtInputFile.getText().trim());
        } catch (Exception e) {
```

```
            JOptionPane.showMessageDialog(this, "Morate da unesete fajl iz koga se
ucitavaju koordinate (fajl mora da bude u istom direktorijumu!)\n" + e.getMessage());
            return;
        }

        ArrayList sredjenaLista = Util.clustering(lista);

        for (int i = 0; i < sredjenaLista.size(); i++) {
            this.jtOutput.setText(this.jtOutput.getText() + "Lista " + (i + 1) + "
");
            for (int j = 0; j < ((ArrayList) sredjenaLista.get(i)).size(); j++) {
                if (j == 0) {
                    this.jtOutput.setText(this.jtOutput.getText() + " [ grupa " +
((Pozicija) ((ArrayList) sredjenaLista.get(i)).get(j)).getGroup() + " ] ");
                }

                this.jtOutput.setText(this.jtOutput.getText() + "(" + ((Pozicija)
((ArrayList) sredjenaLista.get(i)).get(j)).getX() + "," + ((Pozicija) ((ArrayList)
sredjenaLista.get(i)).get(j)).getY() + "), ");
            }
            this.jtOutput.setText(this.jtOutput.getText() + "\n");
        }
    }
}
```

Pozicija.java class

```java
package clustering;

public class Pozicija {

    private double x;
    private double y;
    private int group;

    public double getX() {
        return this.x;
    }

    public void setX(double x) {
        this.x = x;
    }

    public double getY() {
        return this.y;
    }

    public void setY(double y) {
        this.y = y;
    }

    public int getGroup() {
        return this.group;
    }

    public void setGroup(int group) {
        this.group = group;
    }
}
```

Util.java class

```
package clustering;

import java.io.BufferedReader;
import java.io.DataInputStream;
import java.io.FileInputStream;
import java.io.InputStreamReader;
import java.util.ArrayList;

public class Util {

    public static ArrayList<ArrayList<Pozicija>> clustering(ArrayList<Pozicija> list)
{
        ArrayList clone = (ArrayList) list.clone();

        ArrayList listaGrupa = new ArrayList();

        int count = 0;
        while (count < 3) {
            ArrayList tempList = new ArrayList();
            tempList.add(clone.get(0));
            listaGrupa.add(tempList);
            clone.remove(0);
            count++;
        }

        while (clone.size() > 0) {
            for (int i = 0; i < clone.size(); i++) {
                double minValue = -1.0D;
                int pozicijaGrupe = -1;

                for (int j = 0; j < listaGrupa.size(); j++) {
                    double rastojanje = vratiRastojanje((Pozicija) clone.get(i),
(Pozicija) ((ArrayList) listaGrupa.get(j)).get(0));

                    if (minValue == -1.0D) {
                        minValue = rastojanje;
                        pozicijaGrupe = j;
                    } else if (rastojanje < minValue) {
                        minValue = rastojanje;
                        pozicijaGrupe = j;
```

```
                    } else if (rastojanje == minValue) {
                        if(((Pozicija) clone.get(i)).getGroup() ==
((Pozicija)((ArrayList)listaGrupa.get(j)).get(0)).getGroup() ) {
                            minValue = rastojanje;
                            pozicijaGrupe = j;
                        }
                    }

                }

                if (((Pozicija) ((ArrayList)
listaGrupa.get(pozicijaGrupe)).get(0)).getGroup() == ((Pozicija)
clone.get(i)).getGroup()) {
                    ((ArrayList) listaGrupa.get(pozicijaGrupe)).add(clone.get(i));
                    clone.remove(i);
                    i--;
                }

            }

            if (clone.size() > 0) {
                ArrayList tempList = new ArrayList();
                tempList.add(clone.get(0));
                listaGrupa.add(tempList);
                clone.remove(0);
            }
        }

        return listaGrupa;
    }

    public static double vratiRastojanje(Pozicija prva, Pozicija druga) {
        double retValue = Math.sqrt((prva.getX() - druga.getX()) * (prva.getX() -
druga.getX()) + (prva.getY() - druga.getY()) * (prva.getY() - druga.getY()));
        return retValue;
    }

    public static ArrayList<Pozicija> loadFile(String fajl)
            throws Exception {
        ArrayList lista = new ArrayList();

        FileInputStream fstream = new FileInputStream(fajl);
```

```java
        DataInputStream in = new DataInputStream(fstream);
        BufferedReader br = new BufferedReader(new InputStreamReader(in));
        String strLine;
        while ((strLine = br.readLine()) != null) {
            String[] tmpNiz = strLine.split("@");
            Pozicija p = new Pozicija();
            p.setX(Double.parseDouble(tmpNiz[0]));
            p.setY(Double.parseDouble(tmpNiz[1]));
            p.setGroup(Integer.parseInt(tmpNiz[2]));

            lista.add(p);
        }

        return lista;
    }
}
```

Biografija

Aleksandar Đoković je rođen 25.03.1974. godine u Beogradu. Osnovnu školu završio je u Medveđi, a Matematičku gimnaziju u Beogradu. Posle odsluženog vojnog roka 1997. godine upisuje Fakultet organizacionih nauka, smer informacioni sistemi, koji završava 2005. godine, sa prosečnom ocenom 8.16 u toku studija i ocenom 10 na diplomskom radu. U toku studija bio je član studentske fudbalske ekipe Fon-a i aktivno se bavio sindikalnim organizovanjem sudenata u okviru Saveza studenata. Bio je predsednik Saveza sudenata Beograda 2001. godine.

U školskoj 2005/2006 godini i prvom semestru školske 2006/2007 godine bio je angažovan u Laboratoriji za statistiku na Fakultetu organizacionih nauka, kao demonstrator u nastavi na predmetima Teorija verovatnoće, Statistika i Linearni statistički modeli. Od 15.02.2007. godine je zaposlen na Fakultetu organizacionih nauka, kao saradnik u nastavi za užu naučnu oblast Računarska statistika. Osnovne studije je završio po starom nastavnom planu i programu, pa je u školskoj 2007/2008 godini upisao doktorske studije na Fakultetu organizacionih nauka – izborno područje Operaciona istraživanja. Prilikom evaluacije od strane studenata, njegov pedagoški rad je redovno ocenjivan visokom ocenom. Prosečna ocena u letnjem semestru školske 2010/2011 godine je iznosila 4.45 (na Likertovoj skali od 1 do 5).Od 2008. godine je sekretar katedre za Operaciona istraživanja i statistiku. Od februara meseca 2009. godine radi na Fakultetu organizacionih nauka u zvanju asistenta za užu naučnu oblast Računarska statistika. Učestvovao je u projektu Evropske Banke za razvoj, tokom 2009. godine, u okviru kojeg je vršio edukaciju kadrova u Republičkom zavodu za statistiku u Beogradu i Novom Sadu. Član je Statističkog društva Srbije.

Aleksandar Đoković objavio je, u saradnji sa drugim autorima, više naučnih radova u međunarodnim časopisima, kao i zbornicima sa domaćih i međunarodnih konferencija.

Kategorija M23:

1. **Đokovic, A.**, Jeremic, V. & Radojicic, Z., (2012). *Towards efficient elementary school education: a Serbian perspective.* Actual problems of economics, vol. 137, pp. 294-300, 2012 (IF 2011– 0.039. (ISSN:1993-6788)
2. Maletic, P., Kreca, M., Jeremic, V., Bulaji,c M. & **Đokovic, A.,** (2012). *The ranking of municipaities in Serbia through the development level of SME in agribusiness*. Int. J. Agricult. Stat. Sci., vol. 8, no. 1, pp. 7-13, 2012 (IF 2011–0.013) (ISSN:0973-1903).
3. Jeremic, V., Vukmirovic, D., Radojicic, Z. & **Đokovic, A.,** (2011). *Towards a framework for evaluating ICT infrastructure of countries: a Serbian perspective.* Metalurgia International, vol 16, no. 9, pp. 15-18, 2011 (IF 2011 – 0.084) (ISSN:1582-2214).

Kategorija M33:

4. Dobrota, M., Jeremic, V., Jovanovic-Milenkovic, M., & **Đokovic, A**., *Students' Satisfaction with Information System of Faculty of Organizational Sciences*, IISES and University of Economics in Prague, compact disc, , Lisbon, Portugal, 2012 (ISBN: 978-80-905241-2-5).

5. Jeremic, V., **Đokovic, A.**, Mladenovic, N. & Radojicic, Z., *New method for ranking chess Olympics teams*. 10th Balkan Conference on Operational Research - BALCOR 2011, Thessalonici, Greece, pp. 36-41, 2011 (ISBN: 978-960-87277-7-9).
6. Maletic, P., Kreca, M., Jeremic, V. & **Đokovic, A**., *Towards uniformly developed municipalities: a Serbian perspective*. 14th Toulon-Verona Conference Excellence in Services, Alicante, Spain, 2011 (ISBN: 978-88904327-1-2).
7. **Đokovic, A.**, Radojicic, Z. & Vukovic, N., *Vector correlation coefficient as an evaluation measure.* BALCOR 2007, Zlatibor, pp. 381-389, 2007 (ISBN:978-86-7680-147-3).

Kategorija M51:

8. Nikodijevic, A., Anđelkovic-Labrovic, J. & **Đokovic**, **A**., *Sindrom sagorevanja među studentima Fakulteta organizacionih nauka,*, Management, vol. 64, pp. 47-53, 2012 (ISSN:0354-8635).

Kategorija M63:

9. Maletić, P., Kreća, M., Jeremić, V. & **Đoković, A.,** *Ranking of municipalities in Vojvodina through development level of SME in agribusiness.* SYM-OP-IS 2011, Zlatibor, pp. 543-546. 2011 (ISBN: 978-86-403-1168-7).
10. Dobrota, M., Milenković, N., Jeremić, V. & **Đoković, A.,** *Primena neuronskih mreža u određivanju stepena ekonomske razvijenosti zemalja*. SPIN 2011, Beograd, pp. 547-553, 2011 (ISBN 978-86-7680-244-9).
11. Milenković, N., Jeremić, V., **Đoković, A.** & Dobrota, M., *Statistički pristup merenju socio-ekonomske razvijenosti MENA zemalja*. SPIN 2011, Beograd, pp. 554-559, 2011 (ISBN 978-86-7680-244-9).
12. Jeremić, V., Bulajić, M., Marković, A. & **Đoković, A.,** *Indeks razvijenosti e-Uprave kao ključni indikator razvijenosti IKT infrastrukture*. SPIN 2011, Beograd, pp. 563-569, 2011 (ISBN 978-86-7680-244-9).
13. Dobrota, M. & **Đokovic, A.,** *Preoperativno predviđanje smrti kod pacijenata koji boluju od sepse.* SymOrg2010, 2010Zlatibor.
14. Petrovic-Đorđevic, D., **Đokovic, A.** & Savic, G., *Merenje tehnicke efikasnosti fudbalske reprezentacije Srbije u utakmicama kvalifikacija za SP 2010*. SymOrg2010, 2010 Zlatibor.
15. Radojicic, Z. & **Đokovic, A.,** *Dinamika lanca snabdevanja*. SPIN 2007, Beograd, pp.217-221, 2007 (ISBN 978-86-7680-131-2).

Prilog 3.

Izjava o autorstvu

Potpisani-a ____

broj indeksa _______________________________________

Izjavljujem

da je doktorska disertacija pod naslovom

____

____

- rezultat sopstvenog istraživačkog rada,
- da predložena disertacija u celini ni u delovima nije bila predložena za dobijanje bilo koje diplome prema studijskim programima drugih visokoškolskih ustanova,
- da su rezultati korektno navedeni i
- da nisam kršio/la autorska prava i koristio intelektualnu svojinu drugih lica.

Potpis doktoranda

U Beogradu, _____________

____

Prilog 4.

Izjava o istovetnosti štampane i elektronske verzije doktorskog rada

Ime i prezime autora ____

Broj indeksa ____

Studijski program ____

Naslov rada ____

Mentor ____

Potpisani/a ___________________________________

Izjavljujem da je štampana verzija mog doktorskog rada istovetna elektronskoj verziji koju sam predao/la za objavljivanje na portalu **Digitalnog repozitorijuma Univerziteta u Beogradu.**
Dozvoljavam da se objave moji lični podaci vezani za dobijanje akademskog zvanja doktora nauka, kao što su ime i prezime, godina i mesto rođenja i datum odbrane rada.
Ovi lični podaci mogu se objaviti na mrežnim stranicama digitalne biblioteke, u elektronskom katalogu i u publikacijama Univerziteta u Beogradu.

Potpis doktoranda

U Beogradu, ________________________

____

Prilog 5.

Izjava o korišćenju

Ovlašćujem Univerzitetsku biblioteku „Svetozar Marković“ da u Digitalni repozitorijum Univerziteta u Beogradu unese moju doktorsku disertaciju pod naslovom:

____

____

koja je moje autorsko delo.
Disertaciju sa svim prilozima predao/la sam u elektronskom formatu pogodnom za trajno arhiviranje.
Moju doktorsku disertaciju pohranjenu u Digitalni repozitorijum Univerziteta u Beogradu mogu da koriste svi koji poštuju odredbe sadržane u odabranom tipu licence Kreativne zajednice (Creative Commons) za koju sam se odlučio/la.

1. Autorstvo
2. Autorstvo - nekomercijalno
3. Autorstvo – nekomercijalno – bez prerade
4. Autorstvo – nekomercijalno – deliti pod istim uslovima
5. Autorstvo – bez prerade
6. Autorstvo – deliti pod istim uslovima

(Molimo da zaokružite samo jednu od šest ponuđenih licenci, kratak opis licenci dat je na poleđini lista).

Potpis doktoranda

U Beogradu, ________________________

____