

Faktorska analiza

Faktorska analiza predstavlja jednu od najpopularnijih multivarijantnih tehnika koja ima dva cilja:

1. Identifikacija i razumevanje osnovne ideje, odnosno zajedničkih karakteristika za više varijabli.
2. Smanjivanje broja varijabli u analizi kada ih je previše, pri čemu se neke od njih „preklapaju“ jer imaju slično značenje i ponašanje.

Faktorska analiza je tehnika međuzavisnosti jer traži grupu varijabli koje su slične u smislu da se „zajedno pomeraju“ i zbog toga imaju veliku međuzavisnost. Kada jedna varijabla ima veliku vrednost, onda i ostale varijable u grupi imaju veliku vrednost. U marketing istraživanjima ova tehnika se vrlo često služi za analiziranje rejtinga proizvoda ili karakteristika brenda, stavova i slično.

Za efikasnu primenu faktorske analize, pa i drugih multivarijantnih tehnika međuzavisnosti, potrebno je da postoji minimalna količina redundancije varijabli, odnosno da se varijable barem malo preklapaju u svom značenju. Zahvaljujući toj redundanciji moguće je otkriti šablon u ponašanju varijabli, odnosno osnovnu ideju (faktor) kojom su prožete.

Sa druge strane, kod multivarijantnih tehnika zavisnosti, gde postoji jedna zavisna i više nezavisnih varijabli, redundancija nije uopšte poželjna jer može da utiče na visinu regresionog koeficijenta svake nezavisne varijable pojedinačno u modelu. Drugim rečima, dolazi do multikolinearnosti pa se ne vidi jasno koliko koja nezavisna varijabla ima uticaja na zavisnu varijablu.

Identifikacija i razumevanje osnovne ideje

Kada se pravi anketa, često dolazi do redundancije između postavljenih pitanja. Kada se dizajnira upitnik, istraživač nikada ne može biti siguran da je pokrio pravu temu u potpunosti sa pitanjima koja je odabrao. Zbog toga se često u upitnik uvrštava više pitanja koja se na isti ili sličan način odnose na temu istraživanja. Na primer, u istraživanju na tržištu šampona za kosu, upitnik sadrži najmanje 12 pitanja koja pokrivaju za nijansu različite aspekte iste teme: pranje kose.

U ovom kontekstu, jedan od glavnih ciljeva faktorske analize je da traži grupu sličnih iskaza od strane respondenata jer oni izražavaju istu osnovnu ideju na načine koji se razlikuju u nijansama. Mi želimo da identifikujemo tu osnovnu ideju i da je izmerimo. Te osnovne ideje se nazivaju faktorima. Faktori se ne mogu identifikovati i izmeriti direktno. Oni se mogu otkriti preko odnosa između varijabli koje ih svojim ponašanjem ispoljavaju.

Pošto se faktori statistički izdvajaju, svi faktori su inicijalno međusobno nepovezani (ortogonalni). Ovo pojednostavljuje razumevanje širokog spektra varijabli koje opisuju neku kategoriju iz sektora usluga ili proizvodnje. Takođe se stvara okvir za dalju analizu podataka.

Iako faktori inicijalno nisu povezani, to ne znači da se to odnosi i na pojedine, originalne iskaze respondenata. Jedan iskaz respondenta može u sebi da sadrži više faktora. Takvi iskazi su faktorski kompleksni. Sami po sebi, ovi iskazi neće jasno definisati ni jedan faktor, ali mogu da doprinesu u opisivanju faktora od kojih se sastoje.

Redukcija podataka

Drugi glavni cilj u faktorskoj analizi je da se smanji redundancija ili preklapanje varijabli, odnosno pitanja u anketi, da bi se smanjili troškovi i opterećenje respondenata u budućim, sličnim istraživanjima. Kada se otkrije grupisanje varijabli uz pomoć faktorske analize, moguće je uraditi jednu od sledećih intervencija:

1. Eliminisanje jedne ili više varijabli (pitanja u anketi) u svakoj grupi.
2. Kombinovanje dva ili više iskaza sa sličnim značenjem u jedan iskaz.
3. Izbor po jedne varijable iz svake grupe koja najbolje karakteriše određeni faktor i koristiti je kao marker varijablu.
4. Korišćenje jednog faktor skora koji predstavlja prosek svih varijabli vezanih za dati faktor.
5. Kombinacija gore nabrojanog.

Faktorska analiza se često koristi da bi se racionalizovao broj pitanja koji se nalaze u anketama. Nakon racionalizacije događa se da se u istraživanju koriste identifikovani faktori a ne originalne varijable.

Rezultati faktorske analize u mnogome zavise od samog istraživača, jer će analiza otkriti obrasce ponašanja bilo kojih varijabli koje istraživač uključi u model. Ukoliko se odluči za čitav niz varijabli koje su vezane za jednu ideju, a kod drugih ideja postoji znatno manje varijabli, definišaće se jedan faktor sa velikom vrednošću. Zbog toga dizajniranje upitnika bitno utiče na rezultate faktorske analize. Cilj faktorske analize je da pronađe grupu povezanih varijabli, ali ne i da utvrdi važnost tog grupisanja.

Primer: Restorani brze hrane

U jednom gradu izvršena je anketa u kojoj su respondenti ocenjivali restorane brze hrane. Cilj studije je bio da se utvrdi relativna važnost pojedinih karakteristika restorana tog tipa. Oko 400 respondenata je anketirano i svaki od njih je morao da oceni 23 različite karakteristike ocenom od 1 do 10, pri čemu je 1 bila najlošija ocena a 10 najbolja. Kao dodatak, tražena je i ukupna ocena za svaki lanac brze hrane, takođe na skali od 1 do 10.

Nakon ankete urađena je faktorska analiza da bi se grupisale karakteristike (osnovne ideje) koje su najslabije. Nakon što su napravljene grupe odnosno određeni faktori (**tabela**), potrebno je svakom faktoru dati odgovarajući naziv tako da povezuje sve karakteristike koje se nalaze u datoj grupi. Faktorsko opterećenje (factor loading) ukazuje na relativnu važnost svake karakteristike u definisanju faktora. To je zapravo koeficijent korelacije između svake karakteristike i samog faktora. Što je veća korelacija, data karakteristika bolje opisuje faktor. Ova korelacija može biti i pozitivna i negativna.

U ovo primeru, faktor 1 ima najveće opterećenje kod karakteristike „Koriste visoko kvalitetne sastojke u hrani“ (0,76), zatim „Njihova hrana zaista ima dobar ukus“ (0,74) itd. Zbog navedenih karakteristika kod prvog faktora, faktoru je dodeljen naziv „Ukusna, visoko kvalitetna hrana“. Ovaj naziv je dodeljen arbitrarno i mogući su i drugi odgovarajući nazivi.

Kod drugog faktora najveće pristustvo je kod karakteristika „Unutrašnjost uvek izgleda lepo“ (0,70) i „Osoblje je uredno i čisto“ (0,66) itd. Usled toga, logičan naziv za faktor 2 je „Zgrada i unutrašnji izgled su lepi“ jer se ovaj faktor odnosi na objekat sam po sebi i njegov izgled.

Tabela: Ocenjivanje restorana brze hrane – glavna pitanja

Faktor	Faktorsko opterećenje (Factor Loading)
1. Ukusna, visoko kvalitetna hrana	
Koriste visoko kvalitetne sastojke u hrani.	0.76
Njihova hrana zaista ima dobar ukus.	0.74
Uvek mogu da pronadem u meniju nešto što mi se sviđa.	0.68
Imaju konstantan kvalitet hrane.	0.66
Hrana se sprema po narudžbi.	0.66
Ovo je mesto od poverenja.	0.65
Porcije su odgovarajuće.	0.55
2. Zgrada i unutrašnji izgled su lepi	
Unutrašnjost uvek izgleda lepo.	0.70
Osoblje je uredno i čisto.	0.66
Imaju puno mesta za parking.	0.66
Toaleti su čisti.	0.61
Zgrade su atraktivne.	0.59
Nikada ne ostaju bez jela iz menija.	0.56
Nikada nisu otvoreni do kasno.	0.55
Osećam se prijatno pored drugih gostiju.	0.52
3. Okruženje	
Možeš biti brzo uslužen.	0.71
Zaposleni su ljubazni.	0.68
Mesto je uvek uredno i čisto.	0.59
Dobio sam pravu vrednost za svoj novac.	0.58
4. Meni ima novih specijaliteta, dodatne pogodnosti	
Povremeno imaju nova jela.	0.73
Često imaju specijalnu ponudu uz popust.	0.73
Meni pruža širok izbor.	0.62
Postoji mnogo odgovarajućih lokacija.	0.49

Kada se dobijeni koeficijent za prvu karakteristiku (0,76) podigne na kvadrat, dobija se da je oko 58% varijacija u oceni zajedničko sa faktorom. Ovako mali koeficijent determinacije je dobijen zato što su mnoga pitanja u anketi izbačena još prilikom dizajniranja upitnika zbog uštede u resursima. To znači da su mnoge sličnosti i redundantnosti između pitanja ranije eliminisane. Veće opterećenje faktora bi moglo lako da se dobije dodavanjem novih pitanja koja su slična već postojećim.

Faktor 3 je mnogo teži za interpretaciju. Karakteristike nisu logički povezane i ne čine celinu. Ipak, faktorska analiza ukazuje da se ove karakteristike „pomeraju zajedno“ u ocenjivanju. To znači kada respondent da relativno visoku ocenu jednoj karakteristici iz grupe, i ostale karakteristike iz grupe imaju tendenciju ka visokoj oceni. Zbog toga nije bilo jednostavno dodeliti pravi naziv faktoru. Na kraju je izbor pao na naziv „Okruženje“.

Četvrtom faktoru je bilo lakše dati ime jer se karakteristike odnose uglavnom na meni restorana. Zbog toga je dodeljen naziv „Meni ima novih specijaliteta, dodatne pogodnosti“.

Izbor imena za faktore je izuzetno bitan jer je kasnije prilikom donošenja bilo kakvih odluka fokus upravo na nazivima. To je ponekad lakše kada postoje i pozitivno i negativno opterećenje faktora jer nam negativni govore šta dati faktor ne predstavlja.

Na osnovu definisanih faktora stvorena je globalna slika o restoranima brze hrane koju strateški menadžment mora da uvaži. Pored toga, 23 karakteristike su smanjene na svega 4 glavne varijable (faktore) koje mogu biti upotrebene za dodatnu analizu podataka. Iako ova četiri faktora ne pokrivaju sve što su pokrivalo 23 karakteristike, ipak pokrivaju veći deo problema istraživanja.

Primer: Auto-dileri

160 kupaca automobila zamoljeno je da oceni auto-dilere koje su posetili na osnovu 20 datih karakteristika. Faktorska analiza je dala vrlo jasne rezultate na osnovu ankete.

Tabela: Ocenjivanje restorana brze hrane – glavna pitanja

Faktor	Faktorsko opterećenje (Factor Loading)
1. Lični pristup	
Jasno su mi odgovarali na pitanja.	0.86
Diler kojem se može verovati.	0.84
Potrude se da shvate moje potrebe.	0.80
Objasne uslove prodaje.	0.76
Prodavci su ljubazni i kulturni.	0.72
Nisu izbegavali moja pitanja.	0.64
Dobro servisno odeljenje.	0.57
2. Izbor robe	
Mnogi modeli ili tipovi su na zalihama.	0.87

Mnogo boja i opcija.	0.86
Otvoreno uveče ili subotama radi servisa.	0.53
3. Uslovi prodaje	
Dobre kamate.	0.80
Dobra cena za razmenu.	0.63
Cene su vrlo konkurentne.	0.56
4. Objekti	
Dovoljno prostora za parking.	0.71
Atraktivna izložbena sala.	0.68
5. Karakteristike vozila.	
Poređenja sa drugim markama.	0.80
Više od jedne marke vozila.	0.54
Ukazano na mnoge osobine vozila.	0.51
6. Nema pritiska	
Nisu navalentni.	0.80
Zovu kasnije da provere da li sam zadovoljan.	-0.46

Identifikovano je šest faktora na osnovu 20 karakteristika. Ovim verovatno problem istraživanja nije u potpunosti pokriven. Faktorska analiza ne može da ukaže na to šta nedostaje u nizu karakteristika. Jedan od načina da se proveri pokrivenost je da se uradi višestruka regresiona analiza gde bi zavisna varijabla bila ukupna ocena auto-dilera i da se izračuna koeficijent determinacije. Ako je on nizak (recimo ispod 70%), može se reći da određene karakteristike, odnosno određena pitanja treba dodati u anketu.

Vrste faktora

Postoje dve vrste faktora u faktorskoj analizi: zajednički i specifični faktori. Zajednički faktori su oni čije varijacije su podeljene između dve ili više varijabli iz skupa varijabli. Specifični faktori su oni čije su varijacije vezane za svaku varijablu pojedinačno i nisu obuhvaćene zajedničkim faktorima. Skoro svaka varijabla ima makar malu količinu specifične varijanse. Faktorska analiza identifikuje samo zajedničke faktore. Bitno je znati da specifični faktori nekad mogu biti od većeg značaja u nekom istraživanju od zajedničkih. Specifični faktori se mogu „izvući na površinu“ dodavanjem novih pitanja u anketi.

Vrste faktorske analize

U osnovi postoje dva pristupa u otkrivanju faktora: preko analize glavnih komponenti i faktorska analiza u užem smislu. Ukratko, analiza glavnih komponenti identifikuje sve izvore varijacija u

skupu varijabli, uključujući zajedničke i specifične faktore. Faktorska analiza pokušava da otkrije i objasni samo specifične faktore koji su zastupljeni kod dve ili više varijabli.

Kod analize glavnih komponenti 50 varijabli koje se analiziraju biće zamenjene sa 50 glavnih komponenti, s tim što će samo nekoliko glavnih komponenti imati veliko prisustvo u varijabilitetu podataka pa će samo one biti interesantne. Zbog toga se analiza glavnih komponenti smatra pre svega tehnikom za redukciju obima podataka. Glavni cilj faktorske analize u užem smislu je da se shvati i objasni varijabilitet podataka.

Model faktorske analize

Razlika između analize glavnih komponenti i faktorske analize je u tome što se faktorska analiza zasniva na matematičkom modelu. Početak razvoja faktorske analize se vezuje za radove Čarlsa Spirmana (Charles Spearman), s početka XX veka.

Opšti faktorski model ima sledeći oblik:

$$X_i = a_{i1} F_1 + a_{i2} F_2 + \dots + a_{im} F_m + e_i$$

gde su:

X – vrednost varijable (skor faktora za varijablu) sa aritmetičkom sredinom nula i varijansom jedan,

i – redni broj varijable,

F – faktori koji su međusobno nezavisni,

m – redni broj faktora,

a – faktorsko opterećenje (konstanta),

e – specifični faktor vezan samo za datu varijablu.

Varijansa varijable X_i je pri tome:

$$\text{Var}(X_i) = 1 = a_{i1}^2 \text{Var}(F_1) + a_{i2}^2 \text{Var}(F_2) + \dots + a_{im}^2 \text{Var}(F_m) + \text{Var}(e_i) = a_{i1}^2 + a_{i2}^2 + \dots + a_{im}^2 + \text{Var}(e_i)$$

pri čemu $a_{i1}^2 + a_{i2}^2 + \dots + a_{im}^2$ predstavlja komunalitet varijable X_i (deo varijanse koji je povezan sa zajedničkim faktorima), a $\text{Var}(e_i)$ je specifična varijansa varijable X_i (deo varijanse koji nije povezan sa zajedničkim faktorima). Takođe je dokazano da je korelacija između varijabli X_i i X_j :

$$r_{ij} = a_{i1} a_{j1} + a_{i2} a_{j2} + \dots + a_{im} a_{jm}$$

Prema tome, dva faktorska skora mogu biti jako povezana ako imaju visoko opterećenje za iste faktore. Pošto komunalitet ne može da bude veći od 1, mora biti zadovoljen uslov:

$$-1 \leq a_{ij} \leq 1.$$

Postupak faktorske analize

Tipična faktorska analiza se izvodi u nekoliko koraka:

1. Izračunavanje kompletne tabele koeficijenata korelacije između svih originalnih varijabli.
2. Izračunavanje faktorskog opterećenja (factor loading) iz matrice koeficijenata korelacije.
3. Rotacija zajedničkih faktora radi veće razumljivosti.
4. Interpretacija zajedničkih faktora, uključujući i izbor adekvatnog naziva.
5. Izračunavanje faktorskih jednačina za svaku jedinicu posmatranja (factore score) radi dobijanja indeksnih brojeva (opciono).

U nastavku, termin „faktor“ će se zapravo odnositi na zajedničke faktore.

Tabela koeficijenata korelacije

U faktorskoj analizi traži se obrazac odnosa između velikog broja varijabli. To znači da moramo početi analizu sa pregledom korelacionih odnosa originalnih varijabli. Najčešće se koristi Pirsonov koeficijent proste korelacije koji pokazuje jačinu i smer veze između dve varijable.

Dobijena tabela koeficijenata korelacija može da doprinese boljoj identifikaciji, imenovanju i razumevanju faktora. Ukoliko istraživač žuri, često se preskače ovaj korak što može da se odrazi na kvalitet cele analize. Neki računarski programi automatski izračunavaju tabelu koeficijenata korelacije, dok se kod neki posebno mora tražiti ova tabela.

Za izračunavanje Pirsonovog koeficijenta proste korelacije potrebno je da obe varijable imaju vrednosti sa intervalne ili racio skale, dok se u slučaju ordinarne skale primenjuje specijalna verzija faktorske analize. Vrednosti sa nominalne skale se mogu analizirati jedino ako varijabla ima samo dva modaliteta (na primer, obeležje je „vlasništvo automobila“, a modalitetu su „ima“ i „nema“).

Izračunavanje faktorskog opterećenja (zajedničkih faktora)

Potrebno je „izvući“ zajedničke faktore koji se nalaze u tabeli koeficijenata korelacije. Ovaj korak se obično izvodi uz pomoć analize glavnih komponenti.

Analiza glavnih komponenti pronalazi grupe varijabli koje imaju visoke koeficijente u okviru grupe a male u odnosu na druge grupe. Ova analiza će izvući onoliko glavnih komponenti koliko ima i varijabli, zato što ona obuhvata i zajedničke i specifične varijacije podataka. Pažnja istraživača se zadržava na nekoliko prvih glavnih komponenti koje imaju najveći uticaj (najveće faktorsko opterećenje) i obuhvataju najveći deo varijabiliteta podataka. Tih nekoliko glavnih komponenti predstavljaju faktore.

Ono što se želi postići je da faktorsko opterećenje ili bude blizu nule, što znači da nije povezano sa datom varijablom ili da bude znatno udaljeno od nule, što bi značilo da je data varijabla znatno

povezana sa faktorom. Ukoliko varijabla ima veliko faktorsko opterećenje samo za jedan faktor a za ostale ne, onda je lako identifikovati taj faktor.

Metod, dakle, polazi od modela glavnih komponenti, gde će glavnih komponenti biti isto onoliko koliko ima i originalnih varijabli. Glavne komponente su zapravo linearna kombinacija originalnih varijabli:

$$Z_1 = b_{11} X_1 + b_{12} X_2 + \dots + b_{1p} X_p$$

$$Z_2 = b_{21} X_1 + b_{22} X_2 + \dots + b_{2p} X_p$$

...

$$Z_p = b_{p1} X_1 + b_{p2} X_2 + \dots + b_{pp} X_p$$

gde su vrednosti b_{ij} ajgenvektori korelacione matrice. Sledi ortogonalna transformacija za dobijanje vrednosti varijabli. Inverzna linearna kombinacija glasi:

$$X_1 = b_{11} Z_1 + b_{21} Z_2 + \dots + b_{p1} Z_p$$

$$X_2 = b_{12} Z_1 + b_{22} Z_2 + \dots + b_{p2} Z_p$$

...

$$X_p = b_{1p} Z_1 + b_{2p} Z_2 + \dots + b_{pp} Z_p$$

Za faktorsku analizu zadržava se samo m komponenti od ukupnog broja komponenti p :

$$X_1 = b_{11} Z_1 + b_{21} Z_2 + \dots + b_{m1} Z_m + e_1$$

$$X_2 = b_{12} Z_1 + b_{22} Z_2 + \dots + b_{m2} Z_m + e_2$$

...

$$X_p = b_{1p} Z_1 + b_{2p} Z_2 + \dots + b_{mp} Z_m + e_p$$

gde je e_i linearna kombinacija ostalih, izostavljenih glavnih komponenti, od Z_{m+1} do Z_p .

Za dobijanje faktorskih jednačina potrebno je Z_i podeliti sa standardnom devijacijom, $\sqrt{\lambda}$, koja je kvadratni koren odgovarajuće ajgenvrednosti u korelacionojoj matrici:

$$X_1 = \sqrt{\lambda_1} b_{11} F_1 + \sqrt{\lambda_2} b_{21} F_2 + \dots + \sqrt{\lambda_m} b_{m1} F_m + e_1$$

$$X_2 = \sqrt{\lambda_1} b_{12} F_1 + \sqrt{\lambda_2} b_{22} F_2 + \dots + \sqrt{\lambda_m} b_{m2} F_m + e_2$$

...

$$X_p = \sqrt{\lambda_1} b_{1p} F_1 + \sqrt{\lambda_2} b_{2p} F_2 + \dots + \sqrt{\lambda_m} b_{mp} F_m + e_p$$

gde je $F_i = Z_i / \sqrt{\lambda_i}$. Iz datog se vidi da su nerotirani faktori zapravo vrednosti glavnih komponenti pošto su transformisani tako da imaju varijansu jednaku jedinici. Nerotirani faktorski model onda glasi:

$$X_1 = a_{11}F_1 + a_{12}F_2 + \dots + a_{1m}F_m + e_1$$

$$X_2 = a_{21}F_1 + a_{22}F_2 + \dots + a_{2m}F_m + e_2$$

...

$$X_p = a_{p1}F_1 + a_{p2}F_2 + \dots + a_{pm}F_m + e_p$$

pri čemu je $a_{ij} = \sqrt{\lambda_j} b_{ji}$.

Rotacija faktora

Nakon analize glavnih komponenti, faktorska analiza počinje da „rotira“ komponente. Cilj je da se redefiniše i pojasni značenje svakog faktora. Postupak se svodi na preraspodelu uticaja faktora sa prve glavne komponente na ostale, tako da je ukupna varijansa koja je objašnjena preko faktora ravnomernije raspoređena između komponenti.

U koordinatnom sistemu, glavne komponente su predstavljene kao prave linije koje prolaze kroz ishodište i između varijabli koje se nalaze u vidu tačaka u prostoru. Ako su varijable u jačoj korelacionoj vezi, nalaze se blizu jedna drugoj. Prave linije (glavne komponente) prolaze kroz grupu bliskih varijabli. Te prave linije zapravo predstavljaju faktore koji se traže. Prave linije komponenti su međusobno pod uglom od 90 stepeni jer komponente nisu međusobno zavisne.

Najpoznatija rotacija je „varimax“ rotacija koja maksimizira sumu varijansi kvadrata faktorskih opterećenja.

Nakon „varimax“ rotacije, ili neke druge, faktorski model ima oblik:

$$X_1 = g_{11}F_1^* + g_{12}F_2^* + \dots + g_{1m}F_m^* + e_1$$

$$X_2 = g_{21}F_1^* + g_{22}F_2^* + \dots + g_{2m}F_m^* + e_2$$

...

$$X_p = g_{p1}F_1^* + g_{p2}F_2^* + \dots + g_{pm}F_m^* + e_p$$

gde F_i^* predstavlja novi, i -ti faktor.

Primer: Rotacija faktora na primeru kafe

Nakon ankete koja je sprovedena među potrošačima dobijeni su podaci o 14 različitih osobina kafe. Urađena je analiza glavnih komponenti i dobijene su četiri glavne komponente koje su prikazane u tabeli.

Najveća opterećenja prve komponente su 0,91; 0,90; 0,89 itd. Ovo su vrlo velika opterećenja i treba u mnogome da nam pomognu prilikom davanja imena prvoj komponenti. Problem je što ima previše varijabli (osobina kafe) kod kojih je opterećenje veliko, pa to čini interpretaciju komplikovanom. Takođe, kod druge glavne komponente, najveće opterećenje je 0,36, a kod treće -0,31. Potrebno je pronaći način za redistribuiranje ovih opterećenja da bi se postigla interpretacija koja ima smisla za sve faktore. To se postiže rotiranjem osa u koordinatnom sistemu koje predstavljaju komponente oko skupa originalnih podataka.

U ovom konkretnom slučaju, rotiraju se četiri komponente kroz „varimaks“ rotaciju i to na sledeći način:

- Sve ose ostaju pod pravim uglom (90°), jedna u odnosu na drugu.
- Svaka par komponenti obuhvata maksimalan broj tačaka (varijabli) u prostoru između njih.
- Varijanse faktorskih opterećenja između svih varijabli, na svakoj osi, su maksimizirane (odatile i naziv „varimaks“, što za posledicu ima nekoliko visoko varijabilnih opterećenja i mnogo malih, čak negativnih).

Tabela: Faktorsko opterećenje pre i posle rotacije

Osobine kafe	Glavne komponente				Varimax rotacija				Komunalitet
	1	2	3	4	1	2	3	4	
1.Prijatna aroma	0.86	-0.01	-0.20	0.04	-0.63	0.38	0.36	0.34	0.78
2.Iskričav ukus	0.91	-0.01	-0.01	-0.09	0.48	0.43	-0.53	0.38	0.83
3.Zreo ukus	0.86	-0.11	-0.28	0.00	-0.70	0.26	0.38	0.36	0.83
4.Bogat ukus	0.91	0.15	-0.001	-0.10	0.46	-0.53	0.54	0.29	0.87
5.Zadovoljavajući ukus	0.87	-0.002	-0.31	0.10	-0.74	0.38	0.30	0.32	0.87
6.Živ ukus	0.93	0.03	-0.02	-0.16	0.49	0.43	-0.59	0.35	0.90
7.Ukus prave kafe	0.90	-0.02	0.04	-0.21	0.42	0.38	-0.64	0.37	0.86
8.Dubok, originalni ukus	0.77	0.36	0.11	0.16	0.31	-0.74	0.27	0.22	0.77
9.Ukus tek skuvane kafe	0.79	-0.28	0.24	-0.09	0.23	0.24	0.52	-0.62	0.76
10.Topao ukus	0.87	0.25	0.22	0.17	0.28	-0.75	0.33	0.39	0.89
11.Čist i jasan ukus	0.89	0.11	-0.05	0.10	0.51	-0.55	0.36	0.36	0.82
12.Prženi ukus	0.76	-0.29	0.04	0.27	0.43	0.28	0.16	-0.67	0.74

13.Svež ukus	0.84	-0.27	0.19	0.12	0.33	0.32	0.36	-0.70	0.83
14.Ukupan utisak	0.90	0.04	0.08	-0.23	0.38	0.43	-0.65	0.34	0.86
Objašnjena varijansa u %	74.6	3.4	2.7	2.6	22.9	21.3	20.4	18.7	-
Kumulativno varijansa u %	74.6	78.0	80.7	83.3	22.9	44.2	64.6	83.3	-

Pod nazivom „varimaks rotacija“ u **tabeli** se nalaze faktorska opterećenja za svaku varijablu posle rotacije. Sada je prikladnije koristiti termin „faktor“ umesto „komponenta“. Uočava se da su visoka opterećenja kod prve komponente pre rotacije uglavnom nestala kod prvog faktora. Sada je lakše utvrditi šta je zapravo faktor 1 i šta nije. Visoka opterećenja, ako su pozitivna, govore šta faktor jeste, a negativna šta nije.

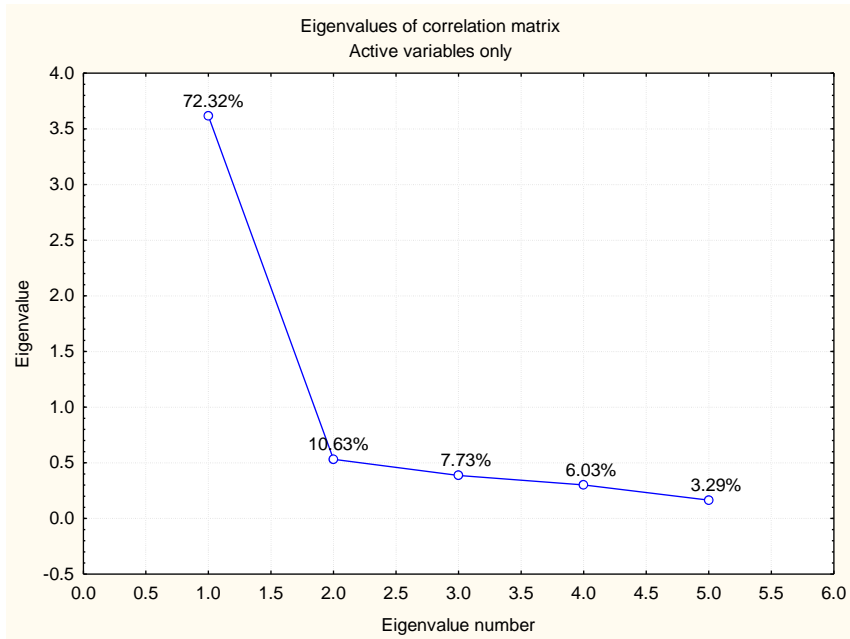
Procenat ukupne varijacije (objašnjena varijansa) se dramatično menja. Pre rotacije, prva komponenta je objašnjavala 74,6% od ukupne varijacije, sledeća 3,4% itd. Nakon rotacije postignuta je ravnomernost u objašnjenju varijansi između faktora (od 18,7% do 22,9%) što govori o redistribuciji faktorskog opterećenja.

Negativna opterećenja su se pojavila kod sva četiri faktora posle rotacije. Ona ukazuju na to što faktor ne predstavlja. Zbog načina na koji se rotacija izvodi, faktoru se dodeljuje ime na osnovu najvećeg opterećenja i to bez obzira na predznak. Najveće faktorsko opterećenje ukazuje na varijable koje imaju najjaču korelaciju sa datim faktorom.

Koliko faktora treba rotirati? Iako glavnih komponenti ima onoliko koliko ima i varijabli, najveći procenat varijacija podataka je objašnjen sa svega nekoliko prvih komponenti. Zbog toga su u primeru sa kafom odabrane četiri komponente za rotaciju. Ostale komponente bi samo doprinele konfuziji i teškoj interpretaciji. Pošto su varijable gotovo uvek standardizovane pre analize, nas interesuju one komponente koje imaju varijansu veću od 1 jer u sebi sadrže veće varijacije nego pojedinačne varijable. Ovo je samo jedan od nekoliko načina. U praksi se preporučuje da se uradi više načina za izbor broja faktora pre donošenja konačne odluke. Zbog toga u nastavku sledi opis još jednog metoda: Scree testa.

Scree test

Scree test je grafički metod za određivanje broja komponenti za rotaciju. Na linijskom dijagramu se predstavljaju ajgenvrednosti komponenti počevši od najveće. Traži se mesto na kojem linija naglo menja pravac i do te tačke se broje komponente koje će biti uključene u analizu. Na slici je prikazan grafik iz primera sa pilićima gde se uočava nagli prelom linije kod druge komponente, što znači da će za rotaciju biti izdvojena samo dva faktora koji zajedno objašnjavaju 82,95% varijacija.



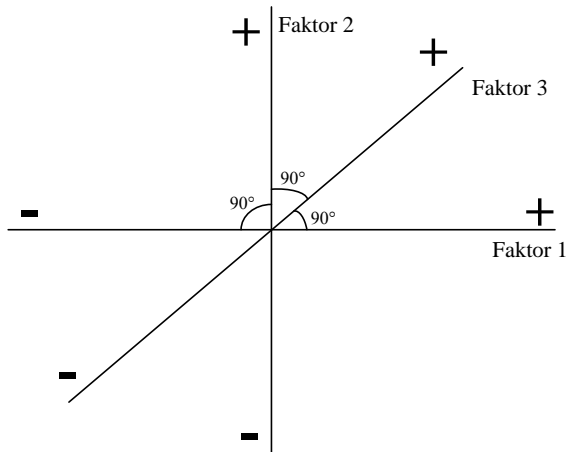
Slika: Grafički prikaz ajgenvrednosti za primenu Scree metoda

Ortogonalna i kosa rotacija

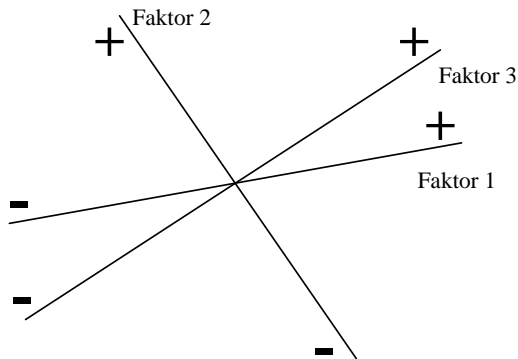
Kod ortogonalne rotacije, ose koje predstavljaju faktore ostaju pod pravim uglom i pre i posle rotacije. Kao posledica toga, faktori su uvek u potpunosti nepovezani. Ovo je tradicionalni pristup koji se prvi pojavio.

Neki analitičari su kasnije zastupali stav da „podaci govore sami za sebe“ i da se zanemari ograničenje o ortogonalnosti faktora. U tom slučaju ose same zauzimaju najbolju poziciju bez obzira na položaj ostalih. Rotacija ove vrste se zove „kosa rotacija“ (oblique rotation). Mnogi programski paketi ovaj oblik rotacije nude kao opciju.

Primeri za ortogonalnu i kosu rotaciju dati su na slikama. Na slici uočava se da su sve tri ose ostale pod pravim uglom nakon rotacije. Na slici, međutim, nakon rotacije ose zauzimaju drugačiji položaj koji nije pod 90° .



Slika: Ortogonalna rotacija faktora



Slika: Kosa rotacija faktora

Različiti programi pružaju različite opcije za kosu rotaciju. Na primer, analitičar može da odabere ciljnu varijablu i program će smestiti ose kroz nju. Ova varijabla oko sebe „okuplja“ ostale bliske varijable dajući im visoko faktorsko opterećenje. Postoji i takozvana „Prokrustova rotacija“ kod koje se prvo izvede ortogonalna rotacija, a zatim se ose rotiraju dok ose ne prođe kroz bilo koju varijablu koja ima najveće faktorsko opterećenje za bilo koji faktor.

Kosa rotacija može da pruži jasnije razumevanje strukture faktora. Ukoliko ortogonalna rotacija nije dovoljno ravnomerno rasporedila faktorska opterećenja na faktore, može se pribеći kosoj rotaciji koja će to efikasnije izvesti.

Nedostatak kose rotacije je taj što se između faktora javlja zavisnost (kolinearnost) pa se dobijeni rezultati ne mogu kvalitetno upotrebiti u višestrukoj regresionoj analizi.

Interpretacija rotiranih faktora

Sledeći korak je da se daju imena faktorima. Imena faktora uvek zavise od najvećeg i najmanjeg (negativnog) faktorskog opterećenja. Na primer, u tabeli, pod varimaks rotacijom, za faktor 1, najveća opterećenja su za „Zadovoljavajući ukus“ (-0,74), „Zreo ukus“ (-0,70) i „Prijatna aroma“ (-0,63), svi sa negativnim opterećenjem. To zapravo znači da ako je određena vrsta kafe visoko ocenjena po jednoj od ovih osobina, verovatno će biti visoko ocenjena i po ostalim osobinama i obrnuto. Pored toga, najveća pozitivna opterećenja kod faktora 1 su „Čist i jasan ukus“ (0,51), „Živ ukus“ (0,49), „Iskričav ukus“ (0,48) i „Bogat ukus“ (0,46). To zapravo znači da kafa koja ima čist, živ, iskričav i bogat ukus obično nema zadovoljavajući, zreo i prijatan ukus i aromu.

Istraživač je za prvi faktor izabrao ime „Zadovoljavajući ukus“ na osnovu najvećeg, negativnog faktorskog opterećenja. Faktor 2 je dobio ime „Toplina ukusa“, faktor 3 „Ukus prave kafe“ i faktor 4 „Svežina“.

Komunalitet

Poslednja kolona tabele pokazuje komunalitet svake varijable. Komunalitet je proporcija varijanse varijable koja je objašnjena zadržanim faktorima. Izračunava se tako što se saberu kvadrati faktorskih opterećenja varijable. Na primer, za varijablu „Prijatna aroma“, komunalitet iznosi:

$$(-0,63)^2 + 0,38^2 + 0,36^2 + 0,34^2 = 0,78.$$

To pokazuje da je 78% od ukupne varijacije varijable „Prijatna aroma“ obuhvaćeno sa četiri zajednička faktora. Isto tako, 22% varijacija se odnosi na specifičnost same varijable plus određeni iznos greške u merenju (e).

Za 14 varijabli kafe komunalitet se kreće od 74% do 90%. To znači da su prilično visoki i da mogu da obuhvate barem jedan faktor na da imaju umereno opterećenje za dva faktora. Ne postoji čisto „nezavisna“ varijabla u analizi.

U pretposlednjem redu tabele nalazi se objašnjena varijabla u procentima. Ona može da posluži kao zavisna varijabla u višestrukoj regresionoj analizi sa ostalim varijablama ili faktorima kao nezavisnim da bi se ocenila relativna važnost faktora.

Problem multikolinearnosti koji se javlja kod regresije može da se reši korišćenjem faktora kao nezavisnih varijabli umesto originalnih varijabli jer su oni, po definiciji, u potpunosti nezavisni.

Uočava se da je komunalitet za svaku varijablu u zbiru isti pre i posle rotacije. To znači da se od ukupne varijanse kod bilo koje varijable nije ništa izgubilo u procesu rotacije. To je zato što se u toku postupka same varijable ne pomeraju. Njihova lokacija u prostoru je određena međusobnim odnosima sa drugim varijablama. Prilikom rotacije osa koje predstavljaju faktore, za koliko se jedna varijabla približi osi, za toliko se druga udalji.

Primer: Zaposlenost u evropskim zemljama

Faktorska analiza je primenjena na podacima o zaposlenosti u evropskim zemljama. Korelaciona matrica kao i ajgenvrednosti i ajgenvektori za ovu seriju su izračunati kod analize glavnih komponenti. Pošto ima ukupno četiri ajgenvrednosti veće od jedan, primenjuje se grubo pravilo da bude i četiri faktora u modelu.

Izračunata su faktorska opterećenja za četiri faktora i devet varijabli:

$$X_1 = 0,90 F_1 - 0,03 F_2 - 0,34 F_3 + 0,02 F_4 + e_1 \quad (0,93)$$

$$X_2 = 0,66 F_1 + 0,00 F_2 + 0,63 F_3 + 0,12 F_4 + e_1 \quad (0,85)$$

$$X_3 = -0,43 F_1 + 0,58 F_2 - 0,61 F_3 + 0,06 F_4 + e_1 \quad (0,91)$$

$$X_4 = -0,56 F_1 + 0,15 F_2 - 0,36 F_3 + 0,02 F_4 + e_1 \quad (0,46)$$

$$X_5 = -0,39 F_1 - 0,33 F_2 + 0,09 F_3 + 0,81 F_4 + e_1 \quad (0,92)$$

$$X_6 = -0,67 F_1 - 0,55 F_2 + 0,08 F_3 + 0,17 F_4 + e_1 \quad (0,79)$$

$$X_7 = -0,23 F_1 - 0,74 F_2 - 0,12 F_3 - 0,50 F_4 + e_1 \quad (0,87)$$

$$X_8 = -0,76 F_1 + 0,07 F_2 + 0,44 F_3 - 0,33 F_4 + e_1 \quad (0,88)$$

$$X_9 = -0,36 F_1 + 0,69 F_2 + 0,50 F_3 - 0,04 F_4 + e_1 \quad (0,87)$$

Vrednosti u zagradama predstavljaju komunalitet. Na primer, komunalitet za varijablu X_1 (AGR, poljoprivreda, šumarstvo i ribarstvo) se izračunava na sledeći način:

$$\langle 0,90 \rangle^2 + \langle 0,03 \rangle^2 + \langle 0,34 \rangle^2 + \langle 0,02 \rangle^2 = 0,93$$

Komunaliteti su veliki za sve varijable osim za X_4 (PS, proizvodnja električne energije, gasa i vode). To znači da je veći deo varijabiliteta podataka osam varijabli obuhvaćeno zajedničkim faktorima.

Faktorska opterećenja koja su veća od 0,50, bez obzira na predznak, predstavljaju velika i umerena opterećenja koja pokazuju kako je varijabla povezana sa faktorom. Očigledno je da je varijabla X_1 gotovo u potpunosti određena faktorom 1, X_2 je mešavina faktora 2 i faktora 3, X_3 je određen faktorima 1 i 2 itd. Nepovoljno je što su pet od devet varijabli jako povezane sa dva faktora. Rotacija faktora će možda pružiti bolje rešenje.

Primenjena je „varimax“ rotacija sa Kaiser-ovom normalizacijom. Dobijen je sledeći model:

$$X_1 = 0,85 F_1 + 0,10 F_2 + 0,27 F_3 - 0,36 F_4 + e_1$$

$$X_2 = 0,11 F_1 + 0,30 F_2 + 0,86 F_3 - 0,10 F_4 + e_1$$

$$X_3 = -0,03 F_1 + 0,32 F_2 - 0,89 F_3 - 0,09 F_4 + e_1$$

$$X_4 = -0,19 F_1 - 0,04 F_2 - 0,64 F_3 + 0,14 F_4 + e_1$$

$$X_5 = -0,02 F_1 + 0,08 F_2 - 0,04 F_3 + 0,95 F_4 + e_1$$

$$X_6 = -0,35 F_1 - 0,48 F_2 - 0,15 F_3 + 0,65 F_4 + e_1$$

$$X_7 = -0,08 F_1 - 0,93 F_2 + 0,00 F_3 - 0,01 F_4 + e_1$$

$$X_8 = -0,91 F_1 - 0,17 F_2 - 0,12 F_3 + 0,04 F_4 + e_1$$

$$X_9 = -0,73 F_1 + 0,57 F_2 - 0,03 F_3 - 0,14 F_4 + e_1$$

Komunalitet je nepromenjen a faktori su i dalje nepovezani. Rešenje je nešto bolje jer je samo varijabla X_9 povezana nešto više sa dva faktora.

U sledećem koraku potrebno je dati nazive faktorima, što podrazumeva određeni stepen inovativnosti.

Faktor 1 ima visoko pozitivno opterećenje za varijablu X_1 (AGR – poljoprivreda, šumarstvo i ribarstvo), gde je opterećenje 0,85 i visoka negativna opterećenja za varijable X_8 (SPS – društvene i lične usluge), sa opterećenjem $-0,91$ i X_9 (TC - transport i komunikacije), sa opterećenjem $-0,73$. To znači da se meri stepen u kojem su ljudi zaposleni u poljoprivredi pre nego u državnoj upravi i komunikacijama. Naziv faktora 1 je zbog toga „ruralna industrija pre nego društveni servis i komunikacije“.

Faktor 2 ima negativno opterećenje za X_7 (FIN – finansije), i to $-0,93$ i dovoljno visoko pozitivno opterećenje za X_9 (TC - transport i komunikacije) od $0,57$. Usled toga je dobio naziv „nedostatak finansija“.

Faktor 3 ima visoko pozitivno opterećenje za X_2 (MIN – rudarstvo) od $0,86$ i negativna opterećenja za X_3 (MAN – prerađivačka industrija) od $-0,89$ i X_4 (PS - Proizvodnja električne energije, gasa i vode) od $-0,64$. Naziv ovog faktora je „rudarstvo pre nego proizvodnja“.

Faktor 4 ima visoka pozitivna opterećenja za varijable X_5 (CON – građevinarstvo) od $0,95$ i X_6 (SER – usluge) od $0,65$. Logično je da naziv bude „građevinarstvo i usluge“.

Na osnovu jednačina izračunavaju se faktorski skorovi za svaku varijablu pojedinačno. Rezultati su dati u tabeli.

Tabela: Rotirani skorovi faktora

Zemlja	Faktor 1	Faktor 2	Faktor 3	Faktor 4
Belgium	0.97	0.56	0.10	0.47
Denmark	0.89	0.47	0.03	0.67
France	0.56	0.78	0.15	0.25
Germany	-0.05	0.57	0.47	-0.58
Ireland	-0.48	-0.19	0.23	-0.02
Greece	-0.28	0.60	0.36	-0.03
Italy	-0.25	0.13	-0.17	-1.00
Luxembourg	0.46	0.36	-0.02	-0.92
Netherlands	1.36	1.56	0.03	2.08
Portugal	-0.66	0.45	0.37	-0.64

Spain	-0.24	0.11	0.09	-0.93
U.K.	0.50	1.14	0.35	0.04
Austria	-0.18	-0.05	0.71	-0.56
Finland	0.78	0.20	0.21	0.52
Iceland	0.18	0.04	0.06	-0.46
Norway	1.36	0.17	-0.20	0.41
Sweden	1.20	0.52	-0.04	0.74
Switzerland	-0.12	0.67	-0.01	-0.65
Albania	-3.16	1.82	-1.76	1.79
Bulgaria	-0.47	-1.56	0.57	0.65
Czech/Slovak Rep.	0.26	-1.45	-3.12	-0.44
Hungary	1.05	-1.70	-2.82	0.14
Poland	-0.97	-0.71	0.37	0.42
Romania	-1.11	-1.73	1.69	0.81
USSR (form.)	-0.08	-2.09	0.11	-0.14
Yugoslavia (form.)	-0.13	-1.48	1.70	-0.17
Cyprus	-0.46	0.32	-0.03	-1.08
Gibraltar	0.05	1.05	-0.08	-3.26
Malta	1.18	-0.49	0.79	1.31
Turkey	-2.15	-0.07	-0.15	0.56

Na osnovu analize faktorskih skorova uočava se da je vrednost faktora 1 visoka u Albaniji i Turskoj, što znači da je tamo akcent na ruralnoj industriji pre nego na društvenom servisu i komunikacijama. Bugarska, Mađarska, Rumunija i bivši SSSR imaju malo zaposlenih u finansijama, dok Holandija i Albanija imaju velik broj zaposlenih u toj oblasti. Ovo je uočljivo na osnovu faktora 2. Faktor 3 stavlja u kontrast Albaniju i bivšu Čehoslovačku sa jedne strane i Rumuniju i bivšu Jugoslaviju sa druge strane zbog broja zaposlenih u rudarstvu. Faktor 4 stavlja u kontrast Gibraltar, sa velikim brojem zaposlenih u građevinarstvu i uslugama, naspram Holandije i Albanije gde je obrnut slučaj.

Možda bi bilo racionalno i izvodljivo nastaviti analizu sa manjim brojem faktora i drugačijim metodima faktorske analize. Različiti softverski paketi mogu da daju ajgenvektore sa suprotnim predznacima. Takođe, obrnuti predznaci mogu da se pojave prilikom rotacije faktora, tako da faktorska opterećenja idu u suprotnom smeru od onog koji je naveden u ovom primeru. U takvom slučaju potrebno je obrnuti i interpretaciju.

Faktorska analiza u statističkom paketu STATISTICA

Koraci za izvođenje analize u programu su sledeći:

Pokretanje analize:

Statistics ▶ Multivariate Exploratory Technique ▶ Factor Analysis

Dobija se početni meni za analizu.

Definisanje tipa podataka koji se analizira:

Input file:

Izabrati „Raw Data“ ako su u pitanju sirovi podaci ili „Correlation Matrix“ ako su u pitanju koeficijenti korelacije koji su ranije izračunati.

Definisanje varijabli:

Quick ▶ Variables

Otvora se prozor sa spiskom varijabli od kojih treba odabrati one koje će biti uvrštene u analizu.

▶ OK

Dobija se drugi glavni meni sa opcijama „Quick“, „Advanced“ i „Descriptives“

Određivanje broja faktora i minimalne ajgen vrednosti:

Quick ▶ Maximum no. of factors (upisati željeni broj faktora)

Quick ▶ Minimum eigenvalue (ostaviti neka bude 1)

Izbor metoda faktorske analize i metoda analize glavnih komponenti:

Advanced ▶ Extraction Method (odabrati „Principal Components“)

Advanced ▶ Principal factor analysis (ostaviti sve neobeleženo)

I u ovom modulu se može definisati broj faktora i minimalna ajgen vrednost.

Izračunavanje korelacione matrice:

Descriptives ▶ Review correlations, means, standard deviations ▶ Quick ▶ Correlations

U modulu „Descriptives“ se nalazi još nekoliko opcija koje mogu biti od koristi prilikom analize.

Izračunavanje ajgen vrednosti:

Quick ▶ OK ▶ Eigenvalues

Izračunavanje faktorskih opterećenja:

Quick ▶ OK ▶ Quick (ili Loadings) ▶ Summary: Factor loadings

Grafički prikaz faktorskih opterećenja:

Quick ▶ OK ▶ Quick (ili Loadings) ▶ Plot of factor loadings, 2D

Rotacija podataka:

Quick ▶ OK ▶ Quick (ili Loadings) ▶ Factor rotation (izabrati vrstu rotacije iz padajućeg menija)

Scree metod određivanja broja faktora:

Quick ▶ OK ▶ Explained variance ▶ Scree plot

Izračunavanje komunaliteta:

Quick ▶ OK ▶ Explained variance ▶ Communalities

Izračunavanje faktorskih skorova:

Quick ▶ OK ▶ Scores ▶ Factor scores

