

UNIVERZITET U BEOGRADU

FAKULTET ORGANIZACIONIH NAUKA

LINEARNI STATISTIČKI MODELI

- Seminarski rad -

Student: Colić Dragan 164/96

Beograd
Decembar 2004.

Matrica podataka

Ekipa	Ukupno					Domaćin					Gost					Gol razlika	
	O	D	N	I	B	O	D	N	I	B	O	D	N	I	B	Dato	Primljeno
Juventus	15	12	2	1	38	7	6	1	0	19	8	6	1	1	19	28	7
Milan	15	10	4	1	34	8	4	3	1	15	7	6	1	0	19	27	9
Udinese	15	8	4	3	28	7	3	2	2	11	8	5	2	1	17	22	13
Cagliari	15	6	4	5	22	8	5	2	1	17	7	1	2	4	5	24	25
Inter	15	3	12	0	21	7	2	5	0	11	8	1	7	0	10	32	24
Lecce	15	5	6	4	21	7	4	1	2	13	8	1	5	2	8	30	26
Palermo	15	5	6	4	21	7	4	2	1	14	8	1	4	3	7	13	11
Sampdoria	15	6	3	6	21	8	3	1	4	10	7	3	2	2	11	11	11
Roma	15	5	5	5	20	7	2	4	1	10	8	3	1	4	10	25	22
Chievo	15	5	5	5	20	8	3	3	2	12	7	2	2	3	8	14	18
Livorno	15	5	4	6	19	8	4	1	3	13	7	1	3	3	6	17	19
Fiorentina	15	4	7	4	19	7	3	3	1	12	8	1	4	3	7	14	17
Reggina	15	5	4	6	19	8	4	1	3	13	7	1	3	3	6	14	17
Messina	15	4	6	5	18	7	3	3	1	12	8	1	3	4	6	18	24
Lazio	15	4	5	6	17	8	2	3	3	9	7	2	2	3	8	19	20
Brescia	15	5	2	8	17	8	2	1	5	7	7	3	1	3	10	13	20
Bologna	15	3	4	8	13	7	2	2	3	8	8	1	2	5	5	13	18
Parma	15	2	6	7	12	8	2	4	2	10	7	0	2	5	2	15	24
Siena	15	2	6	7	12	7	1	4	2	7	8	1	2	5	5	10	23
Atalanta	15	0	7	8	7	8	0	5	3	5	7	0	2	5	2	11	22

O – Odigrano D – Dobijeno

N – Nerešeno I – Izgubljeno

B – Bodovi

Nakon unosa podataka u SPSS *Data View* izgleda ovako

25 : Ekipa

	Ekipa	Ukupno	Ukupno	Ukupno	Ukupno	Ukupno	Domaćin	Domaćin	Domaćin	Domaćin	Domaćin	GostOdi	GostDob	GostNer	GostIzg	GostBod	DatoGol	Primljen
1	Juventus	15	12	2	1	38	7	6	1	0	19	8	6	1	1	19	28	7
2	Milan	15	10	4	1	34	8	4	3	1	15	7	6	1	0	19	27	9
3	Udinese	15	8	4	3	28	7	3	2	2	11	8	5	2	1	17	22	13
4	Cagliari	15	6	4	5	22	8	5	2	1	17	7	1	2	4	5	24	25
5	Inter	15	3	12	0	21	7	2	5	0	11	8	1	7	0	10	32	24
6	Lecce	15	5	6	4	21	7	4	1	2	13	8	1	5	2	8	30	26
7	Palermo	15	5	6	4	21	7	4	2	1	14	8	1	4	3	7	13	11
8	Sampdoria	15	6	3	6	21	8	3	1	4	10	7	3	2	2	11	11	11
9	Roma	15	5	5	5	20	7	2	4	1	10	8	3	1	4	10	25	22
10	Chievo	15	5	5	5	20	8	3	3	2	12	7	2	2	3	8	14	18
11	Livorno	15	5	4	6	19	8	4	1	3	13	7	1	3	3	6	17	19
12	Fiorentina	15	4	7	4	19	7	3	3	1	12	8	1	4	3	7	14	17
13	Reggina	15	5	4	6	19	8	4	1	3	13	7	1	3	3	6	14	17
14	Messina	15	4	6	5	18	7	3	3	1	12	8	1	3	4	6	18	24
15	Lazio	15	4	5	6	17	8	2	3	3	9	7	2	2	3	8	19	20
16	Brescia	15	5	2	8	17	8	2	1	5	7	7	3	1	3	10	13	20
17	Bologna	15	3	4	8	13	7	2	2	3	8	8	1	2	5	5	13	18
18	Parma	15	2	6	7	12	8	2	4	2	10	7	0	2	5	2	15	24
19	Siena	15	2	6	7	12	7	1	4	2	7	8	1	2	5	5	10	23
20	Atalanta	15	0	7	8	7	8	0	5	3	5	7	0	2	5	2	11	22
21																		

SPSS Processor is ready

Variable View:

Untitled - SPSS Data Editor										
File Edit View Data Transform Analyze Graphs Utilities Window Help										
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	Ekipa	String	11	0		None	None	8	Left	Nominal
2	UkupnoOdigrano	Numeric	11	0		None	None	6	Right	Nominal
3	UkupnoDobijeno	Numeric	11	0		None	None	6	Right	Nominal
4	UkupnoNerešeno	Numeric	11	0		None	None	6	Right	Nominal
5	UkupnoIzgubljeno	Numeric	11	0		None	None	6	Right	Nominal
6	UkupnoBodovi	Numeric	11	0		None	None	6	Right	Nominal
7	DomaćinOdigrano	Numeric	11	0		None	None	6	Right	Nominal
8	DomaćinDobijeno	Numeric	11	0		None	None	6	Right	Nominal
9	DomaćinNerešeno	Numeric	11	0		None	None	6	Right	Nominal
10	DomaćinIzgubljeno	Numeric	11	0		None	None	6	Right	Nominal
11	DomaćinBodovi	Numeric	11	0		None	None	6	Right	Nominal
12	GostOdigrano	Numeric	11	0		None	None	6	Right	Nominal
13	GostDobijeno	Numeric	11	0		None	None	6	Right	Nominal
14	GostNerešeno	Numeric	11	0		None	None	6	Right	Nominal
15	GostIzgubljeno	Numeric	11	0		None	None	6	Right	Nominal
16	GostBodovi	Numeric	11	0		None	None	6	Right	Nominal
17	DatoGolova	Numeric	11	0		None	None	6	Right	Nominal
18	PrimljenoGolova	Numeric	11	0		None	None	6	Right	Nominal
19										

Data View Variable View /
 SPSS Processor is ready

Glavne komponente

Zadatak:

Na osnovu podataka iz matrice podataka, a na osnovu promenljivih dobijenih, nerešenih i izgubljenih utakmica kod kuće i na strani naći glavne komponente, odrediti varijanse glavnih komponenti i njihovo učešće u ukupnom varijabilitetu.

Rešenje:

Naredbom iz menija: *Analyze*→*Data Reduction*→*Factor...* pokrećemo analizu glavnih komponenti. U polje *Variable(s)* ubacujemo tražene promenljive. Otvaranjem dialog boksa *Descriptives* i čekiranjem opcije *Coefficients*, kazujemo programu da želimo da vidimo koeficijente korelacione matrice u izveštaju. U dialog boksu *Extraction* određujemo vrstu faktorske analize, mi ćemo ovde izabrati glavne komponente (*Principal components*). Takođe ovde biramo da li će se analiza vršiti na osnovu kovarijacione ili korelacione matrice. Korelacionu matricu uglavnom koristimo kada su varijable u našoj analizi merene na različitim skalama. Možemo podešavati i druge parametre, ali za naš slučaj to nije bitno. Pritiskom na dugme *OK* dobijamo rezultate analize.

Najpre nam je prikazana korelaciona matrica naših podataka.

Correlation Matrix							
		Domaćin Dobijeno	Domaćin Nerešeno	Domaćin Izgubljeno	Gost Dobijeno	Gost Nerešeno	Gost Izgubljeno
Correlation	DomaćinDobijeno	1,000	-,680	-,378	,418	,014	-,481
	DomaćinNerešeno	-,680	1,000	-,359	-,301	,154	,236
	DomaćinIzgubljeno	-,378	-,359	1,000	-,157	-,324	,310
	GostDobijeno	,418	-,301	-,157	1,000	-,484	-,667
	GostNerešeno	,014	,154	-,324	-,484	1,000	-,277
	GostIzgubljeno	-,481	,236	,310	-,667	-,277	1,000

Sledeća tabela prikazuje komunalitete, odnosno sumu varijanse u svakoj varijabli. Za metodu glavnih komponenti inicijalni komunaliteti uvek imaju vrednost 1,00 za korelacionu analizu. Ekstrakovani komunaliteti predstavljaju varijansu svake varijable po komponentama. Svi komunaliteti u ovoj koloni imaju dosta veliku vrednost, što pokazuje da ekstrakovane glavne komponente dobro reprezentuju početne varijable. Ako se desi da je neki komunalitet veoma nizak u ovoj koloni morali bi smo da odredimo još jednu glavnu komponentu.

Communalities

	Initial	Extraction
DomaćinDobijeno	1,000	,829
DomaćinNerešeno	1,000	,984
DomaćinIzgubljeno	1,000	,801
GostDobijeno	1,000	,963
GostNerešeno	1,000	,918
GostIzgubljeno	1,000	,747

Extraction Method: Principal Component Analysis.

U narednoj tabeli su nam prezentovane glavne komponente i njihove varijanse. Koloni *Total* prikazuje iznos varijanse originalnih promenljivih obuhvaćenom odgovarajućom glavnom komponentom. Kolona *% of Variance* prikazuje učešće varijanse glavne komponente u ukupnom varijabilitetu. Sledeća kolona prikazuje to isto samo kumulativno. Zadnje tri kolone prikazuju iste podatke kao i prethodne tri, ali samo za zadržane glavne komponente. Vidimo da zadržane glavne komponente opisuju 87,352% ukupne varijanse. Ovime smo smanjili kompleksnost podataka korišćenjem glavnih komponenti sa samo 12% gubitka originalnih informacija.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2,464	41,064	41,064	2,464	41,064	41,064
2	1,678	27,962	69,026	1,678	27,962	69,026
3	1,100	18,326	87,352	1,100	18,326	87,352
4	,698	11,640	98,992			
5	,061	1,008	100,000			
6	-1,5E-016	-2,55E-015	100,000			

Extraction Method: Principal Component Analysis.

Na kraju su nam dati i vektori glavnih komponenti,

Component Matrix(a)

	Component		
	1	2	3
DomaćinDobijeno	-,851	,071	,315
DomaćinNerešeno	,625	,501	-,586
DomaćinIzgubljeno	,295	-,793	,294
GostDobijeno	-,796	-,215	-,531
GostNerešeno	,107	,790	,531
GostIzgubljeno	,785	-,350	,086

Extraction Method: Principal Component Analysis.
a. 3 components extracted.

Iz ove tabele vidimo da je prva komponenta najviše korelisana sa varijablom *DomaćinDobijeno*, druga sa varijablom *Domaćin izgubljeno*, a treća sa *DomaćinNerešeno*. Ovo pokazuje da je za ekipe najvažnije da kod kuće ne gube utakmice ako žele postići uspeh, što je i logično.

Faktorska analiza

Zadatak:

Na osnovu podataka iz matrice podataka, a na osnovu promenljivih nerešenih i izgubljenih utakmica kod kuće i na strani kao i broja primljenih i postignutih golova naći glavne komponente, odrediti varijanse glavnih komponenti i njihovo učešće u ukupnom varijabilitetu, faktore, rotirati faktore varimax metodom, prikazati transformacionu matricu kao i grafički raspored rotiranih faktora radi interpretacije istih.

Rešenje:

Naredbom iz menija: Analyze→Data Reduction→Factor... pokrećemo faktorsku analizu. U polje Variable(s) ubacujemo tražene promenljive. U dijalog boksu *Extraction* izabiramo metod *Principal axis factoring*. U dijalog boks za rotaciju kao metod rotacije biramo varimax kriterijum i čekiramo opcije *Display rotating solution* i *Display loading plots*. Pritiskom na dugme OK dobijamo rezultate analize.

Kao i u analizi glavnih komponenata prve tri tabele imaju istu funkciju: prikazuju korelacionu matricu, komunalitete, varijanse glavnih komponenti, procentualno učešće varijanse pojedine komponente u ukupnom varijabilitetu, ukupnu opisanu varijansu pomoću svih glavnih komponenti... pa ih nećemo detaljnije objašnjavati već ćemo ih samo dati takve kakve jesu.

Correlation Matrix

		Domaćin Nerešeno	Domaćin Izgubljeno	Gost Nerešeno	Gost Izgubljeno	DatoGolova	Primljeno Golova
Correlation	DomaćinNerešeno	1,000	-,359	,154	,236	,025	,415
	DomaćinIzgubljeno	-,359	1,000	-,324	,310	-,620	,057
	GostNerešeno	,154	-,324	1,000	-,277	,285	,325
	GostIzgubljeno	,236	,310	-,277	1,000	-,650	,517
	DatoGolova	,025	-,620	,285	-,650	1,000	,042
	PrimljenoGolova	,415	,057	,325	,517	,042	1,000

Communalities

	Initial	Extraction
DomaćinNerešeno	,486	,342
DomaćinIzgubljeno	,761	,497
GostNerešeno	,639	,256
GostIzgubljeno	,884	,924
DatoGolova	,874	,700
PrimljenoGolova	,853	,614

Extraction Method: Principal Axis Factoring.

Total Variance Explained

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2,294	38,235	38,235	1,984	33,067	33,067	1,910	31,826	31,826
2	1,822	30,369	68,604	1,349	22,475	55,542	1,423	23,716	55,542
3	,889	14,813	83,416						
4	,570	9,493	92,909						
5	,387	6,445	99,355						
6	,039	,645	100,000						

Extraction Method: Principal Axis Factoring.

Faktorska matrica prikazuje faktore po komponentama

Factor Matrix(a)

	Factor	
	1	2
DomaćinNerešeno	,019	,584
DomaćinIzgubljeno	,600	-,370
GostNerešeno	-,337	,377
GostIzgubljeno	,889	,365
DatoGolova	-,806	,224
PrimljenoGolova	,262	,738

Extraction Method: Principal Axis Factoring.
a. 2 factors extracted. 23 iterations required.

Prvi faktor je najviše korelisan sa promenljivom *GostIzgubljeno*, a drugi sa promenljivom *PrimljenoGolova*. Generalno ima dosta varijabli koje imaju korelaciju sa faktorima veću od 0,2 što malo zamućuje celokupnu sliku. Tabela rotacionih faktora bi trebala ovo da popravi.

Sledeća tabela predstavlja matricu koja se koristi da se izračuna matrica rotacionih faktora iz originalne matrice faktora. U njoj manje vrednosti vandijagonalnih elemenata odgovaraju manjoj rotaciji i obrnuto.

Factor Transformation Matrix

Factor	1	2
1	-,940	,342
2	,342	,940

Extraction Method: Principal Axis Factoring.
Rotation Method: Varimax with Kaiser Normalization.

Rotacijom originalne matrice faktora dobili smo rotiranu matricu faktora koja je predstavljena u sledećoj tabeli.

Rotated Factor Matrix(a)

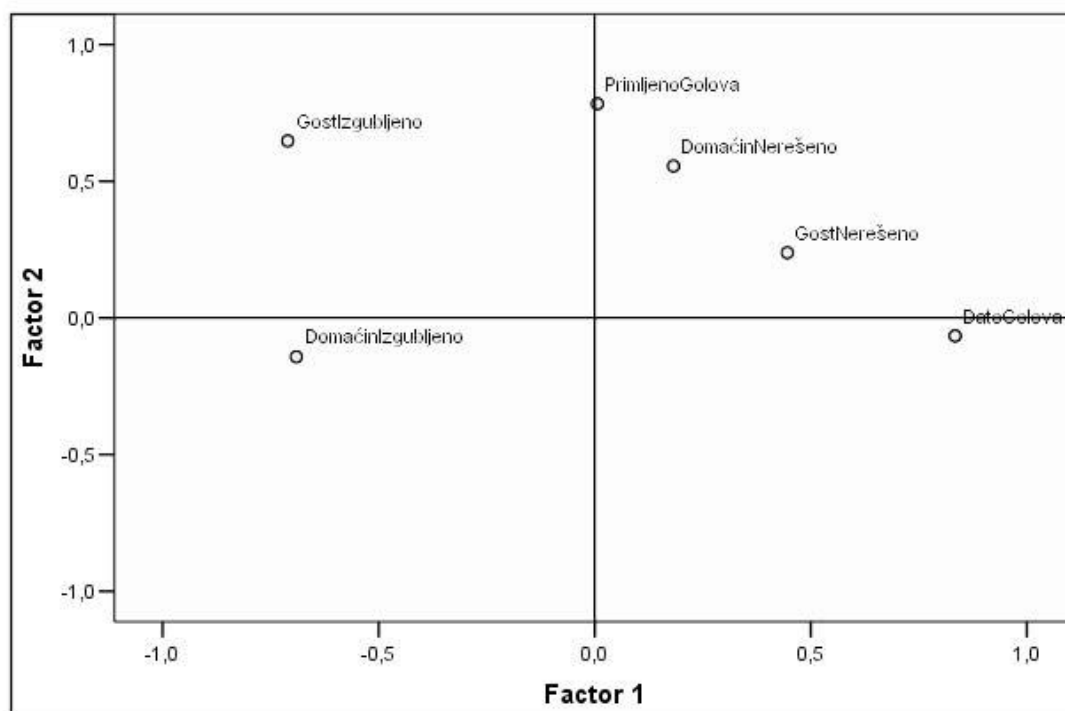
	Factor	
	1	2
DomaćinNerešeno	,182	,556
DomaćinIzgubljeno	-,691	-,142
GostNerešeno	,446	,239
GostIzgubljeno	-,711	,648
DatoGolova	,834	-,066
PrimljenoGolova	,006	,783

Extraction Method: Principal Axis Factoring.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 3 iterations.

Sada je prvi faktor najviše korelisan sa promenljivom *DatoGolova*, a drugi sa promenljivom *PrimljenoGolova*.

Na kraju dajemo grafičku interpretaciju rotiranih faktora.

Factor Plot in Rotated Factor Space



Hijerarhiska klaster analiza

Zadatak:

Na osnovu podataka iz matrice podataka, a na osnovu promenljivih *DomaćinBodovi* i *GostBodovi* formirati hijerarhisku strukturu objekata korišćenjem metoda jednostrukog povezivanja, a kao razdaljinu između objekata koristiti Euklidsko rastojanje

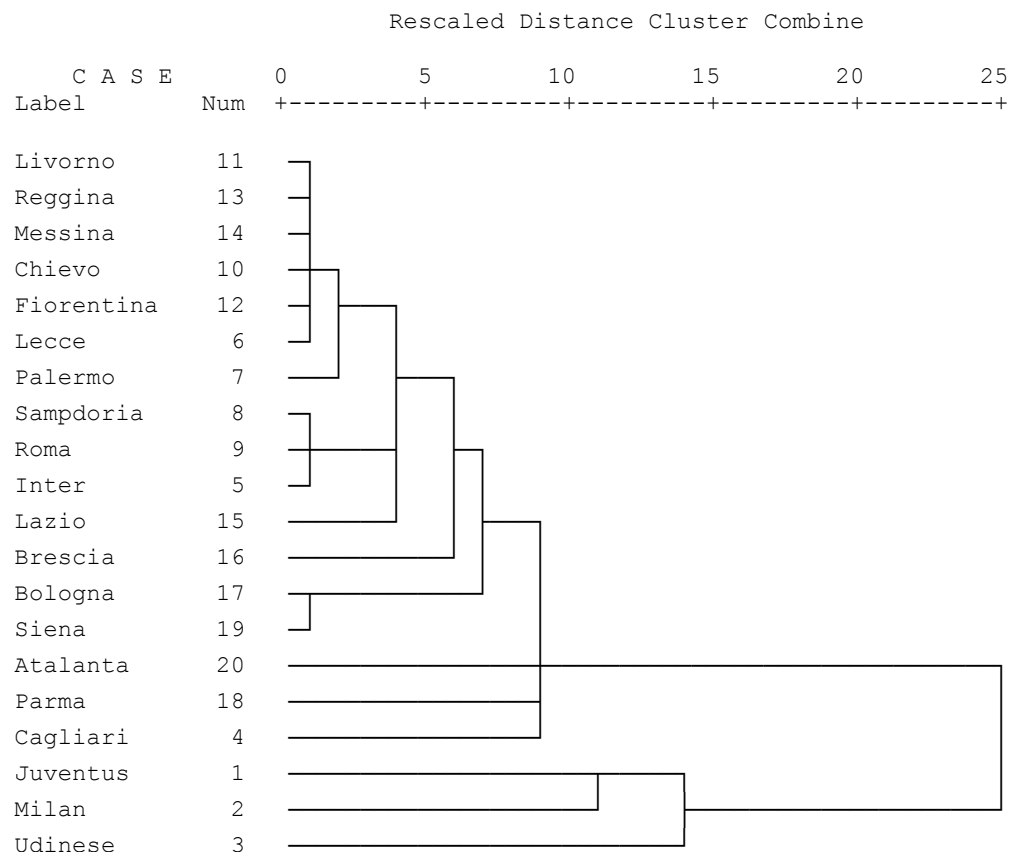
Rešenje:

Naredbom iz menija: *Analyze→Classify→Hierarchical Cluster...* pokrećemo hijerarhisku klaster analizu. U polje *Variable(s)* ubacujemo promenljive na osnovu kojih se vrši analiza. U polje *Label Cases by* ubacujemo varijablu tipa *String* preko koje identifikujemo objekte (u našem slučaju ekipe). U opcijama *Plots* izaberemo opciju *Dendrogram* kako bi na izlazu dobili i dendrogram povezivanja objekata. U opcije metoda (*Method*) biramo metod za analizu (u našem slučaju to je jednostruko povezivanje – Nearest neighbor) i kao interval za meru izabiramo Euklidsko kvadratno odstojanje. Pritiskom na dugme *OK* dobijamo rezultate analize.

Dendrogram je grafički izveštaj rešenja problema. Objekti su poređani po levoj vertikalnoj osi. Horizontalna osa pokazuje razdaljinu između objekata kada se povezuju. Deljenje dendograma kako bi dobili određen broj grupa je subjektivna procena. Generalno tražimo velike skokove između povezivanja na horizontalnoj osi. Vidimo da je najveći skok kada se povezuje objekat broj 20 (Atalanta) sa objektima 1 i 3 (Juventus i Udinese), pa presecanjem dendograma na ovom mestu možemo dobiti dve grupe. Ponavljanjem ovog postupka možemo dobiti proizvoljan broj grupa

* * * * * H I E R A R C H I C A L C L U S T E R A N A L Y S I S * * * * *

Dendrogram using Single Linkage



Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	11	13	,000	0	0	3
2	17	19	1,000	0	0	13
3	11	14	1,000	1	0	5
4	10	12	1,000	0	0	5
5	10	11	1,000	4	3	6
6	6	10	1,000	0	5	9
7	8	9	1,000	0	0	8
8	5	8	1,000	0	7	10
9	6	7	2,000	6	0	11
10	5	15	5,000	8	0	11
11	5	6	5,000	10	9	12
12	5	16	8,000	11	0	13
13	5	17	10,000	12	2	14
14	5	20	13,000	13	0	15
15	5	18	13,000	14	0	16
16	4	5	13,000	0	15	19
17	1	2	16,000	0	0	18
18	1	3	20,000	17	0	19
19	1	4	37,000	18	16	0

U prikazanoj tabeli vidimo izveštaj rešenja problema. Vidimo da se u prvoj fazi povezuju objekti 11 i 13 zato što je njihovo rastojanje najmanje. Grupa kreirana njihovim spajanjem se pojavljuje opet u fazi 3 što nam govori zadnja kolona tabele. U fazi 3 klaser kreiran u fazi 1 se sjedinjuje sa novim objektom. Ako je broj objekata posmatranja veliki lakše je pratiti kolonu koeficijenata i tražiti velike skokove između susedna dva koeficijenta nego posmatrati dendogram. Na osnovu ovih podataka možemo određivati broj grupa.

Dendogram je grafički izveštaj rešenja problema. Objekti su poređani po levoj vertikalnoj osi. Horizontalna osa pokazuje razdaljinu između objekata kada se povezuju. Deljenje dendograma kako bi dobili određen broj grupa je subjektivna procena. Generalno tražimo velike skokove između povezivanja na horizontalnoj osi. Vidimo da je najveći skok kada se povezuje objekat broj 20 (Atalanta) sa objektima 1 i 3 (Juventus i Udineze), pa presecanjem dendograma na ovom mestu možemo dobiti dve grupe. Ponavljanjem ovog postupka možemo dobiti proizvoljan broj grupa

Nehijerarhiska klaster analiza

Zadatak:

Na osnovu podataka iz matrice podataka, a na osnovu promenljivih *DomaćinBodovi* i *GostBodovi* klasifikovati ekipe u tri grupe.

Rešenje:

Naredbom iz menija: *Analyze→Classify→K-Means Cluster...* otvara se novi prozor i započinjemo proces klasifikacije. Unosimo varijable pomoću kojih se vrši klasifikacija u polje *Variable(s)* (u našem slučaju to su *DomaćinBodovi* i *GostBodovi*). U opcijama smo čekali *Initial cluster centers* i *Cluster information for each case* da bi dobili informacije o početnim klasterima i informacije o pripadnosti entiteta klasterima. Pritiskom na dugme *OK* otvara se prozor sa rezultatima datim u tabelama koje ćemo nadalje objasniti,

K-means (K-sredina) metod počinje konstrukcijom centra inicijalnih klastera. Centri inicijalnog klastera su podaci o 3 odabrana entiteta (po jedan entitet za svaku grupu). Ugrađeni algoritam bira tri entiteta po kriterijumu udaljenosti. Kao što možemo primetiti, za prvu grupu kao inicijalni klaster je izabran prvi entitet koji ima i najveće vrednosti promenljivih, a za treću grupu zadnji entitet koji ima najmanje vrednosti izabranih promenljivih. Centri inicijalnih klastera mogu se i ručno zadati.

Initial Cluster Centers

	Cluster		
	1	2	3
DomaćinBodovi	19	17	5
GostBodovi	19	5	2

Tabela istorije iteracije prikazuje poboljšanja koje vrši proces klasteringa u svakom koraku. Generalno, u prvim iteracijama promene u centrima klastera bi trebalo da se razlikuju dosta, za razliku od zadnjih iteracija gde bi te promene trebale da budu znatno manje. Na žalost u našem primeru to nismo u mogućnosti da primetimo jer imamo samo dve iteracije.

Iteration History(a)

Iteration	Change in Cluster Centers		
	1	2	3
1	4,055	5,255	4,269
2	,000	,000	,000

Sledeća tabela prikazuje pripadnost ekipa grupama.

Cluster Membership

Case Number	Cluster	Distance
1	1	4,055
2	1	,667
3	1	4,216
4	2	5,255
5	2	2,775
6	2	,656
7	2	1,671
8	2	4,164
9	2	3,408
10	2	,582
11	2	1,725
12	2	,782
13	2	1,725
14	2	1,698
15	3	2,981
16	3	4,714
17	3	,471
18	3	4,069
19	3	,745
20	3	4,269

Tabela finalnih centara klastera je izračunata kao sredina svih varijabli unutar klastera.. Finalni centri klastera reprezentuju karakteristike tipičnog entiteta za svaki klaster. U našem primeru u prvoj grupi su ekipe koje uglavnom dobijaju i kod kuće i na strani, u drugoj grupi su neke prosečne ekipe, dok su u trećoj grupi slaniije ekipe koje češće gube utakmice.

Final Cluster Centers

	Cluster		
	1	2	3
DomaćinBodovi	15	12	8
GostBodovi	18	8	5

Sledeća tabela prikazuje euklidsko odstojanje između krajnjih centara klastera. Veće odstojanje između klastera odgovara većim razlikama među klubova u tim klasterima. U našem primeri klasteri 1 i 3 se najviše razlikuju, dok su najbližnije grupe 1 i 3.

Distances between Final Cluster Centers

Cluster	1	2	3
1		10,996	14,926
2	10,996		5,313
3	14,926	5,313	

Sledeća tabela pokazuje koliko u kom klasteru ima entiteta. Vidimo da u drugom klasteru ima najveći broj entiteta (11), to su osrednje ekipe. Trećem klasteru pripadaju 6 ekipa koje uglavnom lošije igraju i pretendenti su za eventualno ispadanje iz lige, dok će se, najverovatnije, ekipe iz prvog klastera boriti za skudeto.

Number of Cases in each Cluster

Cluster	1	3,000
	2	11,000
	3	6,000
Valid		20,000
Missing		,000

Diskriminaciona analiza

Zadatak:

Na osnovu podataka iz matrice podataka, a na osnovu promenljivih *UkupnoOdigrano* i *UkupnoBodovi* klasifikovati ekipe u dve grupe tako da se u prvoj grupi nalaze ekipe koje imaju više od 50% ukupnog učinka (imaju više od polovine maksimalnog broja bodova iz svih odigranih utakmica), a u drugoj grupi ostale ekipe (imaju manje od 50% ukupnog učinka). Formirati diskriminacionu funkciju na osnovu promenljivih *DatoGolova* i *PrimljenoGolova*.

Rešenje:

Najpre moramo podeliti objekte u dve grupe po traženom kriterijumu. To ćemo uraditi na taj način što ćemo formirati novu promenljivu koja će se zvati *PromenljivaKlasifikacije* koja će imati vrednost 1 za objekte koje pripadaju prvoj grupi, a vrednost 0 za objekte koji pripadaju drugoj grupi. Izraz na osnovu kojeg se određuje ova vrednost je:

$$\text{UkupnoBodovi} > (\text{UkupnoOdigrano} * 3/2)$$

Objašnjenje izraza: Maksimalan broj bodova iz odigranih utakmica je $\text{UkupnoOdigrano} * 3$. 50% efikasnosti je polovina maksimalnog mogućeg broja bodova, dakle $\text{UkupnoOdigrano} * 3/2$. Ako je broj osvojenih bodova veći od polovine ukupno mogućih bodova nova promenljiva (koju ćemo nazvati *PromenljivaKlasifikacije*) uzeće vrednost 1, a u suprotnom nula.

Da bi ovo postigli idemo u meni *Transform* → *Compute*. Kao novu promenljivu (*Target Variable*) upisujemo *Promenljiva klasifikacije*. U polje *izraz za izračunavanje promenljive* (*Numeric Expression*) upisujemo gore dat izraz. Pritiskom na dugme OK dobijamo još jednu varijablu (*PromenljivaKlasifikacije*) koja ima vrednosti 0 ili 1 i pomoću koje klasifikujemo objekte u grupe. U našem slučaju samo prve tri ekipe pripadaju grupi ekipa koja ima učinak veći od 50%.

Naredbom iz menija: *Analyze* → *Classify* → *Discriminant...* otvara se novi prozor i započinjemo proces diskriminacione analize. Kao varijablu za grupisanje (*Grouping Variable*) unosimo našu novoformiranu promenljivu – *PromenljivaKlasifikacije*. Još jedna stvar koju moramo uraditi sa ovom promenljivom pre nego što krenemo sa diskriminacionom analizom je da definišemo opseg ove promenljive. Objekti koji su izvan definisanog opsega neće biti uzeti u analizu. Kako kod nas svi objekti treba da budu uključeni u analizu definišemo granice opsega (pritiskom na dugme *Define Range*) od 0 do 1.

Iz zahteva zadatka uočavamo da su promenljive *DatoGolova* i *PrimljenoGolova* nezavisne promenljive koje će se koristiti u našoj analizi, pa te promenljive ubacujemo u polje nezavisnih promenljivih (*Independents*).

Da bi mogli da vidimo sredine, standardne devijacije grupa kao i koeficijente klasifikacione funkcije čekirajmo u dialog boksu *Statistics* opcije *Means* i *Fisher's Function coefficients*.

Da bi mogli da vidimo praktične rezultat korišćenja diskriminacionog modela čekirajmo u dialog boksu *Classify* opciju *Leave-one-out classification*.

Da bi zapamtili diskriminacione skorove za svaki objekat (ekipu) čekirajmo opciju u dijalog boksu *Save* opciju *Probabilities of group membership*.

Klasifikaciona funkcija se koristi da pridruži objekte grupama. Postoji posebna funkcija za svaku grupu. Za svaki objekat se formira diskriminacioni skor. Diskriminacioni model svaki objekat pripisuje grupi čija je klasifikaciona funkcija dala najveći skor. Koeficijenti za promenljivu *DatoGolova* su manji za prvu grupu (manje od 50% uspešnosti) nego za drugu (više od 50% uspešnosti), što u našem primeru znači da će ekipe koje su dale manje golova ići u prvu, a ekipe koje su dale više golova ići u drugu grupu. Slično objekti (ekipe) koje imaju veći diskriminacioni skor za promenljivu *PrimljenoGolova* idu u prvu grupu, a objekti koji imaju manji diskriminacioni skor za ovu promenljivu idu u grupu boljih ekipa.

Classification Function Coefficients

	PromenljivaKlasifikacije	
	,00	1,00
DatoGolova	,065	,612
PrimljenoGolova	1,033	,054
(Constant)	-11,611	-8,816

Fisher's linear discriminant functions

Dokaz gore izloženog možemo potražiti i pregledavanjem diskriminacionih skorova koji su sada pridruženi svakom objektu (ekipi) po svakoj grupi.

Na primer izračunati diskriminacioni skorovi za prvu ekipu na tabeli su

Ekipa	Grupa ,00	Grupa 1,00
Juventus	,00001	,99999

što znači da će ova ekipa ući u drugu – efikasniju grupu.

Matrica korelacije unutar grupa pokazuje korelaciju između nezavisnih varijabli

Pooled Within-Groups Matrices

		DatoGolova	PrimljenoGolova
Correlation	DatoGolova	1,000	,515
	PrimljenoGolova	,515	1,000

Tabela grupne statistike pokazuje sredine i standardne devijacije grupa.

Group Statistics

PromenljivaKlasifikacije		Mean	Std. Deviation	Valid N (listwise)	
				Unweighted	Weighted
,00	DatoGolova	17,24	6,666	17	17,000
	PrimljenoGolova	20,06	4,437	17	17,000
1,00	DatoGolova	25,67	3,215	3	3,000
	PrimljenoGolova	9,67	3,055	3	3,000
Total	DatoGolova	18,50	6,932	20	20,000
	PrimljenoGolova	18,50	5,662	20	20,000

Standardizovani koeficijenti nam omogućuju da poredimo promenljive merene na različitim skalama

Standardized Canonical Discriminant Function Coefficients

	Function
	1
DatoGolova	-,908
PrimljenoGolova	1,095

Strukturna matrica pokazuje korelaciju nezavisne promenljive sa diskriminantnom funkcijom.

Structure Matrix

	Function 1
PrimljenoGolova	,628
DatoGolova	-,344

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions.
Variables ordered by absolute size of correlation within function.

Tabela klasifikacije pokazuje praktične rezultate primene diskriminacionog modela.

Classification Results^b

			Predicted Group Membership		Total
			,00	1,00	
Original	Count	PromenljivaKlasifikacije ,00	17	0	17
		100,0	0	3	3
	%	,00	100,0	,0	100,0
		100,0	,0	100,0	100,0

b. 100,0% of original grouped cases correctly classified.

Od objekata korišćenih za kreiranje modela sve 17 ekipe prethodno klasifikovane u prvu grupu su i sada klasifikovane tačno kao i sve 3 ekipe iz druge grupe, što nam kazuje da je 100% objekata tačno klasifikovano.