Faktorska analiza

Srša, Sanja

SVEUČILIŠTE U ZAGREBU
PRIRODOSLOVNO–MATEMATIČKI FAKULTET
MATEMATIČKI ODSJEK

Sanja Srša

Faktorska analiza

Diplomski rad

Voditelj rada:
prof. dr. sc. Miljenko Huzak

Zagreb, 2015

Ovaj diplomski rad obranjen je dana _____ pred ispitnim povjerenstvom u sastavu:

1. _____, predsjednik
2. _____, član
3. _____, član

Povjerenstvo je rad ocijenilo ocjenom _____.

Potpisi članova povjerenstva:

1. _____
2. _____
3. _____

Sadržaj

Uvod

Faktorska analiza je statistička disciplina koja otkriva i uspostavlja korelaciju među opaženim slučajnim varijablama. Motivirana je činjenicom da su mjerene varijable korelirane na takav način da se korelacija može rekonstruirati. Faktorska analiza pomaže odabrati manji broj parametara koji rekonstruiraju osnovnu strukturu u sažetijem i jasnijem obliku. Ti parametri se nazivaju faktori. Važno je dobro odabrati varijable koje ćemo uključiti u model, a koje ne jer to jako utječe već na samo prikupljanje podataka pa onda i na rezultate. Cilj faktorske analize je smanjiti broj parametara tako da varijablu odaziva možemo gotovo jednako dobro opisati kao sa većim brojem parametra, ili čak bolje jer uklonimo podudaranja. Ona nastoji da se što više pojava opiše sa što manje varijabli poticaja. Njezina primjena je najučestalija u humanističkim znanostima, na primjer za mjerenje mentalnih sposobnost, potrošačkih ukusa, političke orijentacije, socijalnih razreda i sl., iako se primjenjuje i u prirodnim znanostima. Glavna motivacija za korištenje faktorske analize je u mogućnosti "smislene" interpretacije podataka. Njezina primjena postaje svakim danom sve veća.

U ovom radu bavimo se metodama faktorske analize. Prvo ćemo u preliminarijama navesti oznake i tvrdnje iz drugih područja koje će nam biti potrebne za razumijevanje rada. Nakon toga ćemo detaljno opisati osnovni faktorski model, sa svim danim pretpostavkama. Vidjet ćemo da je faktorski model zapravo određen matricom koeficijenata te faktorima. Postoji puno načina za određivanje faktorskog modela, odnosno koeficijenata modela, koji se međusobno razlikuju s obzirom na dodatne pretpostavke te način procjene. Te načine zovemo faktorskom analizom. Osnovna podjela je na faktorsku analizu glavnih komponenti te faktorsku analizu maksimalne vjerodostojnosti. Unutar svakog od tih područja predstavit ćemo nekoliko modela, od kojih ćemo svaki detaljno obraditi.

Diskutirat ćemo testove značajnosti koji vrijede pod pretpostavkom normalnosti podataka. Dani testovi zapravo ispituju da li na danim podacima ima smisla tražiti faktorski model, odnosno da li uopće postoje zajednički faktori te koliko ih ima. Klasični test za takvo testiranje je χ^2 test. Osim njega, ili u kombinaciji s njim, koriste se Akaikeov informacijski kriterij (AIC) i Schwarzov informacijski kriterij (SIC). Navedeno testiranje demonstrirat ćemo na primjeru. Također ćemo vidjeti da uz pretpostavku normalnosti možemo testirati značajnost svakog koeficijenta.

Nakon objašnjenja kako pronaći broj zajedničkih faktora te odrediti matricu koeficijenata, objasnit ćemo kako procijeniti faktore. Na faktore se može gledati kao na

slučajne varijable, a i kao na fiksne parametre. Sukladno tome postoje dva različita pristupa procjene.

Na samom kraju rada ćemo na konkretnom primjeru demonstrirati sprovođenje faktorske analize.

Svi teoremi koji nemaju referencu preuzeti su iz knjige [1].

Poglavlje I

Preliminarije

U danom radu koristit ćemo naredne oznake:

ρ - rang matrice
F - funkcija distribucije
S - uzoračka kovarijacijska matrica
R^2 - koeficijent determinacije

Neka je X vektor. Tada je X oblika

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix}.$$

Neka je matrica podataka $\mathbb{X}$ dana sa n opservacija, $X_1, X_2, \ldots, X_n$. Tada je $\mathbb{X}$ oblika

$$\mathbb{X} = \begin{bmatrix} X_1^T \\ X_2^T \\ \vdots \\ X_n^T \end{bmatrix}.$$

Matrice sa kojima ćemo raditi često će biti Grammove, stoga nevedimo formalnu definiciju Grammove matrice:

Definicija 0.1. *Neka su $x_1, x_2, \ldots, x_k$ vektori unitarnog prostora U. Gramova matrica je $k \times k$ matrica definirana sa*

$$G(x_1, x_2, \ldots, x_k) = \begin{bmatrix} (x_1|x_1) & (x_1|x_2) & \ldots & (x_1|x_k) \\ (x_2|x_1) & (x_2|x_2) & \ldots & (x_2|x_k) \\ \vdots & \vdots & & \vdots \\ (x_k|x_1) & (x_k|x_2) & \ldots & (x_k|x_k) \end{bmatrix}$$

Kada imamo dvije slučajne varijable koje su zavisne od interesa nam je njihova uvjetna gustoća.

Definicija 0.2. *(vidi [3]) Uvjetna funkcija gustoće slučajne varijable Y za dano $X = x$ jest funkcija $y \mapsto f_{(Y|X)}(y|x)$ definirana na $\mathbb{R}$ sa*

$$f_{(Y|X)}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}, \text{ ako je } f_X(x) > 0.$$

Za $f_X(x) = 0$ funkcija $y \mapsto f_{(Y|X)}(y|x)$ nije definirana.

Nakon što smo definirali uvjetnu gustoću, možemo navesti Bayeseovu formulu koja se vrlo često koristi u statistici. Navest ćemo Bayesovu formulu u terminima vjerojatnosti te u terminima funkcije gustoće.

Teorem 0.1 (Bayesova formula)**.** *(vidi [3]) Neka je $H_i\,(i = 1, 2, \ldots)$ potpun sistem događaja u vjerojatnosnom prostoru $(\Omega, \mathscr{F}, P)$ i $A \in \mathscr{F}$ takav da je $P(A) > 0$. Tada za svako i vrijedi*

$$P(H_i|A) = \frac{P(H_i)P(A|H_i)}{\sum_j P(H_j)P(A|H_j)}.$$

Neka su sada X i Y slučajne varijable. Tada je Bayesova formula u terminima funkcije gustoće

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{\int f_{Y|X}(y|x')f_X(x')dx'}.$$

Neka je sada

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \epsilon$$

opći regresijski model, pri čemu su $\beta_0, \beta_1, \ldots, \beta_k$ parametri modela, $x_1, \ldots, x_k$ varijable poticaja a ϵ slučajna greška.
Za model kažemo da je homoskedastičan ako je varijanca slučajne greške modela konstantna za sve opservacije. Ako model nije homoskedastičan, tj. varijance slučajnih grešaka razlikuju se po opservacijama, tada kažemo da je model heteroskedastičan.
Koliko je neki model dobar ovisi i o tome koliko varijabilnosti objašnjava. To se mjeri koeficijentom determinacije koji se definira kao

$$R^2 = \frac{objašnjena\ varijabilnost}{ukupna\ varijabilnost} = \frac{\sum_{i=1}(y - \hat{y})^2}{\sum_{i=1}(y - \overline{y})^2}$$

pri čemu je y opažena vrijednost a $\hat{y}$ vrijednost dobivena iz regresijskog modela.

Pretpostavimo da nam je normalna razdioba poznata, pa navedimo u nastavku uvjetnu distibuciju normalnih slučajnih vektora.

Teorem 0.2. *Neka je $X = (X_1, X_2)^T$ višedimenzionalni normalni vektor. Tada je uvjetna distribucija od X_1 uz dano X_2, višedimenzionalna normalna sa očekivanjem*

$$\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2)$$

i kovarijacijskom matricom

$$\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^{T}$$

pri čemu je $\mu_i\,(i = 1, 2)$ *vektor očekivanja od* X_i*, a* $\Sigma_{ij}\,(i, j = 1, 2)$ *kovarijacijska matrica kovarijanci svake komponente od* X_i *u odnosu na svaku komponentu od* X_j*.*

Osim normalne distribucije koristit ćemo i Wishartovu distribuciju, čija funkcija gustoće je oblika

$$f(S) = c|\Sigma|^{-n/2}|S|^{1/2(n-p-1)}exp[-\tfrac{1}{2}tr(\Sigma^{-1}S)]$$

pri čemu je

$$c = \frac{n^{n/2}}{2^{np/2}\pi^{p(p-1)/4}\sum_{j=1}^{p}\Gamma(n/2 + (1-j)/2)}$$

gdje je Γ gama funkcija.

Faktorski model se određuje na temelju kovarijacijske matrice, no kada nam ona nije poznata tada umjesto nje koristimo uzoračku kovarijacijsku matricu koja je definirana na sljedeći način:

Definicija 0.3. *Uzoračka kovarijacijska matrica skupa p slučajnih varijabli, na osnovi slučajnog uzorka duljine n, je matrica S čiji* (l, h)*-ti element je uzoračka kovarijanca između l-tog i h-tog stupca matrice podataka* $\mathbb{Y}$*. Matrica S ima elemente*

$$\begin{aligned} s_{lh} &= \sum_{i=1}^{n}\frac{(y_{il} - \overline{y_l})(y_{ih} - \overline{y_h})}{n-1} \\ &= \frac{1}{n-1}[\sum_{i=1}^{n} y_{il}\,y_{ih} - n\overline{y_l}\,\overline{y_h}] \end{aligned}$$

Definirajmo sada S u terminima matrica.
Neka je

$$\overline{Y} = \begin{bmatrix} \overline{y_1} & \overline{y_2} & \dots & \overline{y_p} \\ \overline{y_1} & \overline{y_2} & \dots & \overline{y_p} \\ \vdots & \vdots & & \vdots \\ \overline{y_1} & \overline{y_2} & \dots & \overline{y_p} \end{bmatrix}$$

matrica čiji su stupci uzoračka očekivanja (to jest aritmetičke sredine) vektora $Y_1, Y_2, \dots, Y_p$*, pri čemu je* Y_i *i-ti stupac matrice* $\mathbb{Y}$*. Tada je uzoračka kovarijacijska matrica dana sa*

$$\begin{aligned} S &= \frac{1}{n-1}(\mathbb{Y} - \overline{Y})^T(\mathbb{Y} - \overline{Y}) \\ &= \frac{1}{n-1}\mathbb{X}^T\mathbb{X} \end{aligned}$$

gdje je $\mathbb{X} = \mathbb{Y} - \overline{Y}$*.*

0.1 Derivacije

Neka je X $(n \times 1)$ vektor stupac i neka je $y = f(X)$. Tada vrijedi

$$\frac{\partial y}{\partial X} := \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{bmatrix} y = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix}.$$

Neka je sada $X = [x_{ij}]$ $(n \times m)$ matrica, a $y = f(X)$. Tada vrijedi

$$\frac{\partial y}{\partial X} := \begin{bmatrix} \frac{\partial}{\partial x_{11}} & \frac{\partial}{\partial x_{12}} & \cdots & \frac{\partial}{\partial x_{1m}} \\ \frac{\partial}{\partial x_{21}} & \frac{\partial}{\partial x_{22}} & \cdots & \frac{\partial}{\partial x_{2m}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial}{\partial x_{n1}} & \frac{\partial}{\partial x_{n2}} & \cdots & \frac{\partial}{\partial x_{nm}} \end{bmatrix} y = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{12}} & \cdots & \frac{\partial y}{\partial x_{1m}} \\ \frac{\partial y}{\partial x_{21}} & \frac{\partial y}{\partial x_{22}} & \cdots & \frac{\partial y}{\partial x_{2m}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial y}{\partial x_{n1}} & \frac{\partial y}{\partial x_{n2}} & \cdots & \frac{\partial y}{\partial x_{nm}} \end{bmatrix}.$$

Neka su

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Tada je

$$\frac{\partial Y}{\partial X} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_n}{\partial x_2} \\ \vdots & \vdots & & \vdots \\ \frac{\partial y_1}{\partial x_n} & \frac{\partial y_2}{\partial x_n} & \cdots & \frac{\partial y_n}{\partial x_n} \end{bmatrix}.$$

Neka je X nesingularna matrica i A kvadratna matrica koja ne ovisi o X. Tada vrijedi

$$\frac{\partial}{\partial X} tr(AX^{-1}) = -(X^{-1}AX^{-1})^T.$$

Ako je X nesingularna, tada je

$$\frac{\partial}{\partial X} ln|X| = (X^{-1})^T.$$

Ako elementi nesingularne matrice X ovise o nekoj varijabli y tada vrijedi

$$\begin{aligned} \frac{\partial}{\partial y} ln|X| &= \frac{1}{|X|} \frac{\partial |X|}{\partial y} \\ &= tr(X^{-1} \frac{\partial X}{\partial y}). \end{aligned}$$

Definicija 0.4. *Linearna skalarana funkcija $y = f(X)$ je funkcija oblika*

$$\begin{aligned} y &= a_1x_1 + a_2x_2 + \ldots + a_nx_n \\ &= A^TX \end{aligned}$$

pri čemu su $x_1, x_2, \ldots, x_n$ realni brojevi. Za $X = (x_1, x_2, \ldots, x_n)^T$ tada vrijedi

$$\frac{\partial y}{\partial X} = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = A. \tag{0.1}$$

Definicija 0.5. *Potpuni diferencijal dF funkcije više varijabli $F(x_1, x_2, \ldots, x_n)$ je*

$$dF = \sum_{i=1}^{n} \frac{\partial F}{\partial x_i} dx_i$$

pri čemu su dx_i, $i = 1, 2, \ldots, n$ diferencijali neovisnih varijabli.

U nastavku ćemo navesti leme vezane uz deriviranje koje će nam biti potrebne kasnije.

Lema 0.1. *Neka ja f linearna skalarna funkcija vektora X i neka je dX oznaka za matricu potpunog diferencijala. Ako je $df = tr(c\,dX^T)$, gdje c ovisi o X ali ne ovisi o dX, tada $\partial f/\partial X = c$.*

Lema 0.2. *Neka su X i Y matrice. Tada*

(i) $d(YX) = dY\,X + Y\,dX$

(ii) $dX^{-1} = -X^{-1}\,dX\,X^{-1}$

(iii) $d(tr\,X) = tr(dX)$.

0.2 Glavne komponente

Analiza glavnih komponenata bavi se tumačenjem strukture matrice varijanci i kovarijanci skupa izvornih varijabli pomoću malog broja njihovih linearnih kombinacija. Premda je p ulaznih varijabli odabrano kako bi se opisala varijablnost cijelog sustava, često je velik dio tog varijabiliteta opisan malim brojem k glavnih komponenata ($k < p$). Ako je to ispunjeno, k glavnih komponenata sadrži gotovo jednaku količinu informacija kao p ulaznih varijabli. Glavne komponente ovise samo o matrici varijanci i kovarijanci. Glavne komponente se biraju tako da prva komponenta objašnjava najveći dio varijance, sljedeća komponenta najveći dio od preostale neobjašnjene varijance uz uvjet da je ortogonalna na prethodnu komponentu i tako dalje induktivno dok čitava varijanca nije objašnjena. Za dobivanje glavnih komponenti koriste se svojstveni vektori i svojstvene vrijednosti Grammove matrice.
Navedimo definiciju glavnih komponenti te osnovne teoreme vezane uz njih.

Definicija 0.6. *Neka je* $X = (X_1, X_2, \ldots, X_p)^T$ $(p \times 1)$ *vektor slučajnih varijabli sa očekivanjem* 0 *i kovarijacijskom matricom* $E(XX^T) = \Sigma$. *Tada slučajne varijable* $\zeta_i = \Pi_i^T X \, (i = 1, 2, \ldots, p)$, *pri čemu su* Π_i *jedinični svojstveni vektori od* Σ *koji odgovaraju padajućem nizu svojstvenih vrijednosti od* Σ, *zovemo glavne komponente.*

Teorem 0.3. *Neka su* $\zeta_i = \Pi_i^T X$ *i* $\zeta_j = \Pi_j^T X$ $(i \neq j)$ *dvije linearne kombinacije* p *slučajnih varijabli* $X = (X_1, X_2, \ldots, X_p)^T$, *pri čemu su* Π_i *i* Π_j *jedinični svojstveni vektori matrice* $E(XX^T) = \Sigma$. *Tada vrijedi:*

(i) Π_i *i* Π_j *su ortogonalni vektori ako pripadaju raličitim svojstvenim vrijednostima* λ_i *i* λ_j *i u tom slučaju*

(ii) ζ_i *i* ζ_j *su nekorelirane slučajne varijable takve da je* $var(\zeta_i) = \lambda_i$ $(i = 1, 2, \ldots, p)$.

Teorem 0.4. *Neka je* $\Sigma\Pi = \Pi\Lambda$, *pri čemu je* $\Pi = [\Pi_1, \Pi_2, \ldots, \Pi_p]$ *ortogonalna matrica svojstvenih vektora, a* Λ *dijagonalna matrica svojstvenih vrijednosti* $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$. *Tada vrijedi*

$$\lambda_i = max(\alpha^T \Sigma \alpha) = \Pi_i^T \Sigma \Pi_i \qquad (i = 1, 2, \ldots, p)$$

pri čemu se maksimum traži po svim jediničnim vektorima α *takvima da je* $\alpha^T \Pi_j = 0$ *za* $j = 1, \ldots, i-1$.

0.3 Težinske glavne komponente

Neka je

$$X = \chi + \Delta$$

pri čemu je χ "pravi" dio od X, a Δ greške. Možemo napraviti dekompoziciju matrice kovarijance u dva dijela

$$\Sigma = \Sigma^* + \Psi$$

pri čemu je "pravi" dio linearna kombinacija $r < p$ glavnih komponenti, to jest

$$\chi = \zeta\alpha.$$

Bolje nego dekompoziciju od Σ je napraviti dekompoziciju produkta

$$\Psi^{-1}\Sigma = \Psi^{-1}\Sigma^* + I$$

budući da varijable sa većom (manjom) varijancom greške daju manju (veću) težinu u analizi. Također, takvim množenjem dobivamo da su greške homoskedastične te nekorelirane. Neka je Π $(p \times 1)$ vektor koeficijenata. Želimo maksimizirati funkciju

$$\lambda = \frac{\Pi^T \Sigma \Pi}{\Pi^T \Psi \Pi}$$

po Π uz pretpostavku da je $\Pi^T\Psi\Pi > 0$ za svaki $\Pi \neq 0$. Maksimum se dobiva iz normalnih jednadžbi

$$(\Sigma - \hat{\lambda}\Psi)\hat{\Pi} = 0 \tag{0.2}$$

pri čemu je $\hat{\Pi}$ svojstveni vektor od Σ (u metrici od Ψ) koji pripada najvećoj svojstvenoj vrijednosti $\hat{\lambda}$. Jednadžbu (0.2) možemo također zapisati kao

$$(\Psi^{-1}\Sigma - \hat{\lambda}I)\hat{\Pi} = 0 \tag{0.3}$$

gdje je $\hat{\Pi}$ svojstveni vektor koji pripada najvećoj svojstvenoj vrijednosti težinske matrice $\Psi^{-1}\Sigma$. Sljedeći vektori za koje se postiže maksimum su svojstveni vektori matrice $\Psi^{-1}\Sigma$ koji odgovaraju padajućem nizu preostalih svojstvenih vrijednosti dane matrice, te su ortonormalni na prethodno dobivene vektore.

0.4 Kanonski korelacijski model

Na kanonski korelacijski model može se gledati kao na generaliziranu analizu glavnih komponenti budući da računa korelaciju između dva skupa varijabli promatranih nad istim uzorkom. Dakle, objekt kanonskog korelacijskog modela je proučavanje korelacijske strukture između dva skupa podataka. Pretpostavimo da imamo dva skupa $X_{(1)}$ i $X_{(2)}$ pri čemu prvi sadrži p_1, a drugi p_2 slučajnih varijabli te vrijedi da je $p_1 < p_2$ i $p_1 + p_2 = p$. Slučajne varijable iz $X_{(1)}$ i $X_{(2)}$ se opažaju na istih n opservacija. $(n \times p)$ matricu možemo napisati kao $\mathbb{X} = [\mathbb{X}_{(1)} : \mathbb{X}_{(2)}]$, gdje je $\mathbb{X}_{(1)}$ $(n \times p_1)$ matrica i $\mathbb{X}_{(2)}$ $(n \times p_2)$ matrica, sa kovarijacijskom matricom

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}. \tag{0.4}$$

Teorem 0.5. *Neka je Σ $(p \times p)$ podijeljena kovarijaciska matrica dana sa* (0.4). *Tada su koeficijenti koji maksimiziraju korelaciju između linearnih kombinacija $u = \alpha^T X_{(1)}$ i $v = \beta^T X_{(2)}$ svojstveni vektori sustava jednadžbi:*

$$(\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \lambda^2)\alpha = 0$$
$$(\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} - \mu^2)\beta = 0.$$

U tom slučaju je $\lambda = \mu = \alpha^T\Sigma_{12}\beta$ je maksimalna korelacija.

Poglavlje II

Modeli faktorske analize

1 Osnovni faktorski model

Osnovni faktorski model je oblika

$$Y = \mu + \alpha\Phi + \epsilon \tag{1.1}$$

odnosno

$$X = \alpha\Phi + \epsilon \tag{1.2}$$

pri čemu je $Y = (y_1, y_2, \cdots, y_p)^T$ vektor opažanih slučajnih varijabli, $\mu = EY$, $X = Y - \mu$, a $\Phi = (\phi_1, \phi_2, \cdots, \phi_r)^T$ vektor od $r < p$ neopažanih ili latentnih varijabli koje zovemo faktori. Φ nije zadan, on se procjenjuje iz podataka. α je $(p \times r)$ matrica koeficijenata faktora koji su fiksni, a $\epsilon = (\epsilon_1, \epsilon_2, \cdots, \epsilon_p)^T$ vektor grešaka. ϵ sadrži grešku mjerenja, individualan učinak svake varijable y_j na grešku te grešku uzorkovanja.

Osnovne pretpostavke faktorskog modela su:

(i) $\rho(\alpha) = r < p$

(ii) $E(X \mid \Phi) = \alpha\Phi$

(iii) $E(\epsilon\epsilon^T) = \begin{bmatrix} \sigma_1^2 & & & \mathbf{0} \\ & \sigma_2^2 & & \\ & & \ddots & \\ \mathbf{0} & & & \sigma_p^2 \end{bmatrix}$

(iv) $E(\Phi\epsilon^T) = 0$.

Vidimo da nam iz pretpostavke (iii) slijedi nekoreliranost grešaka, dok nam iz (iv) slijedi da su greške i faktori nekorelirani. Koristiti ćemo sljedeće oznake: $\Sigma = E(XX^T), \Omega = E(\Phi\Phi^T)$ i $\Psi = E(\epsilon\epsilon^T)$. Iako faktori međusobno nisu nužno nekorelirani često pretpostavljamo da je $\Omega = I$. Tada slijedi

$$E(\phi_i\phi_j) = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases}$$

Koristivši $(i) - (iv)$ računamo:

$$\begin{aligned} E(XX^T) = \Sigma &= E(\alpha\Phi + \epsilon)(\alpha\Phi + \epsilon)^T \\ &= E(\alpha\Phi\Phi^T\alpha^T + \alpha\Phi\epsilon^T + \epsilon\Phi^T\alpha^T + \epsilon\epsilon^T) \\ &= \alpha E(\Phi\Phi^T)\alpha^T + \alpha E(\Phi\epsilon^T) + E(\epsilon\Phi^T)\alpha^T + E(\epsilon\epsilon^T) \\ &= \alpha\Omega\alpha^T + E(\epsilon\epsilon^T) \\ &= \Gamma + \Psi \end{aligned} \tag{1.3}$$

gdje je $\Gamma = \alpha\Omega\alpha^T$ kovarijacijska matrica, a $\Psi = E(\epsilon\epsilon^T)$ kovarijacijska matrica grešaka. Γ je Gramova matrica, a Ψ je po pretpostavci modela dijagonalna matrica sa dijagonalnim elementima $\sigma_i^2 > 0\,(i = 1, 2, \ldots, p)$. Koristeći (iii) i (iv) iz jednadžbe (1.2) dobijemo

$$\begin{aligned} E(X\Phi^T) &= E(\alpha\Phi\Phi^T + \epsilon\Phi^T) \\ &= \alpha E(\Phi\Phi^T) + E(\epsilon\Phi^T) \\ &= \alpha\Omega. \end{aligned}$$

U slučaju $\Omega = I$ vrijedi $E(X\Phi^T) = \alpha$.

Proučimo sada slučaj kada je X višedimenzionalni normalni slučajni vektor. U tom slučaju drugi moment sadrži svu informaciju vezanu za faktorski model. Također slijedi da je faktorski model (1.2) linearan. Naime, neka je $\Phi \sim N(0, I)$. Tada prema teoremu 0.2 za uvjetnu distibuciju od X vrijedi

$$\begin{aligned} X|\Phi &\sim N[\alpha\Phi, (\Sigma - \alpha\alpha^T)] \\ &\sim N[\alpha\Phi, \Psi] \end{aligned}$$

gdje zbog dijagonalnosti od Ψ slijedi uvjetna nezavisnost varijabli uvjetno na faktore Φ. Zajednički faktori Φ reproduciraju čitavu kovarijancu (korelaciju), ali samo dio varijance. Također je moguće odrediti procjenitelje maksimalne vjerodostojnosti (ML procjenitelje) za α, Φ i Ψ koji imaju asimptotsku efikasnost pa možemo računati pouzdane intervale za ML procjenitelje od α. Dakle, uvijek je moguće odabrati α tako da reprezentira korelaciju između X i Φ (uz pretpostavku da faktori postoje) pri čemu je Σ kovarijacijska (korelacijska) matrica.

Općenito, postavlja se pitanje pod kojim uvjetima (i ako) se može naći dekompozicija od Σ kao u jednadžbi (1.3)? Pod pretpostavkom da takva dekompozicija postoji, pod kojim uvjetima je moguće naći jedinstvene koeficijente α? Uzevši $1 < r < p$ zajedničkih faktora može se pokazati da općenito nije moguće odrediti α i Φ na jedinstven način. Pomoću rotacija dobivamo beskonačno mnogo rješenja. Pretpostavimo da postoji $1 < r < p$ zajedničkih faktora takvih da su $\Gamma = \alpha\Omega\alpha^T$ i Ψ Gramova, odnosno dijagonalna matrica. Tada kovarijacijska matrica Σ ima $\binom{p}{2} + p = \frac{1}{2}p(p+1)$ različitih elemenata, što je jednako broju normalnih jednadžbi koje treba riješiti. Broj rješenja je beskonačan. Pokažimo to. Zbog pozitivne definitnosti od Ω, postoji nesingularna $(r \times r)$ matrica B takva da $\Omega = B^T B$ i

$$\begin{aligned}\Sigma &= \alpha\Omega\alpha^T + \Psi \\ &= \alpha(B^T B)\alpha^T + \Psi \\ &= (\alpha B^T)(\alpha B^T)^T + \Psi \\ &= \alpha^*\alpha^{*T} + \Psi. \end{aligned} \tag{1.4}$$

Obje faktorizacije (1.3) i (1.4) od Σ imaju iste greške Ψ, pa moraju imati jednako valjana rješenja. Dakle, supstitucija $\alpha^* = \alpha C$ i $\Omega^* = C^{-1}\Omega(C^T)^{-1}$ također zadovoljava faktorski model koji je jednak jednadžbi (1.3). Vidimo da možemo načiniti beskonačno mnogo transformacija, to jest dobiti beskonačno mnogo rješenja koja su jednako valjana. Da bi došli do jedinstvenoga rješenja nužno je uvesti restrikciju.

Restrikcija se najčešće odnosi na Ω, kovarijacijsku matricu faktora. Jednostavna i često korištena restrikcija jest $\Omega = I$ čime se faktori definiraju kao ortogonalni jedinični vektori.
Tada je

$$\Sigma = \alpha\alpha^T + \Psi \tag{1.5}$$

i broj slobodnih parametara m u danoj jednadžbi je $pr + p$ (r broj faktora), broj nepoznatih parametara u α i Ψ, umanjen za broj ne-dijagonalnih elemenata matrice Ω koji su jednaki 0. Zbog simetričnosti od Ω broj parametara u Ω jednakih 0 je $1/2r(r-1)$. Dakle,

$$\begin{aligned} m &= (pr + p) - 1/2(r^2 - r) \\ &= p(r+1) - 1/2r(r-1) \end{aligned}$$

gdje su stupci od α po pretpostavci ortogonalni. Broj stupnjeva slobode d je broj jednadžbi impliciranih relacijom (1.5), što je jednako broju različitih elemenata u Σ umanjenom za broj slobodnih parametar m. Imamo

$$\begin{aligned} d &= 1/2p(p+1) - [pr + p - 1/2(r^2 - r)] \\ &= 1/2[(p-r)^2 - (p+r)] \end{aligned} \tag{1.6}$$

pri čemu d da bi imao smisla mora biti strogo pozitivan. Zbog toga imamo ograničenje na maksimalan broj faktora r s obzirom na p. Uočimo da se u jednadžbi (1.6) pretpostavlja da su normalne jednadžbe linearno nezavisne, što vrijedi samo ako je Σ nesingularna i $\rho(\alpha) = r$. Za $d > 0$ imamo više jednadžbi nego slobodnih parametara, pa r glavnih faktora postoji samo ako uvedemo neka dodatna ograničenja na elemente od Σ. Kada je problem egzistencije riješen, ostaje pitanje jedinstvenosti faktora.

Čak i uz ograničenje $\Omega = I$ faktorski model još uvijek nije jedinstveno određen zbog mogućih rotacija α. Problem nejedinstvenosti se obično rješava na način da se fiksira koordinatni sustav. Način ograničenja definira tip faktorskog modela od kojih ćemo neke obraditi u narednim poglavljima. Za $r = 1$, α je uvijek jedinstveno određen. U sljedećem teoremu izreći ćemo nužne i dovoljne uvjete postojanja faktora.

Teorem 1.1 (Reiersol,1950.)**.** *Neka je Σ $(p \times p)$ kovarijacijska matrica. Nužan i dovoljan uvjet za postojanje r zajedničkih faktora je postojanje nenegativne dijagonalne matrice Ψ, takve da je $\Sigma - \Psi = \alpha\alpha^T$ pozitivno definitna matrica ranga r.*

Teorem 1.2. *Neka $X = \alpha\Phi + \epsilon$ sadrži strukturu takvu da je $r = r_0 < p$, gdje je r_0 minimalni rang od α i Ψ nesingularna matrica. Tada dani faktorski model ima beskonačno mnogo ekvivalentnih struktura za $r = r_0 + 1$.*

Prema teoremu 1.2 slijedi da za određeni r postoji beskonačno mnogo nenegativnih nesingularnih dijagonalnoh matrica Ψ koje zadovoljavaju faktorski model (1.3).

Faktorska analiza se može gledati i u općenitijem obliku, preko vjerojatnosnih distribucija. Početna točka razmatranja je relacija

$$f(X) = \int_R h(\Phi)g(X|\Phi)d\,\Phi$$

gdje su $f(X)$ i $h(\Phi)$ gustoće od X i Φ, a $g(X|\Phi)$ je uvjetna gustoća od X uz dano Φ. R je parametarski prostor od Φ. Koristeći Bayesovu formulu (teorem 0.1) dobivamo da je uvjetna gustoća od Φ uz dano X

$$h(\Phi|X) = h(\Phi)g(X|\Phi)/f(X).$$

$f(X)$ ne određuje $g(X|\Phi)$ i $h(\Phi)$ jedinstveno, pa su naredne pretpostavke nužne. Po pretpostavci su komponente od X uvjetno nezavisne uz dano Φ, to jest uvjetnu distribuciju od X uz dano Φ možemo napisati kao

$$g(X|\Phi) = \prod_{i=1}^{n} g(X_i|\Phi).$$

pa je

$$h(\Phi|X) = \frac{h(\Phi)\prod_{i=1}^{n} g(X_i|\Phi)}{f(X)}.$$

Kako bi dobili $g(X|\Phi)$ potrebne su nam određene pretpostavke na $g(X_i|\Phi)$ i $h(\Phi)$, ali te pretpostavke ne daju jedinstvenost faktorskog modela budući da je distribucija $h(\Phi)$ i dalje proizvoljna. Možemo, recimo, napraviti transformaciju Φ do novih faktora η koja ne utječe na $f(X)$, pa tada ni jedna količinu empirijske informacije ne čini razliku između transformacija i $h(\Phi|X)$. Zbog toga su nam potrebne pretpostavke na $h(\Phi)$. One proizlaze iz naših pretpostavki o prirodi problema ili se temelje na praktičnosti. Naša temeljna pretpostavka postavljena već na smom početku rada je linearnost faktorskog modela (1.2), iako se zapravo mogu razmatrati i nelinearne funkcije. Ako pretpostavimo da je $\Phi \sim N(0, I)$ tada je $X|\Phi$ također pretpostavljeno normalno distribuirano s očekivanjem $\alpha\Phi$ i kovarijacijskom matricom Ψ, to jest $X|\Phi \sim N(\alpha\Phi, \Psi)$. Budući da je Ψ dijagonalna, možemo faktorsku analizu karakterizirati preko glavnih komponenti, pomoću općih svojstva uvjetne nezavisnosti. Korištenjem teorema 0.1 i leme 3.3 iz trećeg poglavlja može se pokazati da vrijedi $\Phi|X \sim N[(\alpha^T\Sigma{-}1\alpha X), (\alpha^T\Psi^{-1}\alpha + I)]$. Prednosti teoretskog pristupa su ukazivanje na bitne pretpostavke za identikaciju problema, lakše je doći do generaliziranih struktura te otkriva moguća neslaganja između pretpostavki i postupka procjene. Kada uzmemo uzorak duljine n faktorski model (1.2) možemo napisati kao

$$\mathbb{X} = FA^T + e \tag{1.7}$$

pri čemu je $\mathbb{X}$ $(n \times p)$ matrica od n opservacija jednako distribuiranih slučajnih varijabli $X_1, X_2, \ldots, X_p$. Uobičajeno se pretpostavlja da su faktori slučajni, no također se može pretpostaviti da su fiksni. Ako pretpostavimo da su faktori ortonormalni i nekorelirani sa greškama, tada radeći sa uzoračkim vrijednostima iz (1.7) slijedi

$$\begin{aligned} F^T\mathbb{X} &= F^TFA^T + F^Te \\ &= A^T \end{aligned}$$

i

$$\begin{aligned} \mathbb{X}^T\mathbb{X} &= (FA^T + e)^T(FA^T + e) \\ &= AF^TFA^T + AF^Te + e^TFA^T + e^Te \\ &= AA^T + e^Te. \end{aligned} \tag{1.8}$$

2 Faktorska analiza glavnih komponenti

Faktorska analiza se od glavnih komponenti razlikuje po tome što se $r < p$ neopaženih ili latentnih zajedničkih faktora opaža pod pretpostavkama da su greške međusobno nekorelirane, heteroskedastične i nekorelirane sa zajedničkim faktorima.Također, glavne komponente za razliku od faktora u faktorskoj analizi ne pokazuju važnost individualne varijable u prisustvu drugih varijabli. Zajednički faktori po pretpostavci reproduciraju kovarijancu između opaženih varijabli, ali ne i varijancu koju reproduciraju glavne komponente. Analiza glavnih komponenti često služi kao međukorak za provođenje faktorske analize.

2.1 Model homoskedastičnih reziduala

Faktorski model homoskedastičnih reziduala je poseban slučaj modela (1.2), s pretpostavkom da greške imaju jednaku varijancu.

Teorem 2.1. *Neka je $\Sigma = \alpha\alpha^T + \Psi$ faktorski model, takav da je $\Psi = \sigma^2 I$ za neki skalar $\sigma^2 > 0$. Tada je model $\Sigma = \alpha\alpha^T + \sigma^2 I$ odredljiv.*

Dokaz. Neka su λ_i i Π_i svojstvene vrijednosti, odnosno svojstveni vektori od Σ $(i = 1, 2, \ldots, r)$. Tada vrijedi

$$\begin{aligned} 0 = (\Sigma - \lambda_i I)\Pi_i &= [(\alpha\alpha^T + \sigma^2 I) - \lambda_i I]\Pi_i \\ &= [\Gamma - (\lambda_i - \sigma^2)I]\Pi_i \\ &= [\Gamma - \lambda_i^* I]\Pi_i. \qquad\qquad (i = 1, 2, \ldots, r) \end{aligned} \tag{2.1}$$

Dakle, $\lambda_i^* = \lambda_i - \sigma^2$ je svojstvena vrijednost od $\Gamma = \alpha\alpha^T$. Sada imamo dekompoziciju glavnih komponenti od $\alpha\alpha^T$ što smo objasnili u poglavlju 0.2 u preliminarijama. Zbog $\lambda_i \geq \sigma^2$ za σ^2 se može uzeti najmanja svojstvena vrijednost od Σ čija je kratnost $p-r$, pa je faktorski model odredljiv. U slučaju $\rho(\Gamma) = r$, glavne komponente od Γ sadrže

r zajedničkih faktora, koji su jedinstveni do na predznak. Jednom kad su faktori poznati, tada ih možemo rotirati do tražene interpretabilnosti. Uz poznate (λ_i^*, Π_i), imamo matrični oblik

$$\Pi^T \Gamma \Pi = \Pi^T \alpha \alpha^T \Pi = \Lambda^*$$

pri čemu je $\alpha\alpha^T$ $(p \times p)$ matrica ranga r, $\Pi = [\Pi_1, \Pi_2, \ldots, \Pi_r]$ je $(p \times r)$ matrica, a

$$\Lambda^* = \begin{bmatrix} \lambda_1^* & & & \mathbf{0} \\ & \lambda_2^* & & \\ & & \ddots & \\ \mathbf{0} & & & \lambda_r^* \end{bmatrix}$$

dijagonalna matrica čiji dijagonalni elementi su svojetvene vrijednosti od Γ različite od 0. Koeficijenti faktora su dani sa

$$\alpha = \Pi \Lambda^{*1/2}$$

odnosno po komponentama $\alpha_i = \Pi_i \sqrt{\lambda_i^*}$ $(i = 1, 2, \ldots, r)$. □

Uzevši uzorak duljine n imamo dekompoziciju

$$\begin{aligned} \mathbb{X} &= \mathbb{X}^* + e \\ &= FA^T + e \end{aligned}$$

gdje gledajući uzoračke vrijednosti imamo da je $\mathbb{X}^T\mathbb{X} = A^TA + e^Te$ i $e^Te = s^2I$ je matrica homoskedastične uzoračke varijance grešaka. Tada je uzorački analogon jednadžbe (2.1)

$$(\mathbb{X}^T\mathbb{X} - l_iI)P_i = (AA^T - l_i^*I)P_i = 0 \tag{2.2}$$

gdje je $l_i^* = l_i - s^2$ $(i = 1, 2, \ldots, r)$. Jednakost (2.2) vrijedi samo ako je posljednjih $(p - r)$ svojstvenih vrijednosti jednako. s^2 možemo procijeniti sa

$$s^2 = \frac{\sum_{i=r+1}^{p} l_i}{p - r}.$$

Jednadžba (2.2) se također može napisati kao

$$P^TAA^TP = L^* \tag{2.3}$$

gdje je $L^* = L - s^2I$ dijagonalna. Jednadžba (2.3) je model glavnih komponenti sa $(p - r)$ jednakih svojstvenih vrijednosti i koeficijentima $A^T = L^{*1/2}P^T$.

2.2 Besteżinski modeli najmanjih kvadrata

Glavno ograničenje jednadžbe (2.1) je pretpostavka da su varijance grešaka jednake. Općenitiji model je oblika $\Sigma = \alpha\alpha^T + diag(\sigma_1^2, \ldots, \sigma_p^2)$, iz kojeg slijedi dekompozicija

$$0 = (\Sigma - \lambda_iI)\Pi_i = [\Gamma - (\lambda_i - \sigma_i^2)I]\Pi_i$$

gdje je $\Psi = diag(\sigma_1^2, \ldots, \sigma_p^2)$ matrica heteroskedastične varijance grešaka. Uočimo, Γ je Gramova matrica pa vrijedi $\lambda_i \geq \sigma_i^2$ $(i = 1, 2, \ldots, p)$. Praksa minimizacije sume kvadrata reziduala može se primijeniti kod široke klase faktorskih modela, koji se mogu riješiti korištenjem dekompozicije glavnih komponenti. Minimizira se (po Σ)

$$U = tr(S - \Sigma)^2.$$

Potpuni diferencijal od U je

$$\begin{aligned} dU &= d[tr(S-\Sigma)^2] \\ &= tr[d(S-\Sigma)^2] \\ &= -2\,tr[(S-\Sigma)\,d\Sigma] \\ &= -2\,tr[(S-\Sigma)(\alpha\,d\alpha^T + d\alpha\alpha^T)] \\ &= -4\,tr[(S-\Sigma)\alpha\,d\alpha^T] \end{aligned}$$

to jest

$$\frac{\partial U}{\partial \alpha} = -4(\Sigma - S)\alpha$$

prema lemi 0.1 i lemi 0.2. Izjednačavanjem sa 0 dobivamo normalne jednadžbe

$$\begin{aligned} 0 = (\hat{\Sigma} - S)\hat{\alpha} &= [(\hat{\alpha}\hat{\alpha}^T + \hat{\Psi}) - S])\hat{\alpha} \\ &= \hat{\alpha}(\hat{\alpha}^T\hat{\alpha}) - (S\hat{\alpha} - \hat{\Psi}\hat{\alpha}) \end{aligned} \tag{2.4}$$

gdje je $\hat{\alpha}^T\hat{\alpha} = L$ dijagonalna. Jednadžba (2.4) se može napisati kao

$$(S - \hat{\Psi})\hat{\alpha} = \hat{\alpha}L \tag{2.5}$$

gdje su L i $\hat{\alpha}$ svojstvene vrijednosti, odnosno svojstveni vektori od $(S-\hat{\Psi})$. Jednadžba (2.5) predstavlja analizu glavnih komponenti korigirane kovarijacijske matrice $(S-\hat{\Psi})$ koja sadrži najviše r svojstvenih vrijednosti različitih od 0. Po pretpostavci, greške nisu poznate unaprijed. Treba procijeniti koeficijente i varijancu grešaka istodobno. To možemo učiniti na dva načina. Možemo uzeti $\hat{\Psi}_0 = diag(S^{-1})$, te uvrstiti dobiveno u (2.5) čime dobivamomo da su $\hat{\alpha}_0$ svojstveni vektori od $S - diag(S^{-1})$. Sada možemo izračunati $\hat{\Psi}_{(1)}$. Imamo

$$\hat{\Psi}_{(1)} = diag(S - \hat{\alpha}_0\hat{\alpha}_0^T)$$

gdje zbog nekoreliranosti grešaka $\hat{\Psi}$ mora biti dijagonalna. Kada znamo $\hat{\Psi}_{(1)}$, uvrstimo ga u jednadžbu (2.5) te dobivamo novi $\hat{\alpha}_1^T$ te poboljšani procjenitelj $\hat{\Psi}_{(2)}$. Ponavljamo postupak sve dok varijanca grešaka i koeficijenti ne konvergiraju prema nekoj fiksnoj vrijednosti. Također smo mogli uzeti $\hat{\Psi}_0 = 0$. Tada $\hat{\alpha}_0$ predstavlja svojstvene vektore matrice S. Druga metoda je model poznat pod nazivom "*image* faktorski model" koji ćemo obraditi u nastavku.

2.3 *Image* faktorski model

Varijanca grešaka se može također procijeniti regresijskom analizom. Za dijagonalne elemente matrice $S - \hat{\Psi}$ se uzima koeficijent determinacije (vrijednost R^2). Neka je Y $(n \times 1)$ vektor opservacija zavisne varijable i neka je X $(n \times p)$ matrica sa n opservacija p nezavisnih varijabli tako da vrijedi $Y = \chi\beta + \epsilon$ i $X = \chi + \delta$. δ je matrica grešaka mjerenja za nezavisne varijable, χ je matrica "pravih" vrijednosti od X, a ϵ je greška u Y. Kada za Y uzmemo jednu od varijabli $X_1, X_2, \ldots, X_p$ imamo

$$\begin{aligned} Y &= \chi\beta + \epsilon \\ &= (X - \delta)\beta + \epsilon \\ &= X\beta + \eta \end{aligned} \tag{2.6}$$

pri čemu je $\eta = \epsilon - \delta\beta$ greška u zavisnoj i nezavisnim varijablama, a X reprezentira matricu podataka za preostalih $p-1$ varijabli. Procjenitelj za β metodom najmanjih kvadrata je

$$\begin{aligned} \hat{\beta} &= (X^T X)^{-1} X^T Y \\ &= (X^T X) X^T (X\beta + \eta) \\ &= (X^T X)^{-1} X^T X\beta + (X^T X)^{-1} X^T \eta \\ &= \beta + (X^T X)^{-1} X^T \eta \end{aligned}$$

Vrijedi

$$\begin{aligned} E[(X^T X)^{-1} X^T \eta] &= E[(X^T X)^{-1} (\chi + \delta)^T (\epsilon - \delta\beta)] \\ &= E[(X^T X)^{-1} (\chi^T \epsilon - \chi^T \delta\beta + \delta^T \epsilon - \delta^T \delta\beta)] \end{aligned}$$

pa čak i kada je $\chi^T \epsilon = X^T \delta = \delta^T \epsilon = 0$ imamo $\delta^T \delta \neq 0$ zbog čega je $E[(X^T X)^{-1} X^T \eta] \neq 0$. Stoga slijedi

$$\begin{aligned} E(\hat{\beta}) &= \beta + E[((X^T X)^{-1} X^T \eta)] \\ &\neq \beta \end{aligned}$$

što znači da je $\hat{\beta}$ pristrani procjenitelj za β. Slijedi da su predviđene vrijednosti (također i R^2) pristrane te da dekompozicija glavnih komponenti korelacijske matrice reduciranog faktorskog modela daje pristrane procjenitelje. Pristranost je još veća ako se koristi nereducirana kovarijacijska (korelacijska) matrica.

Druga varijanta *image* faktorske analize koristi težinski model glavnih komponenti pri čemu se komponentna analiza od $\mathbb{X}^T\mathbb{X}$ izvodi iz

$$\begin{aligned} 0 = ((e^T e)^{-1} \mathbb{X}^T \mathbb{X} - l_i I) P_i &= [(e^T e)^{-1} (\mathbb{X} - e)^T (\mathbb{X} - e) - (l_i + 1) I] P_i \\ &= (e^T e)^{-1} (\mathbb{X}^T \mathbb{X} - l_i e^T e) P_i \end{aligned}$$

pri čemu je $e^T e$ dijagonalna matrica. $(n \times p)$ matrica reziduala e se obično računa regresijom, ali još uvijek ostaje utjecaj grešaka u regresijskim jednadžbama. Težinski *image* faktorski model također daje pristrane procjenitelje.

Postoji još nekoliko vrsta *image* faktorskih modela. Oni se razlikuju po načinu na koji procijenjuju grešku. Uglavnom, *image* faktorski modeli su faktorski modeli koji koriste kombinaciju glavnih komponenti i regresiju metodom najmanjih kvadrata.

2.4 *Whittle* model

Faktori glavnih komponenti se uglavnom smatraju neopaženim varijablama koje variraju slučajno u populaciji. Whittle (1953.) je pokazao da se faktori također mogu smatrati fiksnim varijablama koje su rješenje modela težinskih glavnih komponenti

$$(\Sigma - \lambda_i \Psi)\Pi_i = 0 \qquad (i = 1, 2, \ldots, r) \tag{2.7}$$

pri čemu je Ψ dijagonalna matrica varijance reziduala. Kada Ψ nije poznata postavlja se $\Psi = I$ te se računa njegova vrijednost iterativnim postupkom. Izračuna se $r < p$ komponenti te se iskoriste za izračun varijance reziduala koja se uvrštava u jednadžbu (2.7). Ponovo se dobiju zajednički faktori pa varijanca reziduala, sve dok $\hat{\Psi}$ i koeficijenti ne konvergiraju prema nekoj stabilnoj vrijednosti.

3 Faktorska analiza maksimalne vjerodostojnosti

Kada podaci dolaze iz višedimenzionalne normalne razdiobe može se koristiti metoda maksimalne vjerodostojnosti (ML metoda) za rješavanje normalnih jednadžbi. Prednosti ML procjenitelja su da su efikasni, konzistentni te se mogu sprovesti statistička testiranja parametara. ML procjena ima danas široku primjenu. Metoda maksimalne vjerodostojnosti može se koristiti samo kada su faktori slučajni; za fiksne faktore ML procjenitelji ne postoje.

3.1 Model recipročne proporcionalnosti

Razmatramo faktorski model oblika (1.2) pri čemu je Ψ dijagonalna matrica heteroskedastičnih varijanci grešaka. Kao što smo već rekli, zbog postojanja beskonačno mnogo rješenja nužno je uvesti dodatne pretpostavke kako bi došli do jedinstvenog rezultata. Uz osnovnu pretpostavku da su zajednički faktori ortogonalni, Lawley (1953) i Joreskog (1962,1963) su predložili procijenu modela na način da stavimo

$$\begin{aligned} \Psi &= \sigma^2 (diag\, \Sigma^{-1})^{-1} \\ &= \sigma^2 \Delta^{-1} \end{aligned} \tag{3.8}$$

pri čemu je σ^2 proizvoljan skalar a $\Delta = diag\, \Sigma^{-1}$. Iz (3.8) slijedi da je varijanca reziduala proporcionalna recipročnoj vrijednosti dijagonalnih elemenata od Σ^{-1}, po čemu je dani model i dobio ime. Dakle, sada faktorski model možemo napisati kao

$$\begin{aligned} \Sigma &= \alpha\alpha^T + \Psi \\ &= \Gamma + \sigma^2 \Delta^{-1} \end{aligned} \tag{3.9}$$

Množenjem slijeva i s desna sa $\Delta^{1/2}$ dobivamo

$$\begin{aligned} \Delta^{1/2}\Sigma\Delta^{1/2} &= \Delta^{1/2}\Gamma\Delta^{1/2} + \sigma^2 I \\ &= \Sigma^* \end{aligned}$$

pri čemu Σ^* zovemo težinska kovarijacijska matrica. Vidimo da restrikcija (3.8) dovodi do transformacije početnog modela do modela homoskedastičnih reziduala, čije rješavanja je objašnjeno u poglavlju 2.1.

Teorem 3.1. *Neka je Σ kovarijacijska matrica populacije takva da je je $\Sigma = \alpha\alpha^T + \sigma^2\Delta^{-1}$ i $\Delta = diag(\Sigma^{-1})$. Tada je prvih $r < p$ svojstvenih vrijednosti $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r$ matrice $\Delta^{1/2}\Sigma\Delta^{1/2}$ veće od σ^2. Preostalih $p - r$ svojstvenih vrijednosti je jednako σ^2.*

Dokaz. Neka je λ_i i-ta svojstvena vrijednost od $\Sigma^* = \Delta^{1/2}\Sigma\Delta^{1/2}$, te neka je γ_i i-ta svojstvena vrijednost od $\Delta^{1/2}\alpha\alpha^T\Delta^{1/2} = \Delta^{1/2}\Gamma\Delta^{1/2}$. Tada je

$$\begin{aligned}
|\Sigma^* - \lambda_i I| &= |\Delta^{1/2}\Sigma\Delta^{1/2} - \lambda_i I| \\
&= |(\Delta^{1/2}\alpha\alpha^T\Delta^{1/2} - \sigma^2 I) - \lambda_i I| \\
&= |\Delta^{1/2}\alpha\alpha^T\Delta^{1/2} - (\lambda_i - \sigma^2)I| \\
&= |\Delta^{1/2}\Gamma\Delta^{1/2} - \gamma_i I|
\end{aligned}$$

gdje je $\gamma_i = \lambda_i - \sigma^2$ za $i = 1, 2, \ldots, p$. Zbog pozitivne semidefinitnosti $(p \times p)$ matrice Γ ranga $r < p$ ima r svojstvenih vrijednosti većih od 0, te 0 kao svojstvenu vrijednost s kratnošću $(p - r)$. Dakle, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r > \lambda_{r+1} = \lambda_{r+2} = \ldots = \lambda_p = \sigma^2$. □

Sada se model može riješiti analizom glavnih komponenti skalirane matrice Σ^*. Alternativno, koeficijenti se mogu dobiti iz jednadžbe

$$\Sigma\Delta\alpha = \alpha\Lambda \tag{3.10}$$

Ograničenje da je $\alpha^T\Lambda\alpha$ dijagonalna omogućuje jedinstvenu identifikaciju koeficijenata, to jest fiksira poziciju koordinatnih osi. U praksi, kada r nije poznat, njegova vrijednost se određuje na temelju pokušaja i pogrešaka u kombinaciji sa testovima značajnosti. Jednadžba (3.9) ima još jedno poželjno svojstvo: α su koeficijenti korelacije bez obzira da li radimo sa kovarijacijskom ili korelacijskom matricom.

Pretpostavimo da su faktori $\Phi_1, \Phi_2, \ldots, \Phi_r$ i greške $\epsilon_1, \epsilon_2, \ldots, \epsilon_p$ nezavisni, normalno distribuirani s očekivanjem 0 i kovarijacijskom matricom I odnosno Ψ. Kada su promatrane varijable višedimenzionalne normalne može se pokazati da uzoračka varijanca i kovarijanca imaju Wishartovu distribuciju čija je gustoća

$$f(S) = c|\Sigma|^{-n/2}|S|^{1/2(n-p-1)}exp[-\tfrac{1}{2}tr(\Sigma^{-1}S)].$$

S se može zamjeniti sa $\mathbb{X}^T\mathbb{X}$ uz prikladnu prilagodbu u konstanti proporcionalnosti c. Zamjena Σ sa Σ^*, odnosno S sa

$$\begin{aligned}
S^* &= (diag\, S^{-1})^{1/2} S (diag\, S^{-1})^{1/2} \\
&= D^{1/2} S D^{1/2} \\
&= \frac{1}{n-1} D^{1/2}(\mathbb{X}^T\mathbb{X}) D^{1/2}
\end{aligned}$$

pri čemu je $D = diag\, S^{-1}$ ne zadržava nužno Wishartovu distribuciju. Egzaktna distribucija od S^* nije poznata, no možemo dobiti asimptotski rezultat. Uz pretpostavku $D \to \Delta$ za $n \to \infty$ aproksimativna funkcija vjerodostojnosti je

$$\begin{aligned} L(\Sigma^*) &= k[-ln|\Sigma^*| + tr(S^*(\Sigma^*)^{-1})] \\ &= k[-ln|\alpha\alpha^T + \Psi| + tr(S^*(\alpha\alpha^T + \Psi)^{-1})] \end{aligned} \tag{3.11}$$

pri čemu je k konstanta proporcionalnosti koja ignorira izraze koji ovise o n jer oni teže u 0 zbog $n \to \infty$. Diferenciranjem jednadžbe (3.11) te uz normalne jednadžbe

$$(S^* - \hat{\lambda}_i I)\hat{\alpha}_i = 0 \tag{3.12}$$

imamo

$$\hat{\alpha}_i^T \hat{\alpha}_i = \hat{\lambda}_i - \hat{\sigma}_i^2 \qquad (i = 1, 2, \ldots, r) \tag{3.13}$$

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{p-r}[tr\, S^* - \sum_{i=1}^{r} \hat{\lambda}_i] \\ &= \frac{1}{p-r} \sum_{i=r+1}^{p} \hat{\lambda}_i \end{aligned} \tag{3.14}$$

to jest $\hat{\sigma}_i^2, \hat{\alpha}_i$ i $\hat{\lambda}_i$ su ML procjenitelji. Alternativno, koeficijenti korelacije se mogu dobiti rješavanjem uzoračke verzije jednadžbe (3.10), odnosno rješavanjem

$$SD\hat{\alpha} = \hat{\alpha}\hat{\Lambda}$$

Kada varijable nisu višedimenzionalne normalne, procjenitelji $(3.12 - 3.14)$ su još uvijek optimalni ali u smislu najmanjih kvadrata.

3.2 Lawleyev faktorski model

Originalni ML faktorski model dan je od Lawleya (1940, 1941) i razlikuje se od modela recipročne proporcionalnosti u težinskoj shemi koja se koristi u kovarijacijskoj matrici. Također je invarijantan na skaliranje, pa dobivamo identične koeficijente korelacije bez obzira da li gledamo kovarijacijsku, korelacijsku ili $\mathbb{X}^T\mathbb{X}$ matricu.

Neka je $Y = (y_1, y_2, \ldots, y_p)^T$ normalno distribuiran vektor s očekivanjem μ i kovarijacijskom matricom Σ. Tada korištenjem Wishartove distribucije dobivamo da se vjerodostojnost može izraziti sa

$$L(\Sigma) = ln\, c - \frac{n}{2} ln|\Sigma| + \frac{1}{2}(n - p - 1) ln|S| - \frac{n}{2} tr(\Sigma^{-1} S) \tag{3.15}$$

čiji maksimum je jednak maksimumu od

$$\begin{aligned} L &= -\frac{n}{2}[ln|\Sigma| + tr(\Sigma^{-1} S)] \\ &= -\frac{n}{2}[ln|\alpha\alpha^T + \Psi| + tr((\alpha\alpha^T + \Psi)^{-1} S)] \end{aligned} \tag{3.16}$$

Uz pretpostavku da je Φ slučajan vektor a α fiksna matrica mi zapravo maksimiziramo jednadžbu

$$L = ln|\Sigma| + tr(S\Sigma^{-1}) \tag{3.17}$$

koja ovisi samo o α i Ψ. Parcijalnim deriviranjem po Ψ dobivamo

$$\begin{aligned}\frac{\partial L}{\partial \Psi} &= \frac{\partial ln|\Sigma|}{\partial \Psi} + \frac{\partial (tr\, S\Sigma^{-1})}{\partial \Psi} \\ &= \frac{\partial ln|\alpha\alpha^T + \Psi|}{\partial \Psi} + \frac{\partial [tr\, S(\alpha\alpha^T + \Psi)^{-1}]}{\partial \Psi} \\ &= diag(\alpha\alpha^T + \Psi)^{-1} - (\alpha\alpha^T + \Psi)^{-1} S(\alpha\alpha^T + \Psi)^{-1} \frac{\partial \Psi}{\partial \Psi}\end{aligned}$$

pri čemu je $\partial\Psi/\partial\Psi = I$. Da bi dobili maksimum izjednačimo derivaciju sa 0. Sada imamo

$$\frac{\partial L}{\partial \Psi} = diag(\hat{\alpha}\hat{\alpha}^T + \hat{\Psi})^{-1} - (\hat{\alpha}\hat{\alpha}^T + \hat{\Psi})^{-1} S(\hat{\alpha}\hat{\alpha}^T + \hat{\Psi})^{-1} = 0$$

odnosno

$$diag[\hat{\Sigma}^{-1}(\hat{\Sigma} - S)\hat{\Sigma}^{-1}] = 0 \tag{3.18}$$

pri čemu je $\hat{\Sigma} = \hat{\alpha}\hat{\alpha}^T + \hat{\Psi}$. Jednadžba (3.18) je ekvivalentna

$$diag(\hat{\Sigma}) = diag(S)$$

to jest uvjetu da je varijanca dobivena ML procjeniteljima $\hat{\alpha}$ i $\hat{\Psi}$ jednaka uzoračkoj varijanci dobivenoj iz opaženih podataka.
Vrijedi sljedeća lema:

Lema 3.1. *Maksimizacija jednadžbe* (3.15) *ekvivalentna je minimizaciji*

$$\begin{aligned}F(\Sigma) &= ln|\Sigma| + tr(S\Sigma^{-1}) - ln|S| - p \\ &= tr(\Sigma^{-1}S) - ln|\Sigma^{-1}S| - p.\end{aligned}$$

F iz dane leme je nenegativna i jednaka 0 samo za $\Sigma = S$.
Minimizirajmo sada F po α. Korištenjem leme 0.1 i leme 0.2 iz preliminarija slijedi

$$\begin{aligned}dF &= d\, tr(\Sigma^{-1}S) - d\, ln|\Sigma^{-1}S| \\ &= tr(d\Sigma^{-1}S) - tr(S^{-1}\Sigma\, d\Sigma^{-1}S) \\ &= tr[(S - \Sigma)d\Sigma^{-1}] \\ &= tr[(\Sigma - S)\Sigma^{-1}\, d\Sigma\Sigma^{-1}] \\ &= tr[\Sigma^{-1}(\Sigma - S)\Sigma^{-1}(d\alpha\alpha^T + \alpha\, d\alpha^T)] \\ &= 2tr[\Sigma^{-1}(\Sigma - S)\Sigma^{-1}\alpha\, d\alpha^T]\end{aligned}$$

pa je

$$\frac{\partial F}{\partial \alpha} = 2\Sigma^{-1}(\Sigma - S)\Sigma^{-1}\alpha$$

što izjednačavanjem sa 0 daje normalne jednadžbe

$$(\hat{\Sigma} - S)\hat{\Sigma}^{-1}\hat{\alpha} = 0 \tag{3.19}$$

Dakle, jednadžbe (3.18) i (3.19) čine normalne jednadžbe za Lawleyev ML faktorski model. Zbog nejedinstvenog rješenja jednadžbi (3.19) potrebno je uvesti ograničenje na faktorski model kako bi bio odredljiv. Ograničenje koje se koristi u ovom modelu je pretpostavka da je $\alpha^T\Psi^{-1}\alpha = \eta$ dijagonalna matrica. Time se fiksira koordinatni sustav, pa dobivamo jedinstvene koeficijente korelacije α. Jednom kada dobijemo rješenje, možemo maknuti ograničenje ortogonalnim ili kosim rotiranjem.

Lema 3.2. *Neka je $\Sigma = \alpha\alpha^T + \Psi$ faktorska dekompozicija od Σ. Tada vrijedi*

$$\Sigma^{-1} = \Psi^{-1} - \Psi^{-1}\alpha(I + \eta)^{-1}\alpha^T\Psi^{-1}. \tag{3.20}$$

Dokaz. Množenjem jednadžbe (3.21) sa Σ dobivamo

$$\begin{aligned}
\Sigma^{-1}\Sigma &= [\Psi^{-1} - \Psi^{-1}\alpha(I + \eta)^{-1}\alpha^T\Psi^{-1}](\alpha\alpha^T + \Psi) \\
&= \Psi^{-1}(\alpha\alpha^T + \Psi) - \Psi^{-1}\alpha(I + \eta)^{-1}\alpha^T\Psi^{-1}(\alpha\alpha^T + \Psi) \\
&= \Psi^{-1}\alpha\alpha^T + I - \Psi^{-1}\alpha(I + \eta)^{-1}\alpha^T\Psi^{-1}\alpha\alpha^T - \Psi^{-1}\alpha(I + \eta)^{-1}\alpha^T\Psi^{-1}\Psi \\
&= \Psi^{-1}\alpha\alpha^T + I - \Psi^{-1}\alpha(I + \eta)^{-1}\eta\alpha^T - \Psi^{-1}\alpha(I + \eta)^{-1}\alpha^T \\
&= \Psi^{-1}\alpha\alpha^T + I - \Psi^{-1}\alpha\alpha^T \\
&= I
\end{aligned}$$

pri čemu je $\eta = \alpha^T\Psi^{-1}\alpha$ dijagonalna. Dakle, vrijedi jednakost (3.20). □

Množenjem (3.20) sa α dolazimo do sljedeće leme:

Lema 3.3. *Prepostavimo da vrijede uvjeti leme 3.2. Tada je*

$$\Sigma^{-1}\alpha = \Psi^{-1}\alpha(I + \eta)^{-1}. \tag{3.21}$$

Dokaz. Množenjem jednadžbe (3.20) sa α dobivamo

$$\begin{aligned}
\Sigma^{-1}\alpha &= \Psi^{-1}\alpha - \Psi^{-1}\alpha(I + \eta)^{-1}\alpha^T\Psi^{-1}\alpha \\
&= \Psi^{-1}\alpha - \Psi^{-1}\alpha(I + \eta)^{-1}\eta.
\end{aligned}$$

Budući da je η dijagonalna sa dijagonalnim elementima različitim od 0, množenjem sa $\eta^{-1}(I + \eta)$ slijedi

$$\Sigma^{-1}\alpha\eta^{-1}(I + \eta) = \Psi^{-1}\alpha\eta^{-1}(I + \eta) - \Psi^{-1}\alpha$$

odnosno

$$\begin{aligned}
\Sigma^{-1}\alpha\eta^{-1} + \Sigma^{-1}\alpha &= \Psi^{-1}\alpha\eta^{-1} + \Psi^{-1}\alpha - \Psi^{-1}\alpha \\
&= \Psi^{-1}\alpha\eta^{-1}.
\end{aligned}$$

Množenjem sa η slijedi

$$\Sigma^{-1}\alpha + \Sigma^{-1}\alpha\eta = \Psi^{-1}\alpha$$

to jest

$$\Sigma^{-1}\alpha = \Psi^{-1}\alpha(I + \eta)^{-1}.$$

□

Iskoristivši jednakost (3.21) u jednadžbi (3.19) imamo

$$(\hat{\Sigma} - S)\hat{\Psi}^{-1}\hat{\alpha}(I + \hat{\eta})^{-1} = 0$$

pri čemu množenjem sa $(I + \eta)$ te uz $\hat{\Sigma} = \hat{\alpha}\hat{\alpha}^T + \hat{\Psi}$ dobivamo

$$\hat{\alpha}\hat{\alpha}^T\hat{\Psi}^{-1}\hat{\alpha} + \hat{\alpha} - S\hat{\Psi}^{-1}\hat{\alpha} = 0$$

odnosno

$$S\hat{\Psi}^{-1}\hat{\alpha} = \hat{\alpha}(\hat{\eta} + I).$$

Množenjem s lijeva sa $\hat{\Psi}^{-1/2}$ i grupiranjem imamo

$$[\hat{\Psi}^{-1/2}S\hat{\Psi}^{-1/2} - (\hat{\eta} + I)]\hat{\Psi}^{-1/2}\hat{\alpha} = 0. \tag{3.22}$$

Neka je $S^* = \hat{\Psi}^{-1/2}S\hat{\Psi}^{-1/2}$ težinska uzoračka kovarijacijska matrica. Tada možemo dane normalne jednadžbe napisati kao

$$[S^* - (\hat{\eta}_i + 1)I]\hat{\Psi}_i^{-1/2}\hat{\alpha}_i = 0 \qquad (i = 1, 2, \ldots, r) \tag{3.23}$$

gdje je jasno da je $\hat{\eta}_i + 1$ i-ta svojstvena vrijednost od S^*.
Koeficijenti $\hat{\alpha}_i$ se dobiju iz svojstvenih vektora $\hat{\Psi}_i^{-1/2}\hat{\alpha}_i$ i uz uvjet da je $\hat{\alpha}^T\hat{\Psi}^{-1/2}\hat{\alpha} = \hat{\eta}$ dijagonalna matrica, kojim je osigurana odrediljivost danog faktorskog modela. Model recipročne proporcionalnosti i Lawleyev faktorski model uglavnom se razlikuju po početnim ograničenjima koja koriste kako bi model bio odredljiv. U ML faktorskoj analizi kada imamo normalno distribuirane slučajne varijable, ograničenje da je $\alpha^T\Psi^{-1}\alpha$ dijagonalna matrica je ekvivalentno ograničenju da je kovarijacijska matrica od $\Phi|X$ dijagonalna, to jest da su faktori Φ uz dano X nezavisni. To slijedi iz činjenice da ako je $\alpha^T\Psi^{-1}\alpha$ dijagonalna matrica, tada su vandijagonalni elementi

$$\sum_{i=1}^{p} \alpha_{ik}\alpha_{il}/\psi_i$$

jednaki nula. Također dobivamo da su dijagonalni elementi oblika α_{ik}^2/ψ_i i predstavljaju dio varijance varijable X_i koja je objašnjena k-tim faktorom ζ_k. Dijagonalni elementi od $\alpha^T\Psi^{-1}\alpha$ su poredani u padajućem poretku tako da ζ_1 objašnjava najveći dio varijance, ζ_2 objašnjava najveći dio od preostale neobjašnjene varijance i tako sve do ζ_r koji objašnjava najmanji dio ukupne varijance.

Teorem 3.2. *Neka je $X = (X_1, X_2, \ldots, X_p)^T$ p-dimenzionalan slučajan vektor takav da je $X \sim N(\Phi, \Sigma)$ i $\Sigma = \alpha\alpha^T + \Psi$. Tada*

(i) Ako postoji jedinstvena dijagonalna matrica Ψ sa pozitivnim dijagonalnim elementima takva da r najvećih svojstvenih vrijednosti matrice $\Sigma^* = \Psi^{-1/2}\Sigma\Psi^{-1/2}$ je veće od jedan, a preostalih $p-r$ je jednako jedan, tada α možemo jedinstveno definirati sa

$$[\Sigma^* - (\eta_i + 1)I]\psi_i^{-1/2}\alpha_i = 0. \tag{3.24}$$

(ii) Neka je

$$\Sigma^* = \Psi^{-1/2}\Gamma\Psi^{-1/2} + I \tag{3.25}$$

skalirani model. Tada je $\eta_i = \alpha_i^T \Psi_i^{-1}\alpha_i$ $(i = 1, 2, \ldots, r)$ *i-ta svojstvena vrijednost od* $\Psi^{-1/2}\Gamma\Psi^{-1/2}$ *takva da vrijedi* $\eta_1 \geq \eta_2 \geq \ldots \geq \eta_r > 0$ *i* $\eta_{r+1} = \eta_{r+2} = \ldots = \eta_p = 0$.

Dokaz. (i) Budući da je Σ^* Gramova te nesingularna matrica, možemo napisati

$$(\Sigma^* - \lambda_i I)\Pi_i = 0$$

za $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r > \lambda_{r+1} = \lambda_{r+2} = \ldots = \lambda_p = 1$. Kada je Ψ jedinstvena, tada su i matrice α i Π jedinstvene. Postavljanjem $\lambda_i = \eta_i + 1$ i $\Pi_i = \Psi_i^{-1/2}\alpha_i$ direktno dobivamo jednadžbu (3.24).

(ii) Za $\Sigma^* = \Psi^{-1/2}\Gamma\Psi_i^{-1/2} + I$ iz jednadžbe (3.22) slijedi

$$[(\Psi^{-1/2}\Gamma\Psi^{-1/2} + I) - (\eta_i + 1)I]\Psi^{-1/2}\alpha_i = 0$$

odnosno

$$(\Psi^{-1/2}\Gamma\Psi^{-1/2} - \eta_i I)\Psi X_i^{-1/2}\alpha_i = 0.$$

Budući da je $\Gamma = \alpha\alpha^T$ $(p \times p)$ matrica i $\rho(\Gamma) = r$, posljednjih $p - r$ svojstvenih vrijednosti od Γ je jednako nula, što odgovara nula svojstvenim vrijednostima od $\Psi^{-1/2}\Gamma\Psi^{-1/2}$. Time također slijedi da je posljednjih $p-r$ svojstvenih vrijednosti od Σ^* jednako jedan.

□

Kada podaci ne dolaze iz normalne razdiobe, normalne jednadžbe (3.23) još uvijek daju optimalne procjenitelje, ali u smislu najmanjih kvadrata. Također treba imati na umu da svojstvene vrijednosti od Σ^* moraju biti veće od jedan ako želimo da Γ bude pozitivno definitna matrica.
Lawleyev model također možemo gledati preko dekompozicije glavnih komponenti težinske kovarijacijske matrice Σ^* pri čemu su matrica Σ i težine $\Psi^{-1/2}$ procijenjene sljedećim iterativnim postupkom:

(1) Izračunamo S (ili $\mathbb{X}^T\mathbb{X}$), izdvojimo prvih r glavnih komponenti i korištenjem (1.8) izračunamo procjenu za Ψ, recimo $\hat{\Psi}_{(1)}$.

(2) Konstruiramo težinsku kovarijacijsku matricu $S_{(1)} = \hat{\Psi}_{(1)}^{-1/2} S \hat{\Psi}_{(1)}^{-1/2}$, izračunamo glavne komponente te dobijemo sljedeći procjenitelj za Ψ, recimo $\hat{\Psi}_{(2)}$.

(3) Ponavajamo postupak sve dok ne dobijemo konvergenciju procjenitelja od Ψ prema nekoj fiksnoj vrijednosti $\hat{\Psi}$.

Neka je sada

$$L_{(r)} = \begin{bmatrix} ł_1 & & & \mathbf{0} \\ & ł_2 & & \\ & & \ddots & \\ \mathbf{0} & & & ł_r \end{bmatrix}$$

i $P_1, P_2, \ldots, P_r$ prvih r svojstvenih vrijednosti, odnosno svojstvenih vektora matrice $S^* = \hat{\Psi}^{-1/2} S \hat{\Psi}^{-1/2}$. Tada je

$$S^* P_{(r)} = P_{(r)} L_{(r)}$$

pri čemu je $P_{(r)}$ $(p \times r)$ matrica prvih r svojstvenih vektora od S^*. ML procjenitelj $\hat{\eta}$ se dobije kao

$$\hat{\eta} = L_{(r)} - I$$

pri čemu je

$$\hat{\alpha}^T \hat{\Psi}^{-1} \hat{\alpha} = \hat{\eta}$$

i

$$\begin{aligned} \hat{\Psi}^{-1/2} \hat{\alpha} &= P_{(r)} \hat{\eta}^{1/2} \\ &= P_{(r)} (L_{(r)} - I) \end{aligned}$$

pa je

$$\hat{\alpha} = \hat{\Psi}^{1/2} P_{(r)} (L_{(r)} - I)$$

$(p \times r)$ matrica koeficijenata korelacije. Problem sa ML faktorskim modelom je da funkcija vjerodostojnosti možda nema maksimum koji zadovoljava tražene uvjete: da su greške striktno pozitivne ili da su varijable višedimenzionalne normalne. Čak i kada postoji jedinstveni globalni maksimum koji zadovoljava uvjete, može se desiti da ga numerički algoritam ne locira točno budući da je moguće da iterativni slijed konvergira prema lokalnom maksimumu. Uočimo da je vrlo važno provjeriti da li dobiveni maksimum zadovoljava sve dane pretpostavke. Prednost ovog modela su dobra svojstva procjenitelja $\hat{\alpha}$: velika efikasnost te mala pristranost kao i invarijantnost na jedinice mjerenja.

Teorem 3.3. *Neka je X slučajni vektor sa nesingularnom kovarijacijskom matricom Σ. Tada je težinska kovarijacijska matrica $\Sigma^* = \Psi^{-1/2} \Sigma \Psi^{-1/2}$ invarijantna na transformacije skale od X.*

Dokaz. Neka je H dijagonalna matrica takva da je $Z = HX$, gdje matrica Z predstavlja reskaliranje slučajnih varijabli. Sada je faktorska dekompozicija oblika

$$\begin{aligned} H \Sigma H^T &= H \Gamma H^T + H \Psi H^T \\ &= H \Gamma H^T + (H \Psi^{1/2})(H \Psi^{1/2})^T \end{aligned}$$

te težinska varijanta sa reskaliranom varijancom grešaka je oblika

$$(H \Psi^{1/2})^{-1} H \Sigma H^T (\Psi^{1/2} H^T)^{-1} = (H \Psi^{1/2})^{-1} H \Gamma H^T (\Psi^{1/2} H^T)^{-1} + I$$

odnosno

$$\Psi^{-1/2} \Sigma \Psi^{-1/2} = \Psi^{-1/2} \Gamma \Psi^{-1/2} + I$$

što je zapravo težinska kovarijacijska matrica originalnih, neskaliranih, varijabli. □

3.3 Raov model kanonske korelacije

Alternativni oblik ML faktorskog modela je dao Rao (1955). Izvod tog modela je zanimljiv stoga što pruža drugačiji pogled na faktorski model te pokazuje optimalnost Lawleyovog modela bez pretpostavke da su varijable višedimenzionalne normalne. Neka je X $(p \times 1)$ vektor slučajnih varijabli i $X^* = E(X|\Phi) = \alpha\Phi$ dio koji je predviđen sa skupom od r zajedničkih faktora. Tada je $X = \alpha\Phi + \epsilon = X^* + \epsilon$. Nama je cilj izračunati koeficijente koji maksimiziraju korelaciju između $X = (X_1, X_2, \ldots, X_p)^T$ i $X^* = (X_1^*, X_2^*, \ldots, X_p^*)^T$. Sada, prema poglavlju (0.4) iz preliminarija, znamo da je maksimalna korelacija između X i X^* najveća kanonska korelacija između linearnih kombinacija $U = \beta^T X$ i $V = \gamma X^*$ pri čemu su β i γ koeficijenti koje moramo izračunati. Imamo

$$\begin{aligned}
var(U) &= E(U^2) = E(\beta^T X X^T \beta) = \beta^T E(XX^T)\beta = \beta^T \Sigma \beta \\
var(V) &= E(V^2) = E(\gamma^T X^* X^{*T} \gamma) = \gamma^T E(X^* X^{*T})\gamma = \gamma^T \Gamma \gamma \\
cov(U,V) &= E(UV) = E(\beta^T X X^{*T} \gamma) = \beta^T E(X X^{*T})\gamma = \beta^T \Gamma \gamma
\end{aligned}$$

budući da je

$$\begin{aligned}
XX^{*T} &= (\alpha\Phi + \epsilon)(\alpha\Phi)^T \\
&= \alpha\Phi\Phi^T\alpha^T + \epsilon\Phi^T\alpha^T \\
&= \alpha\alpha^T \\
&= X^* X^{*T}
\end{aligned}$$

uz $\Phi\Phi^T = I$ i $\epsilon\Phi^T = 0$ po pretpostavci. Stoga maksimiziramo $cov(U,V)$ uz ograničenje da su varijance od U i V jedinične. Iz kanonskog korelacijskog modela (vidi poglavlje 0.4) imamo $\Sigma = \Sigma_{11}, \Sigma_{12} = \Sigma_{22} = \Sigma_{21} = \Gamma$ i

$$\begin{aligned}
(\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \lambda_i^2 I)\Pi_i &= (\Sigma^{-1}\Gamma\Gamma^{-1}\Gamma - \lambda_i^2 I)\Pi_i \\
&= (\Sigma^{-1}\Gamma - \lambda_i^2 I)\Pi_i
\end{aligned}$$

odnosno

$$(\Gamma - \Sigma\lambda_i^2)\Pi_i = 0 \tag{3.26}$$

za $i = 1, 2, \ldots, r$ maksimalnih korelacija. λ_i^2 su rješenja determinantnih jednadžbi

$$\begin{aligned}
0 = |\Gamma - \Sigma\lambda_i^2| &= |(\Sigma - \Psi) - \lambda_i^2\Sigma| \\
&= |(1 - \lambda_i^2)\Sigma - \Psi| \\
&= |\Sigma - \xi_i\Psi|
\end{aligned}$$

gdje je $\xi_i = 1/(1 - \lambda_i^2)$, te se maksimalna korelacija postiže uz $\Pi_1 = \beta$. Jednadžba (3.26) se također može napisati kao

$$(\Sigma - \xi_i\Psi)\beta_i = 0 \tag{3.27}$$

za $i = 1, 2, \ldots, r$ zajedničkih faktora. Kanonski korelacijski faktorski model (3.27) je također rješenje Lawleyevih normalnih jednadžbi. Svojstveni parovi (ξ_i, β_i) od Σ

(u metrici od Ψ) daju nam koeficijente α tako da je korelacija između varijabli i zajedničkih faktora maksimalna. Za višedimenzionalan normalan uzorak rješenje od

$$(S - \hat{\xi}_i \hat{\Psi}) P_i = 0$$

su također ML procjenitelji. U specijalnom slučaju kada je $\Psi = \sigma^2 I$ dobivamo model glavnih komponenti sa singularnom $(p \times p)$ matricom Σ ranga r , te možemo zapisati

$$|\Sigma - \frac{1}{(1-\lambda_i^2)i}\Psi| = \begin{cases} |\Sigma - \frac{\sigma^2}{(1-\lambda_i^2)I}| & (1 \leq i \leq r) \\ |\Sigma - \sigma^2 I| & (r < i \leq p). \end{cases}$$

3.4 Općeniti model najmanjih kvadrata

Model najmanjih kvadrata može se poopćiti tako da metodu najmanjih kvadrata koristimo i za minimiziranje težinskog kriterija

$$\begin{aligned} G &= tr[S^{-1}(S-\Sigma)S^{-1}(S-\Sigma)] \\ &= tr[S^{-1}(S-\Sigma)]^2 \\ &= tr(I - S^{-1}\Sigma)^2 \end{aligned}$$

pri čemu je S^{-1} matrica težina. Imamo potpuni diferencijal

$$\begin{aligned} dG &= d[tr(S^{-1}\Sigma - I)^2] \\ &= tr[d(S^{-1}\Sigma - I)^2] \\ &= 2\,tr[(S^{-1}\Sigma - I)d(S^{-1}\Sigma - I)] \\ &= 2\,tr[(S^{-1}\Sigma - I)S^{-1}\,d\Sigma] \end{aligned}$$

te uz Ψ fiksno

$$\begin{aligned} d\Sigma &= d(\alpha\alpha^T + \Psi) \\ &= \alpha(d\alpha^T) + (d\alpha)\alpha^T \\ &= 2\alpha(d\alpha^T) \end{aligned}$$

pa je

$$dG = 4\ tr[(S^{-1}\Sigma - I)S^{-1}\alpha(d\alpha^T)].$$

Koristeći lemu 0.1 iz preliminarija dobivamo

$$\frac{\partial G}{\partial \alpha} = 4S^{-1}(\Sigma - S)S^{-1}\alpha$$

odnosno normalne jednadžbe

$$\begin{aligned} S^{-1}(\hat{\Sigma} - S)S^{-1}\hat{\alpha} &= 0 \\ \hat{\Sigma}S^{-1}\hat{\alpha} &= \hat{\alpha}. \end{aligned} \tag{3.28}$$

Množenjem danih normalnih jednadžbi sa $\hat{\Sigma}^{-1}$ dolazimo do Lawleyevih normalnih jednadžbi (3.19). Uvjetni minimum normalnih jednadžbi (3.28), uz dano Ψ, je različit od uvjetnog minimuma dobivenog iz Lawleyevog ML modela. Ipak, koeficijenti dobiveni iz jednadžbi (3.28) su također konzistentni procjenitelji za α te su asimptotski ekvivalentni Lawleyevim procjeniteljima. Kod općeg model najmanjih kvadrata, kao i kod ML faktorskog modela, pretpostavljamo da je matrica varijanci grešaka dijagonalna kako bi model bio odredljiv. Korištenjem leme 3.2 slijedi da je (3.28) ekvivalentno

$$(\Psi^{1/2}S^{-1}\Psi^{1/2})\Psi^{-1/2}\alpha = \Psi^{-1/2}\alpha(I + \alpha^T\Psi^{-1}\alpha)^{-1}$$

gdje, kao i kod ML modela, možemo uzeti da je $\alpha^T\Psi^{-1}\alpha$ dijagonalna. Tada su stupci od $\Psi^{-1/2}\alpha$ svojstveni vektori od $\Psi^{1/2}S^{-1}\Psi^{1/2}$, a dijagonalni elementi od $(I + \alpha^T\Psi^{-1}\alpha)^{-1}$ su svojstvene vrijednosti. Dakle, budući da se uvjetni minimum od G uz dano $\Psi^{1/2}$ dobije uzimanjem svojstvenih vektora od $\Psi^{1/2}S^{-1}\Psi^{1/2}$ koji pripadaju najmanjim svojstvenim vrijednostima, općeniti model najmanjih kvadrata također se može gledati uz ograničenje da je $\alpha^T S^{-1}\alpha$ dijagonalna.

4 Testovi značajnosti

Kada je $X \sim N(0, \Sigma)$ možemo testirati nultu hipotezu

$$H_0 : \Sigma = \alpha\alpha^T + \Psi$$

naspram alternative

$$H_a : \Sigma \neq \alpha\alpha^T + \Psi$$

kako bi odredili da li Σ sadrži $r > 0$ zajedničkih faktora (i dijagonalnu matricu varijance grešaka).

4.1 χ^2 test

Klasični test za r zajedničkih ML faktora je χ^2 test. Razmatramo gore navedene hipoteze. Pod H_0 funkcija vjerodostojnosti je oblika

$$\begin{aligned} L(\omega) &= c|\hat{\Sigma}|^{-n/2}exp[-\frac{n}{2}tr(\hat{\Sigma}^{-1}S)] \\ &= c|\hat{\alpha}\hat{\alpha}^T + \hat{\Psi}|^{-n/2}exp[tr(\hat{\alpha}\hat{\alpha}^T + \hat{\Psi})S] \end{aligned}$$

a uz H_a imamo

$$\begin{aligned} L(\Omega) &= c|S|^{-n/2}exp[-\frac{n}{2}tr(S^{-1}S)] \\ &= c|S|^{-n/2}exp(-\frac{np}{2}) \end{aligned}$$

jer pod H_a vrijedi $\hat{\Sigma} = S$. Statistika omjera vjerodostojnosti za testiranje H_0 je $\lambda = L(\omega)/L(\Omega)$ pri čemu $-2\,ln\,\lambda$ ima asimptotski χ^2 razdiobu.

Za veliki n, gdje je n veličina uzorka, imamo

$$\begin{aligned}\chi^2 \simeq -2\,ln\,\lambda &= -2\,ln\,L(\omega) + 2\,ln\,L(\Omega) \\ &= n[ln|\hat{\Sigma}| + tr(S\hat{\Sigma}^{-1}) - ln|S| - p].\end{aligned}$$

Budući da iz (3.18) imamo $diag\ \hat{\Psi} = diag\ S$, kriterij se pojednostavljuje do

$$\begin{aligned}\chi^2 &\simeq n[ln|\hat{\Sigma}| + p - ln|S| - p] \\ &= n\,ln\left(\frac{|\hat{\Sigma}|}{|S|}\right).\end{aligned} \tag{4.29}$$

Kako je testiranje potpune nezavisnosti zapravo ekvivalentno testiranju da je $r = 0$, iz (4.29) dobivamo da za velike uzorke vrijedi

$$\chi^2 \simeq -n\,ln\left(\frac{|S|}{|\hat{\Sigma}|}\right) = -n\,ln|R| \tag{4.30}$$

jer kada je $r = 0$ imamo $\hat{\Sigma} = diag(S)$ pa je omjer determinanti jednak $|R|$, determinanti korelacijske matrice, to jest $R = (diag\,S)^{-1/2}S(diag\,S)^{-1/2}$. Statistika (4.30) je valjana samo za velike uzorke. Kada imamo nešto manji uzorak, χ^2 aproksimacija je bolja ako n zamijenimo sa $(n-1) - 1/6(2p+5)$ uz $d = (1/2)[(p-r)^2 - (p+r)]$ stupnjeva slobode, pri čemu su stupnjevi slobode razlika između broja parametara u Σ i broja linearnih ograničenja koja su nametnuta nultom hipotezom. U praksi, test omjera vjerodostojnosti često daje veći broj zajedničkih faktora nego ih se može smisleno interpretirati, stoga je korisno prije sprovesti rotacije nego što se odlučimo za vrijednost od r. Treba paziti da r ne premašuje maksimalnu vrijednost za koju je d pozitivan.

Kada ne odbacujemo pretpostavku potpune nezavisnosti, to jest, kada asimptotska χ^2 statistika ukazuje da postoji barem jedan zajednički faktor, test se ponavlja za veće vrijednosti r budući da je cilj procijeniti "točan" broj zajedničkih faktora $\Phi_1, \Phi_2, \ldots, \Phi_r$ $(1 \leq r < p)$. Kao što smo vidjeli u teoremu 3.2, broj svojstvenih vrijednosti matrice $\Sigma^* = \Psi^{-1/2}\Sigma\Psi^{-1/2}$ jednakih 1 je jednak broju svojstvenih vrijednosti matrice $\Psi^{-1/2}\Gamma\Psi^{-1/2}$ jednakih 0. Dakle, testiranje postojanja $0 < r < p$ zajedničkih faktora je ekvivalentno testiranju da posljednjih $p - r$ svojstvenih vrijednosti $(\hat{\eta}_i + 1)$ od $\Psi^{-1/2}S\Psi^{-1/2}$ je jednako jedan. Statistika za testiranje postojanja r zajedničkih faktora je sada dana sa

$$\chi^2 = -[n - 1 - \frac{1}{6}(2p + 4r + 5)] \sum_{i=r+1}^{p} ln(\eta_i + 1) \tag{4.31}$$

budući da najmanje svojstvene vrijednosti daju mjeru koliko je prilagodba faktorskog modela podacima dobra. Statistika iz (4.31) ima asimptotski χ^2 razdiobu samo kada je $X \sim N(0, \Sigma)$ te kada su uzorci barem umjereno veliki, to jest $n - r \geq 50$. Također, χ^2 test je primjenjiv samo na kovarijacijskim matricama, a ne i na korelacijskim.

4.2 Informacijski kriterij

Jedan od informacijskih kriterija je Akaikeov informacijski kriterij (AIC). Dan je jednadžbom

$$AIC(r) = -2\,ln\,L(r) + 2m \tag{4.32}$$

gdje je m broj slobodnih parametara nakon što je model procijenjen, a $L(r)$ je vjerodostojnost. Kako želimo što veću vjerodostojnost, iz jednadžbe (4.32) je jasno da želimo da AIC bude što manji. Budući da za r zajedničkih faktora imamo $L(r) = (n/2)\sum_{i=r+1}^{p} ln\,\hat{\theta}_i$ i broj slobodnih parametara $m = p(r+1) - (1/2)r(r-1)$, jednadžba (4.33) se može napisati kao

$$AIC(r) = (-2)(\frac{n}{2}\sum_{i=r+1}^{p} ln\,\hat{\theta}_i) + [2p(r+1) - r(r-1)] \tag{4.33}$$

gdje $\hat{\theta}_{r+1}, \hat{\theta}_{r+2}, \ldots, \hat{\theta}_p$ predstavljaju najmanje svojstvene vrijednosti. Ideja korištenja jednadžbe (4.33) je variranje broja zajedničkih faktora, počevši od $r = 1$ te izbor onog r za koji je $AIC(r)$ minimalan.

Kao što je istaknuo Schwarz (1978), $2m$ ne ovisi o vličini uzorka n. To implicira da će jednadžba (4.34) dati isti broj zajedničkihh faktora za male uzorke kao i za velike. Dakle, u tom slučaju $AIC(r)$ kriterij nije konzistentan procjenitelj "točnog" broja faktora r. Schwarzov kriterij koji rješava taj problem se može izraziti kao

$$SIC(r) = -\frac{n}{2}\sum_{i=r+1}^{p} ln\,\hat{\theta}_i + \frac{m}{2}\,ln\,n \tag{4.34}$$

gdje su oznake iste kao u jednadžbi (4.34) samo što je sada uključena i veličina uzorka n. Vrijednost od r se bira tako da je $SIC(r)$ minimalan. Kada je $n > 8$, $SIC(r)$ daje manji broj zajedničkih faktora neko $AIC(r)$.

Primjer 4.0.1. *(vidi [1], str. 388-392)* Imamo podatke (Tablica 1) o $n = 32$ marke automobila i $p = 5$ karakteristika tih automobila. Na tim podacima sprovest ćemo testiranje opisano u danom poglavlju. Budući da je primjer naveden samo u svrhu ilustracije, veličina uzorka i broj varijabli je manji nego što je u većini slučajeva u praksi. Kako je ML faktorski model invarijantan na mjerne transformacije mi ćemo koristiti korelacijsku, a ne kovarijacijsku matricu. Za sprovođenja primjera koritit ćemo statistički program SAS.
Varijable su definirane na sljedeći način

Y_1 = Obujam motora
Y_2 = Snaga motora (ks)
Y_3 = Veličina karburatora (barel)
Y_4 = Masa automobila (lbs)
Y_5 = Vrijeme potrebno da postigne brzinu od 60 milja po satu (sec)

Broj automobila	Y_1	Y_2	Y_3	Y_4	Y_5
1	160.0	110.0	4	2620	16.46
2	160.0	110.0	4	2875	17.02
3	108.0	93.0	1	2320	18.61
4	258.0	110.0	1	3215	19.44
5	360.0	175.0	2	3440	17.02
6	225.0	105.0	1	3460	20.22
7	360.0	245.0	4	3570	15.84
8	146.7	62.0	2	3190	20.00
9	140.8	95.0	2	3150	22.90
10	167.6	123.0	4	3440	18.30
11	167.6	123.0	4	3440	18.90
12	275.8	180.0	3	4047	17.40
13	275.8	180.0	3	3730	17.80
14	275.8	180.0	3	3780	18.00
15	472.0	205.0	4	5250	17.98
16	460.0	215.0	4	5424	17.82
17	440.0	230.0	4	5345	17.42
18	78.7	66.0	1	2200	19.47
19	75.7	52.0	2	1615	18.52
20	71.1	65.0	1	1835	19.90
21	120.1	97.0	1	2465	20.01
22	318.0	150.0	2	3520	16.87
23	304.0	150.0	2	3435	17.30
24	350.0	245.0	4	3840	15.41
25	400.0	275.0	2	3845	17.05
26	79.0	66.0	1	1935	18.90
27	120.3	91.0	2	2140	16.70
28	95.1	113.0	2	1513	16.92
29	351.0	264.0	4	3170	14.50
30	145.0	175.0	6	2770	15.50
31	301.0	335.0	8	3570	14.60
32	121.0	109.0	2	2780	18.80

Tablica 1

U SAS-u pomoću procedure PROC FACTOR izračunamo koeficijente glavnih komponenti, te svojstvene vrijednosti korelacijske matirce za $p = 5$ karakteristika automobila. Podaci su prikazani u Tablici 2, pri čemu je $X_i = Y_i - EY_i$ $(i = 1, 2, \ldots, 5)$.

Varijable	Glavne komponente				
	Z_1	Z_2	Z_3	Z_4	Z_5
X_1	0.878	0.417	-0.207	-0.043	0.107
X_2	0.949	-0.059	-0.106	0.289	-0.040
X_3	0.762	-0.432	0.480	-0.016	0.045
X_4	0.782	0.578	0.172	-0.132	-0.084
X_5	-0.709	0.610	0.308	0.170	0.035
Svojstvene vrijednosti	3.366	1.069	0.409	0.132	0.024
Varijanca (%)	67.31	21.39	8.19	2.64	0.47

Tablica 2

Prvo smo sproveli analizu glavnih komponenti kako bi dobili početnu ideju koliko komponenti nam je potrebno da bi dobro opisali podatke. Prema tablici 2 čini se da su potrebne dvije ili tri komponente. Prvo testiramo potpunu nezavisnost, to jest, da li je prisutan jedan ili više zajedničkih faktora. Koristeći jednadžbu (4.30) imamo aproksimativni χ^2 kriterij

$$\begin{aligned}\chi^2 &= -(n-1)ln|R| \\ &= -31\, ln[(3.366)(1.069)(0.409)(0.132)(0.024)] \\ &= -31\, ln(0.0047) \\ &= -31(-5.3703) \\ &= 166.4798.\end{aligned}$$

Konvergencija je brža ako zamijenimo $(n-1)$ sa $(n-1)-(1/6)(2p+5)$. U našem slučaju mijenjamo 31 sa $31-(1/6)(10+5) = 28.5$, te dobivamo nešto manju vrijednost $\chi^2 = 153.0540$. Imamo

$$\begin{aligned}d &= \tfrac{1}{2}[(p-r)^2 - (p+r)] \\ &= \tfrac{1}{2}[(5-0)^2 - (5)] \\ &= 10\end{aligned}$$

stupnjeva slobode, pa kako je $\chi^2_{0.01,10} = 23.209$ slijedi da su na nivou značajnosti 1% obje χ^2 vrijednosti značajne.

Sljedeće ćemo komentirati proces procjene ML faktorskog modela, počevši od $r = 1$ zajedničkih faktora. Za $r = 1$ dolazimo do poteškoće, dobivamo da samo jedan faktor objašnjava više od 100% ukupne varijance. Do takve pogreške može doći zbog: premalo ili previše zajedničkih faktora, premalo podataka, ako dani model nije prikadan ili nekih drugih razloga. U takvom slučaju trebamo koristiti neki drugi model, najčešće model glavnih komponenti. Budući da ML metodom ne možemo reprezentirati podatke samo jednim zajedničkim faktorom (i greškom), nastavljamo sa $r = 2$ faktora. Podaci za taj slučaj su dani u tablici 3, pri čemu je kao i prije

$X_i = Y_i - EY_i$. Ako usporedimo tablicu 2 i tablicu 3, vidimo da oba modela daju slične vrijednosti koeficijenata.

Varijable	Zajednički faktori ML metodom		R^2
	$\hat{\phi}_1$	$\hat{\phi}_2$	
X_1	0.931	0.229	91.90
X_2	0.902	-0.173	84.32
X_3	0.63	-0.338	51.06
X_4	0.833	0.489	93.34
X_5	-0.639	0.727	93.65

Tablica 3

Svojstvene vrijednosti težinske "reducirane" korelacijske matrice $\hat{\psi}^{-1/2}\hat{\Gamma}\hat{\psi}^{-1/2}$ su $\hat{\eta}_1 = 33.558364, \hat{\eta}_2 = 12.960902, \hat{\eta}_3 = 0.687469, \hat{\eta}_4 = 0.122094$ i $\hat{\eta}_5 = -0.809567$, a svojstvene vrijednosti od $\hat{\psi}^{-1/2}S\hat{\psi}^{-1/2}$ su dane sa $\hat{\theta}_i = (\hat{\eta}_i + 1)$. Iz jednadžbe (4.31) imamo

$$\begin{aligned}\chi^2 &= [31 - \tfrac{1}{6}(10+8+5)]\sum_{i=3}^{5} ln\,\hat{\theta}_i \\ &= -27.167(ln\,1.687469 + ln\,1.122094 + ln\,0.190433) \\ &= -27.167(-1.02) \\ &= 27.71\end{aligned}$$

što aproksimativno ima χ^2 razdiobu sa $d = (1/2)(3^2-7) = 1$ stupnjeva slobode. Kako je $\chi^2_{0.01,1} = 6.635$, vrijednost statistike je još uvijek značajna na nivou značajnosti od 1%. Za sljedeći r, $r = 3$, više ne možemo sprovesti postupak, jer u tom slučaju dobivamo da je broj stupnjeva slobode negativan.

Sada ilustrirajmo korištenje Akaike $AIC(r)$ i Schwarzovog $SIC(r)$ kriterija. Jednom kada su svojstvene vrijednosti $\hat{\theta}_i$ poznate, AIC kriterij se može jednostavno izračunati korištenjem jednadžbe (4.34).

$$\begin{aligned}AIC(r) &= (-2)(\frac{n}{2}\sum_{i=3}^{5} ln\,\hat{\theta}_i) + [2p(r+1) - r(r-1)] \\ &= -32(-1.02) + [10(3) - 2(1)] \\ &= 32.64 + 28 \\ &= 60.64\end{aligned}$$

a prema jednadžbi (4.35) imamo da je Schwarzov kriterij

$$\begin{aligned}SIC(r) &= -\frac{n}{2}\sum_{i=3}^{5} ln\,\hat{\theta}_i + \frac{m}{2} ln\,n \\ &= -16(-1.02) + 7(3.46574) \\ &= 40.58.\end{aligned}$$

Pošto r u našem slučaju može biti jedino 2, ne možemo uspoređivati dana tri kriterija za različite r-ove. U stvarnijim primjerima, kada je p veći pa r može varirati, tada biramo onaj r za koji je vrijednost kirterija najmanja. Općenito, χ^2 statistika daje najveći broj značajnih faktora, a zatim $AIC(r)$ pa $SIC(r)$ kriterij.

4.3 Testiranje koeficijenata

Pod pretpostavkom normalnosti možemo dobiti točne asimptotske druge momente koeficijenata faktora procjenjenih ML metodom. To nam omogućuje testiranje hipoteza oblika

$$H_0 : \alpha_{ij} = 0$$
$$H_a : \alpha_{ij} \neq 0$$

kao i određivanje pouzdanih intervala te procjenjivanje tih parametara. Za model obrnute proporcionalnosti (poglavlje 3.1), može se pokazati (Joreski, 1963) da ako $D \to \Delta$ za $n \to \infty$, da su tada uzoračka varijanca i kovarijanca asimptotski

$$\begin{aligned} nE[(\hat{\alpha}_s - \alpha_s)[(\hat{\alpha}_s - \alpha_s)^T] \sim \frac{\lambda_s}{(\lambda_s - \sigma^2)}\{\Sigma - \frac{\lambda_s}{2(\lambda_s - \sigma^2)}\alpha_s\alpha_s^T + \sum_{j \neq s}\frac{\lambda_s}{(\lambda_j - \sigma^2)} \\ \times [\frac{(\lambda_j - \sigma^2)}{(\lambda_s - \lambda_j)} - 1]\alpha_j\alpha_j^T\} \\ nE[(\hat{\alpha}_s - \alpha_s)[(\hat{\alpha}_t - \alpha_t)^T] \sim -\frac{\lambda_s\lambda_t}{(\lambda_s - \lambda_t)^2}\alpha_s\alpha_t^T \text{ za } s \neq t \end{aligned} \tag{4.35}$$

pri čemu je α_i i-ti stupac matrice koeficijenata α.

Za Lawley-Raove ML procjenitelje razmotrimo vjerodostojnost (jednadžba (3.15)) sa matricom očekivanja drugih derivacija $E[\partial^2 F/\partial\Psi_i\partial\Psi_j] = G$ pri čemu je F kao u lemi 3.1. Također, neka je Λ dijagonalna matrica prvih r svojstvenoh vrijednosti od $\Psi^{-1/2}\Sigma\Psi^{-1/2}$ (teorem 3.2). Ako je uzorak normalan, distirbucija od $\hat{\Psi}$ se približava distribuciji $N[\Psi, (2/n)G^{-1/2}]$. Neka je b_{iq} vektor stupac čiji elementi, $b_{1,iq}, b_{2,iq}, \ldots, b_{p,iq}$, su koeficijenti regresije od $\hat{\alpha}_{iq}$ u odnosu na $\hat{\Psi}_1, \hat{\Psi}_2, \ldots, \hat{\Psi}_p$. Lawley je pokazao (Lawley and Maxwell, 1971.) da je kovarijanca između dva koeficijenta asimptotski

$$nE[(\hat{\alpha}_{is} - \alpha_{is})(\hat{\alpha}_{jt} - \alpha_{jt})^T] \sim -\frac{\lambda_s\lambda_t}{(\lambda_s - \lambda_t)^2}\alpha_{is}\alpha_{jt} + 2b_{is}^T(G^{-1/2})b_{jt} \tag{4.36}$$

pri čemu, ako stavimo $i = j$ i $t = s$, dobivamo asimptotsku varijancu. Regresijski koeficijenti $\hat{b}_{j,iq}$ mogu se izračunati kao

$$\begin{aligned} \hat{b}_{j,iq} = -\hat{\alpha}_{jq}(\hat{\lambda}_q - 1)^{-1}\hat{\psi}_j^{-2} \\ \times [\delta_{ij}\hat{\psi}_j - 1/2\hat{\alpha}_{iq}\hat{\alpha}_{jq}/(\hat{\lambda}_q - 1) + \hat{\lambda}_q\sum_{h \neq q}^{r}\hat{\alpha}_{ih}\hat{\alpha}_{jh}/(\hat{\lambda}_q - \hat{\lambda}_h)] \end{aligned}$$

za r zajedničkih faktora, pri čemu je δ_{ij} Kroneckerov delta. Budući da je ML faktorski model invarijantan na skaliranje, bez smanjena općenitosti možemo pretpostaviti da koristimo kovarijacijsku matricu. Tada za test koristimo normalne tablice jer se $\sqrt{n}(\hat{\alpha}-\alpha)$ približava multivarijantnoj normalnoj razdiobi sa očekivanjem nula i slijedi da je $\hat{\alpha}$ konzistentan procjenitelj od α.

Za prethodne testove značajnosti nužna je pretpostavka multivarijantne normalnosti. Kao i za model glavnih komponenti, također se mogu koristiti metode ponovnog uzorkovanja, kao što su *jackknife*, bootstrap ili krosvalidacija, za testiranje parametara te koliko je model dobar. Potencijalni problem za *jackknife* i bootstrap procjene je da one mogu dati nekonzistentne procjenitelje koeficijenata (na primjer, korelacijske koeficijente veće od jedan) i za obje metode je potrebno duže vrijeme za izvođenje nego za parametarske testove zasnovane na pretpostavci normalnosti. Iz tih razloga, motode uzorkovanja se ne koriste često za faktorske modele. Metoda krosvalidacije se čini atraktivna, no nije još dovoljno istražena.

Nekoliko narednih komentara je vezano uz teorem s početka rada, teorem 1.1, koji govori o nužnim i dovoljnim uvjetima postojanja $1 \leq r < p$ zajedničkih faktora. Spomenuti teorem se može smatrati temeljnim teoremom faktorske analize. Kao što smo vidjeli u prethodnim poglavljima, uvjet je da postoji pozitivno definitna dijagonalna matrica Ψ, takva da je $\alpha\alpha^T = \Gamma$ pozitivno semidefinitna matrica ranga r. Ako su zadovoljena identifikacijska ograničenja, tako dana matrica Ψ je uvjet koji garantira postojanje dekompozicije $\Sigma = \Gamma + \Psi$ za $1 \leq r < p$. U praksi, Ψ gotovo nikada nije poznata, pa treba biti procijenjena zajedno sa koeficijentima α. Vidjeli smo u primjeru 4.0.1, da istovremeno postojanje $1 \leq r < p$ zajedničkih faktora te Gramovih matrica Γ i Ψ općenito govoreći nije garantirano. Posljedice takvih nepravilnih rješenja su: gornja granica za broj zajedničkih faktora koja proizlazi iz jednadžbe (1.6) je samo nužan uvjet, uzorački koeficijenti su nekonzistentni procjenitelji za populacijske parametre. Jedan od načina izlaza iz takvih teškoća je popuštanje oko pozitivne (semi)definitnosti matrica $\Gamma = \alpha\alpha^T$ i Ψ. Također, kada dobijemo da jedan sam faktor pojašnjava više od 100% varijance, to bi mogao biti znak da za dani slučaj faktorski model nije prikladan i vjerojatno bi trebao biti zamjenjen sa modelom glavnih komponenti.

5 Procjena faktora

Faktorski modeli iz prethodnih poglavlja procjenjuju korelacijske koeficijente α, koji se uglavnom mogu dobiti iz korelacijske ili kovarijacijske matrice. To često predstavlja primarni cilj faktorske analize budući da ti koeficijenti određuju manji skup zajedničkih faktora od svih promatranih varijabli te omogućuju identifikaciju i interpretaciju faktora. Iako općenito koeficijenti variraju za svaku varijablu, ipak su nepromjenljivi za svaki element uzorka, pa u tom smislu ne predviđaju relativni položaj danog elementa uzorka u odnosu na faktore. U ortogonalnom faktorskom modelu (jednadžba 1.2) uz poznate (procijenjene) r, α i Ψ, postoji dodatna neodređenost zbog nejedinstvenosti faktora Φ koja je uzrokovana singularnošću matrice α. Završni

korak u faktorskoj analizi matrice podataka je pronalazak optimalne procjene faktora Φ. Za procjenu se koriste dva različita pristupa, ovisno o tome da li razmatramo Φ fiksan ili slučajan.

5.1 Slučajni faktori: regresijski procjenitelj

Procjenitelj faktora Φ može se izvesti u kontekstu teorije maksimalne vjerodostojnosti uz pretpostavku zajedničke normalnosti Φ i X. Razmotrimo prošireni vektor $Z = [\Phi^T \ X^T]^T$, koji po pretpostavci ima $(r+p)$ dimenzionalnu normalnu razdiobu sa vektorom očekivanja $\mu = 0$ i kovarijacijskom matricom

$$\begin{aligned}\Sigma = \begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix} &= E(ZZ^T) = E\begin{bmatrix}\Phi\\ X\end{bmatrix}[\Phi^T X^T]\\ &= \begin{bmatrix}E(\Phi\Phi^T) & E(\Phi X^T)\\ E(X\Phi^T) & E(XX^T)\end{bmatrix}.\end{aligned}$$

Prema teoremu 0.2 iz preliminarija slijedi da je uvjetna distribucija od Φ uz dano X također normalna, to jest

$$\begin{aligned}\Phi|X &\sim N[(\Sigma_{12}\Sigma_{22}^{-1}X),\ (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T)]\\ &\sim N[\alpha^T\Sigma^{-1}X,\ (I - \alpha^T\Sigma^{-1}\alpha)]\end{aligned}$$

gdje je $E(\Phi X^T) = \alpha^T$. Koristeći lemu 3.3 možemo kovarijacijsku matricu od $\Phi|X$ prikazati i u drugačijem obliku.

Lema 5.1. *Ako su zadovoljeni uvjeti teorema 3.2, tada vrijedi*

$$I - \alpha^T\Sigma^{-1}\alpha = (I + \alpha^T\Psi^{-1}\alpha)^{-1}. \tag{5.37}$$

Dokaz. Iz leme 3.3 slijedi

$$\begin{aligned}\alpha^T\Sigma^{-1}\alpha &= \alpha^T[\Psi^{-1}\alpha(I+\eta)^{-1}]\\ &= \eta(I+\eta)^{-1}\end{aligned}$$

gdje je $\eta = \alpha^T\Psi^{-1}\alpha$. Sada imamo

$$\begin{aligned}I - \alpha^T\Sigma^{-1}\alpha &= I - \eta(I+\eta)^{-1}\\ &= (I+\eta)(I+\eta)^{-1} - \eta(I+\eta)^{-1}\\ &= (I+\eta)^{-1}\\ &= (I+\alpha^T\Psi^{-1}\alpha)^{-1}.\end{aligned}$$

□

Dakle, uvjetna distribucija od Φ uz dano X se također može napisati kao

$$\Phi|X \sim N[\alpha^T\Sigma^{-1}X,\ (I+\alpha^T\Psi^{-1}\alpha)^{-1}]. \tag{5.38}$$

Imamo

$$\begin{aligned} E(\Phi|X) &= \alpha^T \Sigma^{-1} X \\ &= (I + \alpha^T \Psi^{-1} \alpha)^{-1} \alpha^T \Psi^{-1} X. \end{aligned} \tag{5.39}$$

Kako bi dobili ML procjenitelj za Φ gledamo $(n \times p)$ matricu podataka $\mathbb{X}$, pri čemu su koeficijenti α poznati (procijenjeni sa $\hat{\alpha}$). Uzorački analogon jednadžbe (5.39) je

$$F = \mathbb{X}\hat{\Psi}^{-1}\hat{\alpha}(I + \hat{\alpha}^T \hat{\Psi}^{-1} \hat{\alpha})^{-1} \tag{5.40}$$

pri čemu je

$$F = \begin{bmatrix} F_1^T \\ F_2^T \\ \vdots \\ F_n^T \end{bmatrix}$$

i $\hat{\Psi} = \hat{\epsilon}^T \hat{\epsilon}$ dijagonalna matrica varijanci reziduala. Neka je

$$F = \mathbb{X}B + \delta \tag{5.41}$$

gdje je B $(p \times r)$ matrica koeficijenata. Za zadano F, B se može procijeniti linearnom regresijom, to jest

$$\hat{B} = (\mathbb{X}^T \mathbb{X})^{-1} \mathbb{X}^T F$$

što se dobije minimiziranjem $tr(\delta^T \delta)$ po B.

Neka je X neki novi podatak. Sada Φ možemo procijeniti sa

$$\hat{\Phi} = \hat{B}^T X$$

pa imamo

$$\begin{aligned} \hat{\Phi}^T &= X^T \hat{B} \\ &= X^T (\mathbb{X}^T \mathbb{X})^{-1} \mathbb{X}^T F \\ &= X^T (\mathbb{X}^T \mathbb{X})^{-1} \hat{\alpha} \end{aligned}$$

Korištenjem jednadžbe (3.21) iz leme 3.3 dobivamo da vrijedi

$$\begin{aligned} \hat{\Phi}^T &= X^T S^{-1} \hat{\alpha} \\ &= X^T \hat{\Psi}^{-1} \hat{\alpha} (I + \hat{\alpha}^T \hat{\Psi}^{-1} \hat{\alpha})^{-1}. \end{aligned} \tag{5.42}$$

imajući na umu da radimo sa uzoračkim vrijednostima, te uz ograničenje da je $\hat{\alpha}\hat{\Psi}^{-1}\hat{\alpha}^T$ dijagonalna. Dakle, kada su opažane varijable i faktori po pretpostavci normalno distribuirani, dani regresijski procjenitelj je ekvivalentan ML procjenitelju. ML procjenitelj (jednadžba (5.42)) je asimptotski efikasan, iako nije nepristran. Kada nemamo pretpostavku normalnosti od X, procjenitelj (5.42) je optimalan u smislu najmanjih kvadrata. Procjenitelj je međutim pristran, jer iako jednadžba (5.42) pretpostavlja da je greška u Φ, u stvarnosti greška mjerenja pripada X; regresijski procjenitelj mijenja ulogu zavisnih i nezavisnih varijabli. Kada $p \to \infty$, tada se uzoračka varijanca od $\hat{\Phi}$ smanjuje. To isto vrijedi za faktore koji su povezani sa velikim vrijednostima u matrici $\hat{\alpha}\hat{\Psi}\hat{\alpha}^T$, to jest faktori koji objašnjavaju veći dio varijance mogu se preciznije odrediti nego oni koji objašnjavaju manji dio varijance.

5.2 Fiksni faktori: procjenitelj minimalne udaljenosti

Procijenitelj (5.40) od Φ pretpostavlja da je Φ slučajan i procjenjuje ga nakon što su X i α poznati. Na faktore se može gledati i na drugačiji način, pri čemu se koristi pretpostavka da je svaki element uzorka karakteriziran sa fiksnim vektorom od r parametara Φ_i. Tada za pojedini i $(i = 1, 2, \ldots, n)$ imamo $x_i \sim N(\alpha\Phi_i, \Psi)$ pri čemu je $\alpha\Phi_i + \epsilon_i$ $(p \times 1)$ vektor observacije p varijabli. Sada, dok su Φ_i fiksni populacijski parametri, pogodnije je tražiti linearne procjenitelje koji su nepristrani i daju najmanju varijancu.

Neka je x_i $(p\times 1)$ vektor opservacije za i-ti element uzorka takav da je $x_i = \alpha\Phi_i + \epsilon_i$, pri čemu pretpostavljamo da je α poznat. Tada je razuman kriterij minimizacija težinske udaljenosti između x_i i njegove predviđene vrijednosti $\hat{x}_i = \alpha\Phi_i$, to jest minimizacija

$$\begin{aligned} d_i &= (x_i - \alpha\Phi_i)^T \Psi^{-1}(x_i - \alpha\Phi_i) \\ &= x_i^T\Psi^{-1}x_i - 2x_i^T\Psi^{-1}\alpha\Phi_i + \Phi_i^T\alpha^T\Psi^{-1}\alpha\Phi_i. \end{aligned} \tag{5.43}$$

Uočimo da d_i također reprezentira težinsku sumu kvadrata reziduala.
Diferenciranjem s obzirom na nepoznate parametre Φ_i, te izjednačavanjem normalnih jednadžbi sa nula dobivamo

$$\frac{\partial d_i}{\partial \Phi_i} = -2\alpha^T\Psi^{-1}x_i + 2\alpha^T\Psi^{-1}\alpha\Phi_i = 0$$

odnosno

$$\tilde{\Phi}_i = (\alpha^T\Psi^{-1}\alpha)^{-1}\alpha^T\Psi^{-1}x_i. \tag{5.44}$$

Minimiziranje jednadžbe (5.43) je također ekvivalentno minimiziranju

$$tr[(x_i - \alpha\Phi_i)^T\Psi^{-1}(x_i - \alpha\Phi_i)].$$

Uočimo da je procijenitelj

$$\hat{\Phi}_i = (\alpha^T\alpha)^{-1}\alpha^T x_i \tag{5.45}$$

dobiven metodom najmanjih kvadrata netočno određen kada je $\Psi \neq \sigma^2 I$ i da je $var(\hat{\Phi}_i) > var(\tilde{\Phi}_i)$. Rješenja za $\tilde{\Phi}_i$ su dobivena korištenjem generaliziranog inverza $(\alpha^T\Psi^{-1}\alpha)^{-1}\alpha^T\Psi^{-1}$. Iako procjenitelj (5.44) ima poželjna svojstva, nepristranost i efikasnost, on pretpostavlja da su faktori parametri a ne slučajne varijable što može dovesti do poteškoća prilikom ML procjene. Uočimo također da su Φ_i korelirani, osim ako je $(\alpha^T\Psi^{-1}\alpha)$ dijagonalna.

Može se također uvesti ograničenje da su faktori ortogonalni. Minimiziranjem

$$tr[(x_i - \alpha\Phi_i)^T\Psi^{-1}(x_i - \alpha\Phi_i)] - \lambda(\Phi_i^T\Phi_i - 1) \tag{5.46}$$

slijedi da je tada procijenitelj za ML faktore

$$\Phi_i^* = [(\alpha^T\Psi^{-1}\alpha)(I + \alpha^T\Psi^{-1}\alpha)]^{-1/2}\alpha^T\Psi^{-1}x_i. \tag{5.47}$$

Procijenitelj (5.47) ima veće očekivanje kvadrata greške nego procijenitelj (5.39).

Kada su α i Ψ poznati, procijenitelji fiksnih faktora su također ML procijenitelji. Kada se α i Ψ procjenjuju, tada ML procijenitelji faktora ne postoje jer normalne jednadžbe nemaju minimuma.

Poglavlje III

Primjer

Na primjeru ćemo demonstrirati sprovođenje faktorske analize. Primjer ćemo sprovesti na podacima iz ekologije. Za sprovođenje primjera koristi se softver SAS.

Primjer 2.0.2. *[2]* Singa i Lee (1970) razmatrali su poljoprivrednu ekologiju. Uzeli su složene uzorke pšenice, zobi, ječma i raži sa različitih lokacija Kanadske prerije. Cilj istraživanja je bio odrediti vezu, ako postoji, između prisustva člankonožaca i okoline zrna. Opaženo je sljedećih 9 varijabli na $n = 165$ uzorka:

Y_1 = Ocjena uzroka s obzirom na kvalitetu zrna (1 najveća, 6 najmanja)
Y_2 = Vlaga zrna (postotak)
Y_3 = Prisutnost korova, slomljenih zrna i ostalih stranih tvari
Y_4 = Broj pronađenih člankonožaca *Acarus* u zrnu
Y_5 = Broj pronađenih člankonožaca *Cheyletus* u zrnu
Y_6 = Broj pronađenih člankonožaca *Glycyphagus* u zrnu
Y_7 = Broj pronađenih člankonožaca *Larsonemus* u zrnu
Y_8 = Broj pronađenih člankonožaca *Cryptolestes* u zrnu
Y_9 = Broj pronađenih člankonožaca *Psocoptera* u zrnu

Varijable Y_1 i Y_3 su transformirane tako da je uzet korijen pravih vrijednosti. Dok su varijable $Y_4 - Y_9$ transformirane sa log funkcijom (s bazom 10). U nastavku dalje radimo sa transformiranim vrijednostima te za transformirane varijable koristimo iste oznake kao za netransformirane. U tablici 4 je dana korelacijska matrica podataka.

Korelacijska matrica $p = 9$ varijabli									
Varijable	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9
Y_1	1.000								
Y_2	0.441	1.000							
Y_3	0.441	0.342	1.000						
Y_4	0.107	0.250	0.040	1.000					
Y_5	0.194	0.323	0.060	0.180	1.000				
Y_6	0.105	0.400	0.082	0.123	0.220	1.000			
Y_7	0.204	0.491	0.071	0.226	0.480	0.399	1.000		
Y_8	0.197	0.158	0.051	0.019	0.138	-0.114	0.154	1.000	
Y_9	-0.236	-0.220	-0.073	-0.199	-0.084	-0.304	-0.134	-0.096	1.000

Tablica 4

Sprovedimo ML faktorsku analizu. Iz jednadžbe (1.6) sljedi da maksimalni broj faktora može biti 5. Pomoću softvera SAS izvrtit ćemo ML metodu za $r = 1, 2, \ldots, 5$ te usporediti rezultate. Za sve naredne rezultate koristi se PROC FACTOR procedura.
Prvo, koristeći χ^2 test, testirajmo nultu hipotezu

$$H_0 : \text{Nema zajedničkih faktora}$$

naspram alternative

$$H_a : \text{Postoji barem jedan zajednički faktor.}$$

Dobivamo da je vrijednost statistike 267.485. P-vrijednost uz 36 stupnjeva slobode je manja od 0.0001. Dakle, odbacujemo hipotezu H_0, da nema zajedničkih faktora. Sljedeće, testirajmo hipoteze:

$$H_0 : r \text{ faktora je dovoljno}$$
$$H_a : \text{Potrebno je više od } r \text{ faktora.}$$

Pregled vrijednosti test statistika, p-vrijednosti te AIC statistika dan je u tablici 5.

r	χ^2	df	p-vrijednost	AIC
1	82.962	27	<0.0001	31.303
2	36.636	19	0.0088	-0.173
3	12.503	12	0.4062	-11.036
4	4.014	6	0.6748	-7.821
5	0.615	1	0.4329	-1.357

Tablica 5

Budući da je p-vrijednost za $r \leq 2$ jako mala, zaključujemo da trebamo više od 2 glavna faktora. Za $r = 3, 4$ i 5 ne odbacujemo pretpostavku da je to dovoljan broj

faktora, pa sada gledamo vrijednosti AIC statistike kako bi odabrali najbolji r. Kao što smo rekli u četvrtom poglavlju, bira se r za koji je AIC najmanji. Dakle, prema tablici 5, vidimo da su nam dovoljna tri glavna faktora za opisati podatke. Pomoću procedure PROC FACTOR dobili smo koeficijente faktora, odnosno matricu α. Dobiveni koeficijenti, bez rotacije i sa ortogonalnom (varimax) rotacijom, su prikazani u tablici 6.

ML koeficijenti						
	Nerotirani koeficijenti			Rotirani koeficijenti		
	1	2	3	1	2	3
Y_1	0.105	0.591	-0.464	0.203	0.731	0.029
Y_2	0.4	0.616	-0.109	0.508	0.473	0.264
Y_3	0.082	0.372	-0.45	0.038	0.586	0.054
Y_4	0.123	0.26	0.075	0.275	0.102	0.052
Y_5	0.22	0.461	0.26	0.562	0.08	0.078
Y_6	1	0	0	0.253	0.035	0.967
Y_7	0.399	0.607	0.424	0.816	0.044	0.197
Y_8	-0.114	0.327	0.008	0.226	0.188	-0.184
Y_9	-0.304	-0.16	0.157	-0.106	-0.232	-0.278

Tablica 6

Dakle, nerotirani ML koeficijenti impliciraju da je prisustvo člankonožaca *Glychipagusa* povezano sa faktorom 1. Također vidimo da je prisutnost *Psocoptera* negativno povezana sa svim faktorima, te vidimo da drugi faktor ima najveće vrijednosti koeficijenata za ocjenu, vlagu te prisutnost *Tarsonemusa*. Treći faktor je negativno koreliran sa ocjenom, vlagom, prisutnošću korova, slomljenih zrna i ostalih stranih tvari.
Sa rotacijom dobivamo da je faktor 1 ima najveće koeficijente za vlagu, te prisutnost *Glychipagusa* te *Cheyletusa*, te sada, za razliku od prije, dobivamo da je 3. faktor pozitivno povezan sa svim varijablama. Također, sada je *Cryptolestes*, iako ne naročito značajno, negativno povezan sa 3. faktorom, te prisustvo *Glychipagusa* je uglavnom povezano sa istim faktorom. Kako nam je SAS prilikom izvođenja javio grešku " Communality greater than 1.0", dane rezultate moramo uzeti s oprezom.

Sprovedimo sada na danim podacima faktorsku analizu glavnih komponenti za $r = 3$. Vrijednosti rotiranih koeficijenata za tu metodu su dani u tablici 7.

ML koeficijenti			
	Rotirani koeficijenti		
	1	2	3
Y_1	0.196	0.762	0.025
Y_2	0.450	0.497	0.283
Y_3	0.024	0.561	0.061
Y_4	0.269	0.097	0.111
Y_5	0.576	0.079	0.065
Y_6	0.254	0.041	0.893
Y_7	0.795	0.055	0.208
Y_8	0.232	0.185	-0.184
Y_9	-0.133	-0.217	-0.266

Tablica 7

Vidimo da se rezultati ne razlikuju jako od rezultata dobivenih ML metodom za rotirane koeficijente. Vidimo da faktor 1 ima njaveće koeficijente za prisustvom člankonožaca *Cheyletus* i *Tarsonemus*, koeficijenti 0.576 i 0.795, pa bi prvi faktor možda mogli interpretirati kao razvoj spomenutih člankonožaca. Drugi faktor ima znatno najveće vrijednosti koeficijenata (0.762, 0.497, 0.561) za prve tri varijable pa bi ga mogli interpretirati kao okolina zrna. Treći faktor ima veliki koeficijent samo za jednu varijablu, Y_6, koja je broj prisutnih *Glychipagusa* pa dani faktor možemo interpretirati kao razvoj *Glychipagusa*. Preostale varijable, koje nismo spomenuli, nemaju naročitog utjecaja na razvoj zrna.

Bibliografija

[1] A. Basilevsky, *Statistical Factor Analysis and Related Methods: Theory and Applications*, John Wiley and Sons, 1994.

[2] R. Khattree i D.N. Naik, *Multivariate Data Reduction and Discrimination*, John Wiley and Sons, 2000.

[3] N. Sarapa, *Teorija vjerojatnosti*, Školska Knjiga, 2002.

Zahvale

Zahvaljujem mentoru rada prof. dr. sc. Miljenku Huzaku na iznimnoj strpljivosti i velikoj pomoći tijekom izrade diplomskog rada.

Posebno zahvaljujem obitelji i Roku na neizmjernoj podršci i razumijevanju tijekom studija.

Također, zahvaljujem svim prijateljima na pomoći i velikoj podršci.

Sažetak

U ovom radu, uveli smo osnovni faktorski model te predstavili faktorske modele glavnih komponenti i faktorske modele maksimalne vjerodostojnosti. Vidjeli smo da je polazna točka faktorske analiza kovarijacijska (korelacijska) matrica podataka. Određivanje faktorskog modela se svodi na određivanje matrice koeficijenata te zajedničkih faktora. Uočili smo da se modeli međusobno razlikuju po dodatnim pretpostavkama i načinu određivanja matrice koeficijenata.
Nakon predstavljanja modela, obradili smo testove za testiranje da li postoje zajednički faktori te koliko ih ima. Za to se koristi χ^2 test te Akaikeov (AIC) i Schwarzov (SIC) informacijski kriterij. Spomenuto smo potkrijepili primjerom. Nakon toga smo opisali procjenu faktora. Ona se razlikuje s obzirom da li na faktore gledamo kao na slučajne varijable ili fiksne parametre.
Na kraju smo dali primjer u kojem smo sproveli faktorsku analizu.

Summary

In this work, we have introduced the general factor model and presented principal components factor models and maximum likelihood factor models. We have seen that the starting point of the factor analysis is the data covariance (correlation) matrix. Determining the factor model comes down to determining coefficient matrix and the common factors. We have observed that there si a difference between various factor models in additional constraints and the method of determining the coefficient matrix.
After presenting the models we elaborate tests for testing if common factors exist and how many common factors there are. For that purpose, we used χ^2 test and Akaike (AIC) and Schwarz (SIC) information criteria. We showed mentioned in the example. After that we have described the factor estimation. It differs depending on whether we consider the factors as random variables or fixed parameters.
Finally, we gave an example where we implemented a factor analysis.

Životopis

Rođena sam 2. ožujka 1992. godine u Čakovcu. Nakon završene osnovne škole upisala sam gimnaziju Josipa Slavenskog Čakovec, prirodoslovno-matematički smjer. Maturirala sam 2010. godine.
2010. godine sam upisala preddiplomski studij matematike na Matematičkom odsjeku Prirodoslovno-matematičkog fakulteta u Zagrebu. Preddiplomski studij sam završila 2013. godine. Te godine sam upisala diplomski sveučilišni studij Matematička statistika na istom fakultetu. Dobitnica sam petogodišnje stipendije tvrtke Procter & Gamble d.o.o..