Faktorska analiza

Faktorska analiza predstavlja jednu od najpopularnijih multivarijacionih 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 vrlo često služi za analiziranje rejtinga proizvoda ili karakteristika brenda, stavova i slično.

Kod tehnika međuzavisnosti ne postoji podela na zavisne i nezavisne varijable jer su sve zapravo nezavisne. Ovim tehnikama se zapravo traži model odnosa između varijabli koji ima smisla sa aspekta problema istraživanja. Tehnike međuzavisnosti su zapravo heuristički, aproksimativni metodi kojima se traga za razumnim, smislenim, optimalnim rešenjima.

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

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

Kada se pravi anketa, često dolazi do redundantnosti 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 dečije hrane, upitnik može da sadrži najmanje 15 pitanja koja pokrivaju za nijansu različite aspekte iste teme: dečiju hranu.

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.

Identifikovani faktori reprezentuju osnovne ideje odnosno komponente koje su bitne, na primer, potrošačima prilikom vrednovanja nekog proizvoda. Za istraživača je mnogo lakše da se fokusira na nekoliko najvažnijih karakteristika proizvoda koje reprezentuju faktori nego na sve moguće karakteristike koje su posmatrane. Faktorska analiza na taj način pruža dobar osnov za razumevanje najvažnijih, suštinskih dimenzija ili ideja vezanih za posmatranu pojavu.

Faktorska analiza ima i svoja ograničenja koja se često navode u literaturi (Hair, Black, Babin, & Anderson, 2010):

- Pošto postoji mnogo tehnika kojima može da se izvede faktorska analiza, ne postoji konsenzus koja od njih je najbolja.
- Subjektivni aspekt je veoma izražen (koliko faktora, koja rotacija, kolika statistička značajnost faktorskih opterećenja) što dovodi do velikog razmimoilaženja u stavovima istraživača.
- Problem pouzdanosti je prisutan.

Racionalizacija preko faktorske analize

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 odnosno surogat.
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, definisaće 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 najsličnije. 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 (varijable, pitanja) i samog faktora. Što je veća korelacija, data karakteristika bolje opisuje faktor. Ova korelacija može biti i pozitivna i negativna. Ako je korelacija pozitivna, onda pokazuje u kojoj meri određena varijabla doprinosti tom faktoru, a ako je negativna, pokazuje koliko varijable ne učestvuje u datom faktoru.

U ovo primeru, faktor 1 ima najveće opterećenje kod karakteristike „Koriste visoko kvalitetne sastojke u hrani“ (0,78), zatim „Njihova hrana zaista ima dobar ukus“ (0,72) itd. Zbog navedenih karakteristika kod prvog faktora, faktoru je dodeljen naziv „Hrana“. Ovaj naziv je dodeljen arbitrarno i mogući su i drugi odgovarajući nazivi.

Kod drugog faktora najveće prisustvo je kod karakteristika „Unutrašnjost uvek izgleda lepo“ (0,72) i „Osoblje je uredno i čisto“ (0,69) itd. Usled toga, logičan naziv za faktor 2 je „Objekat“ 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. faktor	**Hrana**	
	Koriste visoko kvalitetne sastojke u hrani.	0.78
	Njihova hrana zaista ima dobar ukus.	0.72
	Uvek mogu da pronađem u meniju nešto što mi se sviđa.	0.69
	Imaju konstantan kvalitet hrane.	0.59
	Hrana se sprema po narudžbi.	0.59
	Ovo je mesto od poverenja.	0.59
	Porcije su odgovarajuće.	0.53
2. faktor	**Objekat**	
	Unutrašnjost uvek izgleda lepo.	0.72
	Osoblje je uredno i čisto.	0.69
	Imaju puno mesta za parking.	0.68
	Toaleti su čisti.	0.60
	Zgrade su atraktivne.	0.58

	Nikada ne ostaju bez jela iz menija.	0.55
	Nikada nisu otvoreni do kasno.	0.53
	Osećam se prijatno pored drugih gostiju.	0.52
3. faktor	**Okruženje**	
	Možeš biti brzo uslužen.	0.77
	Zaposleni su ljubazni.	0.65
	Mesto je uvek uredno i čisto.	0.58
	Dobio sam pravu vrednost za svoj novac.	0.57
4. faktor	**Meni**	
	Povremeno imaju nova jela.	0.77
	Često imaju specijalnu ponudu uz popust.	0.74
	Meni pruža širok izbor.	0.59
	Postoji mnogo odgovarajućih lokacija.	0.48

Kada se dobijeni koeficijent za prvu karakteristiku (0,78) podigne na kvadrat, dobija se da je oko 61% 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“.

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 upotrebljene za dodatnu analizu podataka. Iako ova četiri faktora ne pokrivaju sve što su pokrivale 23 karakteristike, ipak pokrivaju veći deo varijacija.

Primer: Auto-dileri

140 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.	0.90
Diler kojem se može verovati.	0.86
Potrude se da shvate moje potrebe.	0.79
Objasne uslove prodaje.	0.76
Prodavci su ljubazni i kulturni.	0.75
Nisu izbegavali moja pitanja.	0.59
Dobro servisno odeljenje.	0.57
2. Izbor robe	
Mnogi modeli ili tipovi su na zalihama.	0.90
Mnogo boja i opcija.	0.87
Otvoreno uveče ili subotama radi servisa.	0.63
3. Uslovi prodaje	
Dobre kamate.	0.83
Dobra cena za razmenu.	0.73
Cene su vrlo konkurentne.	0.66
4. Objekti	
Dovoljno prostora za parking.	0.75
Atraktivna izložbena sala.	0.69
5. Karakteristike vozila.	
Poređenja sa drugim markama.	0.79
Više od jedne marke vozila.	0.64
Ukazano na mnoge osobine vozila.	0.61
6. Nema pritiska	
Nisu navalentni.	0.90
Zovu kasnije da provere da li sam zadovoljan.	-0.48

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.

Nekada je jedno od pitanja u anketi ponuđeno da se da „opšta ocena“ o posmatranoj pojavi pored svih ostalih pojedinačnih karakteristika i onda se ta varijabla takođe uključuje u faktorsku analizu. U tim slučajevima često se dešava da takva varijabla nema visoko faktorsko opterećenje ni kod jednog faktora nego je podjednako „raspršena“ na više faktora sa malim faktorskim opterećenjem.

Ponekad se dešava da gotovo sve varijable imaju visoko faktorsko opterećenje za prvi faktor. Taj slučaj se javlja, na primer, kada respondenti treba da ocene veliki broj performansi nekog proizvoda koji je za njih zapravo nov i nedovoljno poznat. Tada se dešava da respondenti daju odgovore na osnovu nekog opšteg utiska. Ova pojava se naziva „halo efekat“ i postoji više načina da se on izbegne. Jedno rešenje je da standardizuju odgovori za svakog respondenta posebno, a drugi je eliminisanje halo efekta uz pomoć parcijalne korelacije. Više o ovim tehnikama može se naći u Myers i Mullet (2003).

Izbor varijabli za analizu i veličina uzorka

Bez obzira koji je krajnji cilj faktorske analize, mora se voditi računa o tome kakve se varijable koriste u postupku.

Izbor varijabli i njihove karakteristike imaju direktnog uticaja na konačni ishod analize. Na primer, ako se istražuje imidž neke prodavnice, i ako se u analizu ne uključi ni jedna varijabla koja je vezana za osoblje koje radi u prodavnici, onda faktorska analiza neće moći da identifikuje tu dimenziju. Prema tome, veoma je važno uključiti varijable koje pokrivaju sve važne dimenzije određenog predmeta istraživanja.

Mora se imati u vidu da će faktorska analiza uvek, bez obzira na to kakvi su podaci, kao rezultat izračunati faktore. Zbog toga treba voditi računa da se ne desi slučaj GIGO (garbage in-garbage out), odnosno da će se na osnovu beskorisnih ulaznih podataka dobiti i beskorisni faktori. Ukoliko analitičar misli da će ubacivanjem ogromnog broja varijabli faktorska analiza biti u stanju da tu šumu „raščisti“ i ponudi smisleno rešenje onda je on na pogrešnom putu. Kvalitet dobijenih faktora je u direktnoj vezi sa konceptualnim značajem varijabli uključenih u analizu.

Prilikom odabira varijabli dva pitanja se moraju postaviti: „Koji tip varijabli može da se analizira?“ i „Koliko varijabli treda ukupno da bude?“.

Kad je u pitanju tip varijabli treba imati na umu da je za faktorsku analizu bitno da može da se izračuna koeficijent korelacije. Metričke varijable se mogu lako izmeriti, dok su nemetričke problematične jer ne može da se koristi isti tip korelacije kao kod metričkih. Iako postoje specijalni metodi za izračunavanje korelacije između nemetričkih varijabli, najracionalniji pristup jeste da se takve varijable izbegavaju. Ako već nemetrička varijabla mora da bude uvrštena, jedan pristup je da se izračunaju kodirane varijable (dummy variables) koje su kodirane sa 0, 1 itd. i koje na taj način reprezentuju nemetričke varijable. Ako su sve varijable u bazi kodirane, onda je bolje koristiti posebnu vrstu faktorske analize, kao što je Boolean analiza.

Analitičar treba da se trudi da smanji broj varijabli ali takođe i da zadrži razuman broj varijabli po faktoru. Ako analiza ima za cilj da objasni određenu strukturu, potrebno je uključiti nekoliko varijabli koje mogu da reprezentuju određeni faktor – barem pet. Značaj faktorske analize je u pronalaženju šablona unutar grupe varijabli i mala je korist od faktora koji se zasniva na jednoj varijabli.

Što se tiče veličine uzorka, nije preporučljivo analizirati uzorak koji ima manje od 50 jedinica i poželjno je da uzorak ima barem 100 jedinica. Generalno je pravilo da postoji barem pet puta više jedinica nego što ima varijabli u bazi, a najbolje bi bilo da je taj odnos 10:1. U slučajevima kada je taj odnos manji od 5:1, potrebno je rezultate objašnjavati sa velikom rezervom.

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 pojedinačne varijable i te varijacije 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.

Varijable koje imaju malo zajedničkog sa drugim varijablama i imaju malo faktorsko opterećenje (manje od 0,30) prema svim zajedničkim faktorima nazivaju se često „nezavisnim varijablama". Često se dešava da se previdi važnost takvih varijabli i da se one isključe iz interpretacije rezultata što može da bude velika greška.

Vrste faktorske analize

U osnovi postoje dva pristupa u otkrivanju faktora: preko analize glavnih komponenti i faktorska analiza u užem smislu (common factor analysis). Ukratko, analiza glavnih komponenti identifikuje sve izvore varijacija u skupu varijabli, uključujući zajedničke i specifične faktore. Faktorska analiza pokušava da otkrije i objasni samo zajedničke varijacije koji su zastupljene kod dve ili više varijabli.

Kod analize glavnih komponenti 40 varijabli koje se analiziraju biće zamenjene sa 40 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 u kojoj je cilj dobiti minimalni broj faktora koji imaju maksimalni udeo u ukupnoj varijansi originalnih varijabli.

Faktorska analiza u užem smislu (common factor analysis) služi da se podstakne razumevanje i značenje posmatrane pojave. Osnovni cilj faktorske analize jeste da se smanji broj povezanih, preklapajućih varijabli na manji broj nepovezanih komponenti koje bi se mogle efikasnije koristiti u daljoj analizi.

Pošto je faktorska analiza u užem smislu složenija, postoji tendencija u svetu da se mnogo više primenjuje analiza preko glavnih komponenti. U svakom slučaju, empirijski rezultati pokazuju da se rezultati dve vrste analize značajno podudaraju ako broj varijabli prelazi 30 ili komunalitet prelazi 0,60 za veći broj varijabli.

Postoji još jedna podela, i to na R faktorsku analizu i Q faktorsku analizu. R faktorska analiza podrazume otkrivanje latentnih dimenzija u skupu varijabli, odnosno služi za redukciju broja varijabli u modelu.

Q faktorska analiza ima isti zadatak kao i klaster analiza, a to je da grupiše jedinice posmatranja prema svojoj sličnosti u grupe ili klase. Razlika između klaster analize i Q analize je u tome što se u klaster analizi posmatraju stvarne udaljenosti između jedinica posmatranja i spajaju se najbliži parovi, dok kod se kod Q analize posmatraju slične strukture kovarijansi.

Na primer, u tabeli se vide četiri jedinice posmatranja i tri varijable.

Jedinica posmatranja	Varijable		
	V_1	V_2	V_3
A	8	8	12
B	9	9	8
C	4	4	6
D	3	3	2

Claster analiza bi zbog blizine tačaka u prostoru zajedno grupisala jedinice A i B u jednu grupu a C i D u drugu, dok bi Q faktorska analiza grupisala zajedno jedinice A i C u jednu a B i D u drugu grupu jer se pomeraju zajedno.

Model faktorske analize

Razlika između analize glavnih komponenti i faktorske analize je u tome što se faktorska analiza zasniva na matematičkom modelu sa faktorima koji su dobijeni kao standardizovane glavne komponente. Početak razvoja faktorske analize se vezuje za radove Čarlsa Spirmana (Charles Spearman), s početka XX veka.

Opšti faktorski model ima sledeći oblik:

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

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:

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

pri čemu $a_{i1}^2 + a_{i2}^2 + \ldots + a_{im}^2$ predstavlja komunalitet varijable X_i (deo varijanse koji je povezan sa zajedničkim faktorima), a *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:

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

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. Evaluacija i eventualno redefinisanje modela.
5. Interpretacija zajedničkih faktora, uključujući i izbor adekvatnog naziva.
6. Izračunavanje faktor skorova, da bi svaki zajednički faktor bio predstavljen jednim, vaganim indeksim brojem.

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 korelacije 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 nekih posebno mora tražiti ova tabela. Ona može da bude korisna pri razumevanju i davanju imena faktorima, posebno kod manje važnih faktora.

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. Preporučuje se da se vrednosti sa nominalne skale analiziraju jedino ako varijabla ima samo dva modaliteta (na primer, obeležje je „vlasništvo automobila“, a modaliteti su „ima“ i „nema“ odnosno „1“ i „0“).

Analitičar mora da se uveri da u korelacionoj matrici ima dovoljno visokih koeficijenata korelacije da bi imalo smisla primeniti faktorsku analizu. Ako su svi koeficijenti niski, ili su svi jednaki (što znači da nije moguće grupisati varijable), pitanje je da li treba raditi faktorsku analizu. Ako ne postoji značajan broj koeficijenata korelacije koji su veći od 0,30, onda faktorsku analizu ne treba primenjivati.

Korelacija između varijabli se može analizirati i preko parcijalnih koeficijenata korelacije između varijabli. Parcijalna korelacija je korelacija koja je neobjašnjena kada se uzmu u obzir uticaji ostalih varijabli. Ako postoje značajni faktori u strukturi podataka, onda bi parcijalni koeficijenti trebali biti mali, jer se varijabla može objasniti preko učešća varijabli u faktoru. Ako su parcijalni koeficijenti veliki, onda ne treba raditi faktorsku analizu.

Bartlett-ov test sferičnosti je još jedan način da se analizira korelaciona matrica. Ovaj test analizira postojanje statističke značajnosti odnosno da li postoji korelacija barem između nekih varijabli. Treba imati u vidu da sa povećanjem uzorka Bartlett-ov test postaje sve osetljiviji na otkrivanje korelacije.

MSA (measure of sampling adequacy) je još jedan način da se kvantifikuje stepen korelacije između varijabli i opravdanost faktorske analize. Indeks se kreće u granicama od 0 do 1. Što je MSA bliži jedinici, to je lakše predvideti određen varijablu uz pomoć ostalih varijabli. MSA se posmatra prema sledećoj skali:

- preko 0,80 – vrlo jaka korelacija
- između 0,70 i 0,80 – jaka
- između 0,60 i 0,70 – srednja
- između 0,50 i 0,60 – slaba
- ispod 0,50 – neprihvatljiva.

MSA može da se poveća u sledećim slučajevima:

- povećanjem veličine uzorka
- porastom prosečne korelacije
- povećanjem broja varijabli
- smanjivanjem broja faktora.

Opšti MSA mora da bude iznad 0,50 pre nego što se primeni faktorska analiza. Ako opšti MSA padne ispod 0,50 onda specifične MSA vrednosti varijabli mogu da ukažu na one varijable koje treba isključiti iz analize.

Pošto je moguće izračunati specifične MSA za svaku varijablu posebno, moguće je isključiti varijablu sa najnižim specifičnim MSA i onda ponovo uraditi faktorsku analizu. Ovaj postupak potrebno je ponavljati sve dok sve preostale varijable imaju MSA iznad 0,50.

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.

Mnogi statistički programi nude opciju da se umesto faktorske analize na osnovu glavnih komponenti uradi analiza tako da se izvuku samo zajednički faktori a ne svi koji su mogući.

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:

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

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

...

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

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

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

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

...

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

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

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

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

...

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

gde je e_i linearna kombinacija ostalih, izostavljenih glavnih komponenti, od Z_{m+1} do Z_p. Sada treba transformisati preostale glavne komponente da imaju jediničnu varijansu. Za dobijanje faktorskih jednačina potrebno je Z_i podeliti sa standardnom devijacijom, $\sqrt{\lambda}$, koja je kvadratni koren odgovarajuće ajgenvrednosti u korelacionoj matrici:

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

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

...

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

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:

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

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

...

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

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:

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

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

...

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

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

Primer: Rotacija faktora na primeru voćnih sokova

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

Najveća opterećenja prve komponente su 0,93; 0,92; 0,90 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 voćnih sokova) kod kojih je opterećenje veliko, pa to čini interpretaciju komplikovanom. Takođe, kod druge glavne komponente, najveće opterećenje je 0,35, a kod treće -0,30. 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 (odatle i naziv „varimaks“, što za posledicu ima nekoliko visoko varijabilnih opterećenja i mnogo malih, čak negativnih).

Tabela: Faktorsko opterećenje pre i posle rotacije

Osobine voćnih sokova	Glavne komponente pre rotacije				Faktori posle Varimax rotacije				Komunalitet
	1	2	3	4	1	2	3	4	
1.Prijatan ukus	0.84	-0.001	-0.29	0.14	-0.62	0.38	0.36	0.34	0.76
2.Iskričav ukus	0.93	-0.02	-0.02	-0.1	0.48	0.43	-0.53	0.38	0.84
3.Zreo ukus	0.76	-0.11	-0.28	0.1	-0.70	0.26	0.38	0.36	0.83

4.Bogat ukus	0.91	0.16	-0.001	-0.11	0.45	-0.53	0.54	0.29	0.86
5.Zadovoljavajući ukus	0.78	-0.002	-0.20	0.21	-0.74	0.38	0.30	0.32	0.88
6.Živ ukus	0.93	0.12	-0.02	-0.16	0.50	0.43	-0.59	0.35	0.91
7.Ukus pravog voća	0.90	-0.04	0.04	-0.21	0.42	0.38	-0.64	0.37	0.84
8.Dubok, originalni ukus	0.78	0.35	0.11	0.16	0.31	-0.74	0.27	0.22	0.78
9.Ukus tek iscеđenog voća	0.85	-0.28	0.24	-0.09	0.23	0.24	0.52	-0.62	0.81
10.Topao ukus	0.86	0.25	0.22	0.17	0.28	-0.75	0.33	0.39	0.86
11.Čist i jasan ukus	0.89	0.11	-0.05	0.10	0.52	-0.55	0.36	0.36	0.79
12.Sladak ukus	0.86	-0.29	0.04	0.27	0.43	0.28	0.16	-0.67	0.77
13.Svež ukus	0.84	-0.27	0.19	0.12	0.33	0.32	0.36	-0.70	0.81
14.Ukupan utisak	0.92	0.04	0.08	-0.23	0.38	0.43	-0.65	0.34	0.85
Objašnjena varijansa u %	75.2	2.4	3.1	2.5	23.6	20.2	21.3	19.0	-
Kumulativno varijansa u %	75.2	77.6	80.7	83.2	23.6	43.8	65.1	84.1	-

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 75,2% od ukupne varijacije, sledeća 2,4% itd. Nakon rotacije postignuta je ravnomernost u objašnjenoj varijansi između faktora (od 23,6% do 19,0%) 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.

Poslednja kolona tabele pokazuje komunalitet svake varijable. Komunalitet je proporcija varijanse varijable koja je zajednička sa svim ostalim varijablama zajedno. Izračunava se tako što se saberu kvadrati faktorskih opterećenja varijable. Na primer, za varijablu „Prijatan ukus“, komunalitet iznosi:

$(-0{,}62)^2+0{,}38^2+0{,}36^2+0{,}32^2=0{,}76.$

To pokazuje da je 76% od ukupne varijacije varijable „Prijatan ukus“ obuhvaćeno sa četiri zajednička faktora. Isto tako, 24% varijacija se odnosi na specifičnost same varijable plus određeni iznos greške u merenju (*e*).

Za 14 varijabli voćnih sokova komunalitet se kreće od 76% do 91%. To znači da su prilično visoki i da mogu da obuhvat barem jedan faktor i da neke 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.

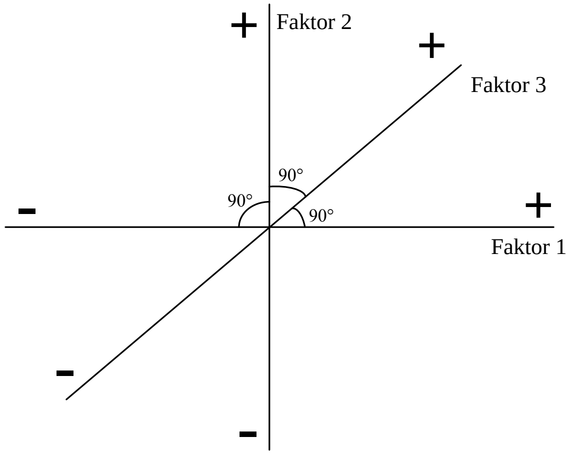
Koliko faktora treba rotirati? Iako glavnih komponenti ima onoliko koliko ima i varijabli, najveći procenat varijacija podataka je objašnjen sa svega nekoliko prvih komponenti. Zbog toga su u primeru sa voćnim sokovima odabrane četiri komponente za rotaciju. Ostale komponente bi samo doprinele konfuziji i težoj 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 će nastavku biti dat njihov detaljniji opis.

Ortogonalna i kosa rotacija

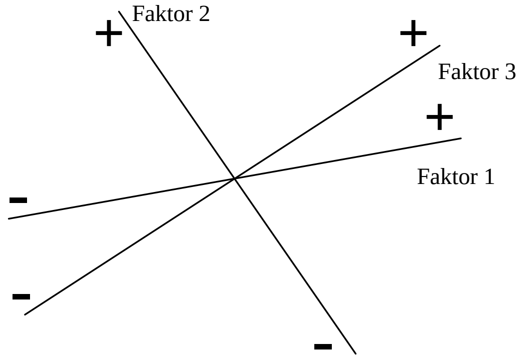
Kod ortogonalne rotacije, ose koje predstavljaju faktore ostaju pod pravim uglom i pre i posle rotacije. Kao posledica toga, faktori su uvek u potpunosti nepovezani. Ovo je tradicionalni pristup koji se prvi pojavio.

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) jer više ne važi pravilo o pravouglom odnosu linija faktora. 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 osa ne prođe kroz bilo koju varijablu koja ima najveće faktorsko opterećenje za svaki ortogonalni faktor.

Kosa rotacija može da pruži jasnije razumevanje strukture faktora. Ukoliko ortogonalna rotacija nije dovoljno ravnomerno rasporedila faktorska opterećenja na faktore, može se pribeći 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. Ipak kosa rotacija neće imati uticaja na prognostičku moć regresionog modela. Ukoliko se za regresionu analizu koriste ortogonalni faktori, onda je problem multikolinearnosti potpuno rešen. Što je veći broj ortogonalnih varijabli, regresioni model je efikasniji.

Pošto rotirani faktor u kosoj rotaciji više nisu pod uglom od 90 stepeni, oni se mogu tretirati kao nove originalne varijable na kojima može da se uradi drugostepena faktorska analiza.

Metod rotacije koji se najčešće koristi je takozvana varimax rotacija. Ona se bazira na pretpostavci da razumljivost datog faktora može da se izmeri varijansom faktorskih opterećenja. Ako je ta varijansa velika onda vrednosti faktorskih opterećenja teže da budu ili blizu nule ili blizu jedinice. Varimax rotacija maksimizira sumu tih varijasni za sve faktore. Ova rotacija može da se radi bez ili sa prethodnom normalizacijom faktorskih opterećenja (Kaiser normalizacija).

Kriterijumi za određivanje broja faktora

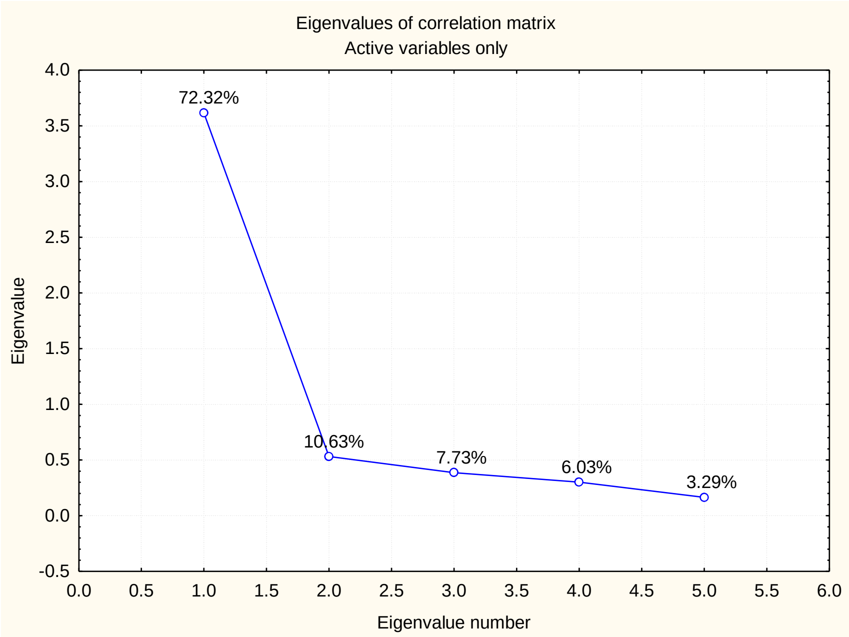
U odlučivanju koliko će faktora biti uzeto u obzir, analitičar mora da kombinuje konceptualna znanja (Koliko faktora treba da bude u datoj strukturi?) sa empirijskim dokazima (Koji je racionalan broj faktora koji se mogu objasniti?). Analitičar polazi od unapred postavljenih kriterijuma, kao što je opšti broj faktora plus opšti prag praktične značajnosti (procenat objašnjenosti varijanse koji je unapred postavljen). Ovi kriterijumi se kombinuju sa empirijskim rezultatima. Tehnika za tačno određivanje broja faktora nije razvijena.

Najčešči metod koji se primenjuje jeste kriterijum latentnog korena (latent root criterion). Prema ovom kriterijumu u obzir se uzimaju samo oni faktori koji imaju ajgenvrednost veću od 1. Faktori koji imaju manju ajgenvrednost od 1 se smatraju nebitnim jer objašnjavaju manje varijabiliteta nego što ga objašnjavaju same varijable. Ovaj metod je najbolji kada je broj varijabli između 20 i 50. Ako ima manje od 20 varijabli, postoji tendencija da se izabere premalo faktora, a ako je broj varijabli veći od 50, tendencija je da se izabere previše faktora.

A priori kriterijum polazi od unapred definisanog broj faktora koji se želi i računaru se prosto da instrukcija da se dati broj faktora izvuče.

Kriterijum procenta objašnjene ukupne varijanse. Polazi se od toga da se odredi procenat ukupne varijanse za koji se želi da bude objašnjen i kada se taj procenat dostigne, tada se utvrdi koliko je faktora potrebno da bi se to dostiglo. Ne postoji čvrsto pravilo koji je to procenat objašnjene ukupne varijanse dovoljan, nego se ide od slučaja do slučaja. U prirodnim naukama obično se zahteva veći procenat (barem 95%), dok kod društvenih nauka, gde je informacija manje precizna, često se analitičari zadovoljavaju i sa 60% od ukupno objašnjene ukupne varijanse.

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 jedan takav grafikon 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

Prilikom konačnog izbora faktora treba voditi računa o tome da broj faktora bude adekvatan. Iako su faktori nezavisni, negativne posledice se javljaju i kad je izabrano previše i kad je izabrano premalo faktora. Ako je izabrano premalo faktora, onda se ne objašnjava prava struktura i važne dimenzije neće biti otkrivene. Ako se zadrži preveliki broj faktora, interpretacija postaje komplikovana kada se rezultati rotiraju. „Po analogiji, izbor broja faktora je nešto kao fokusiranje mikroskopa.“ (Hair, Black, Babin, & Anderson, 2010).

Evaluacija i eventualno redefinisanje modela

Analitičar mora da evaluira dobijeno rešenje. Ukoliko rešenje nije adekvatno ili dovoljno zadovoljavajuće, moguće je da se javi potreba za redefinisanjem celog modela i to upotrebom sledećih koraka:

- isključivanje jedne ili više varijabli iz analize
- upotreba drugačijeg metoda za rotiranje faktora radi bolje interpretacije
- izvlačenje drugačijeg broja faktora u analizi
- upotreba drugačijeg modela za definisanje faktora putem deljenja varijanse.

Potrebno je videti da li su faktorska opterećenja dovoljno značajna. Faktorska opterećenja je potrebno posmatrati na sledeći način:

- Ako su opterećenja u intervalu od ±0,30 do ±0,40 onda oni ispunjavaju minimalne zahteve za učešće u interpretaciji date strukture podataka.
- Opterećenja preko ±0,50 se smatraju praktično signifikantnim.
- Opterećenja preko ±0,70 se smatraju indikativnim za jednu dobro definisanu strukturu i oni su pravi cilj faktorske analize.

Pored navedene skale, postoji i kriterijum za statističku značajnost faktorskih opterećenja uz verovatnoću 95%, odnosno koliko treba da bude velik uzorak da bi se određeni nivo faktorskog opterećenja smatrao značajnim (tabela).

Tabela: Identifikacija statistički značajnih faktorskih opterećenja na osnovu veličine uzorka

Faktorsko opterećenje	Potrebna veličina uzorka da bi se postigla značajnost
0,30	350
0,35	250
0,40	200
0,45	150
0,50	120
0,55	100
0,60	85
0,65	70
0,70	60
0,75	50

Izvor: Hair et al. (Multivariate Data Analysis - A Global Perspective, 2010)

Broj varijabli koje se posmatraju takođe je bitan za donošenje odluke o tome koja su faktorska opterećenja značajna. Kako se broj analiziranih varijabli povećava, prihvatljivi nivo statističke značajnosti opada. Prilagođavanje broju varijabli postaje sve važnije kako se analiza pomera od prvog ka poslednjem faktoru.

Jednom kada su definisana sva statistički značajna opterećenja, potrebno je potražiti varijable koje nisu adekvatno zastupljene dobijenim faktorskim rešenjem. Prvo se potraže sve one varijable koje nemaju ni jedno značajno faktorsko opterećenje. Drugi pristup je da se ispita komunalitet svake varijable, koji reprezentuje zapravo količinu varijanse koja je obuhvaćena faktorskim rešenjem za svaku varijablu. Na primer, analitičar može da odredi da barem jedna polovina varijanse svake varijable mora da bude uzeta u obzir. Prema tome, ona varijabla koja ima komunalitet manji od 0,50 nema dovoljno dobro objašnjenje u modelu.

Kada su istražena faktorska opterećenja i komunalitet, mogu da se jave sledeći problemi:

- varijabla nema značajno faktorsko opterećenje
- komunalitet varijable je nizak, bez obzira što je faktorsko opterećenje značajno
- varijabla ima unakrsno faktorsko opterećenje, odnosno ista varijabla ima značajna faktorska opterećenja za više faktora (cross-loadings).

Potrebno je preduzeti određene mere, koje mogu da se kombinuju (Hair, Black, Babin, & Anderson, 2010):

- Ignorisati problematične varijable i interpretirati faktore takve kakvi su. Ovo je opravdano ako je cilj redukcija podataka, ali mora se imati na umu da su određene varijable loše reprezentovane u faktorskoj strukturi.
- Razmisliti o eventualnom brisanju varijable iz analize, što zavisi od opšteg doprinosa u celokupnom istraživanju i u komunalitetu date varijable. Nakon toga se izrađuje novo faktorsko rešenje bez te varijable. Brisanje variable se često radi i u slučaju unakrsnog faktorskog opterećenja.
- Uraditi neku drugu vrstu rotacije, možda i kosu rotaciju ako je do tada korišćena samo ortogonalna rotacija.
- Smanjiti ili povećati broj faktora da bi se videlo da li će onda problematična varijabla biti bolje reprezentovana.
- Promena vrste faktorske analize (analiza glavnih komponenti vs. faktorska analiza u užem smislu) da bi se videlo da li će se značajnije promeniti faktorska struktura.

Interpretacija rotiranih faktora

Kao finalna faza istraživač posmatra faktorska opterećenja nakon rotacije (ako je ona bila uopšte potrebna) i eventualnog redefinisanja modela. Da bi se dodelili adekvantni nazivi faktorima posmatraju se faktorska operećenja za svaku varijablu sa ciljem da se odredi njena uloga i doprinos u definisanje strukture faktora.

Predznaci faktorskih opterećenjsa se interpretiraju kao kod bilo kog drugog koeficijenta korelacije, što znači da su kod pozitivnih faktorskih opterećenja faktor i varijabla pozitivno povezani a u suprotnom slučaju negativno. Kod ortogonalnih rešenja faktori su nezavisni što znači da se negativno ili pozitivno faktorsko opterećenje kod jednog faktora ne povezuje ni na koji način sa drugim faktorima.

U primeru sa voćnim sokovima, 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 „Prijatan ukus“ (-0,63), svi sa negativnim opterećenjem. To zapravo znači da ako je određena vrsta voćnog soka 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“ (0,52), „Živ ukus“ (0,50), „Iskričav ukus“ (0,48) i „Bogat ukus“ (0,45). To zapravo znači da voćni sok koja ima čist, živ, iskričav i bogat ukus obično nema zadovoljavajući, zreo i prijatan ukus.

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 pravog voća“ i faktor 4 „Svežina“.

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, primenuje se grubo pravilo da bude i četiri faktora u modelu.

Izračunata su faktorska opterećenja za četiri faktora i devet varijabli:

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

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

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

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

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

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

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

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

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

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

$$(0{,}90)^2 + (-0{,}03)^2 + (-0{,}34)^2 + (-0{,}02)^2 = 0{,}93$$

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 varijablia 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:

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

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

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

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

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

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

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

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

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

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 zaposlen 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.

Faktor skorovi

Jednom kada je broj originalnih varijabli smanjen na određeni broj faktora, moguće je izračunati faktor skorove. Oni zapravo predstavljaju rezultate za svaku jedinicu posmatranja po pojedinim faktorima. Skup faktor skorova čini novi, redukovani set podataka. Oni predstavljaju visinu individualnih skorova koji su povezani sa visinom faktorskog opterećenja. Odnosno, visoka vrednost varijable u pogledu faktorskog opterećenja će rezultovati i visokim faktor skorom.

U slučaju sa voćnim sokovima, umesto 14 varijabli sada imamo 4 varijable koje na najbolji način odslikavaju stav ispitanika prema voćnim sokovima.

Postoje slučajevi kada faktor skorove nije poželjno izračunavati:

- Kada struktura faktora nije dovoljno jasna i kada se neki faktori ne mogu dobro interpretirati.
- Kada veliki broj varijabli ima mali komunalitet, što znači malu varijaciju zajedničkih faktora.
- Kada je potrebna velika preciznost u analizi. Tada je bolje primeniti neku od multivarijacionih tehnika zavisnosti, ako je moguće.

Upotreba faktorske analize sa drugim multivarijacionim tehnikama

Pošto faktorska analiza obezbeđuje uvid u međusobni odnos varijabli i otkriva skrivenu strukturu podataka, ona predstavlja dobru početnu osnovu za druge multivarijacione tehnike. Faktorska analiza omogućuje analitičaru jasno razumevanje oko toga koje varijable imaju najvažniji uticaj i koliki je njihov broj. Na primer, u zavisnosti od dobijenih rezultata može se uraditi sledeće:

- Varijable koje su visoko korelisane i pripadaju istom faktoru verovatno imaju iste karakteristike kada su u pitanju statistički značajne razlike između grupa u multivarijacionoj analizi varijanse ili u diskriminacionoj analizi.
- Visoko korelisane varijable, unutar istog faktora, imaju uticaj na stepwise proceduru višestruke regresije i diskriminacione analize prilikom dodavanja novih varijabli i povećavanja moći predviđanja modela. Ako je jedna varijabla vezana za neki faktor već uvrštena u model, onda je malo verovatno da će sledeća varijabla koja je vezana za isti faktor biti uvrštena u model jer će moć predviđanja celog modela biti neznatno uvećana. To ne znači da druge varijable istog faktora nisu bitne ili da imaju manji uticaj, ali njihov efekat je već reprezentovan preko prve varijable koja je uvrštena.

Faktorska analiza pruža empirijsku osnovu za evaluaciju strukture varijabli i uticaj te strukture na interpretaciju rezultata dobijenih nekom drugom multivarijacionom tehnikom. Ako je cilj, naprotiv, da se identifikuju odgovarajuće varijable za primenu u nekoj drugoj analizi, onda će neka od tehnika redukcije biti primenjena. Postoje dve opcije:

- Izbor varijable sa najvećih faktorskim opterećenjem kao surogat koji će reprezentovati određeni faktor odnosno dimenziju.

- Zamena originalnog seta varijabli sa potpuno novim, redukovanim setom varijabli koje su kreirane uz pomoć sumirane skale ili faktor skorova.

O sumiranim skalama više informacija može da se nađe u Hair et al. (2010).

Bilo koja od pomenutih opcija kreira nove varijable koje bi, na primer, mogle da se koriste kao nezavisne varijable u diskriminacionoj analizi, kao zavisne varijable u multivarijacionoj analizi varijanse ili kao klaster varijable u klaster analizi. U svakom slučaju, ako je cilj što veća jednostavnost, onda se favorizuje rešenje sa surogat varijablama, ako se želi replikacija u drugim studijama, favorizuje se sumirana skala, a ako je akcenat na ortogonalnosti, onda se biraju faktor skorovi. Sa empirijskog stanovišta postoji velika sličnost između sumirane skale i faktor skorova.

Ipak, vrlo često su rezultati faktorske analize sami sebi cilj, kada je dovoljno da se identifikuje logička kombinacija varijabli i bolje razumevanje njihovog međusobnog odnosa.

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

Otvara 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 ajgenvrednosti:

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 ajgenvrednost.

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 ajgenvrednosti:

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

Literatura

Hair, J., Black, W., Babin, B., & Anderson, R. (2010). *Multivariate Data Analysis - A Global Perspective*. New Jersey: Pearsib.

Manly, B. F. (2005). *Multivariate Statistical Methods - A primer* (3rd Edition izd.). New York: Chapman & Hall/CRC.

Myers, J. H., & Mullet, G. M. (2003). *Managerial Applications of Multivariate Analysis in Marketing*. Chicago: American Marketing Association.