

De onfatsoenlijke matrix

H. Koppelaar, B. Niemöller

Inleiding

Een praktijkprobleem bij het hanteren van factoranalyse is: is de correlatiematrix geschikt voor deze analyse? In moderne terminologie: "is de correlatiematrix embeddable?". We bespreken een viertal tests voor embeddability. Geen van de vier is geïmplementeerd in de CDC-versie van het SPSS. De test die het SPSS thans geeft, is slechts de berekening van de determinant van de correlatiematrix. We tonen in deze paper in de vierde test, de zogenaamde "conditietest" aan, dat de determinant van een matrix heel goed kan zijn, d.w.z. $\det \neq 0$, terwijl de conditie van die matrix slecht is! Dit resultaat is te vinden in voorbeeld 3 van de paragraaf Conditiegetallen, in het vervolg. Voordat het echter zover is, bespreken we eerst drie embeddability-tests voor factoranalyse aan de hand van Dziuban en Shirkey. Daarna leggen we uit wat matrixnormen zijn, om vervolgens onze nieuwe test op de conditie van een matrix te illustreren en toe te passen.

Bartlett's sfericiteits toets

De nul-hypothese luidt: voor de populatie waaruit de steekproef getrokken is, geldt dat de variabelen niet gecorreleerd zijn, dus voor de populatie correlatiematrix P geldt $P = I$. Dus voor de steekproef correlatiematrix R geldt dat de correlaties niet significant van nul verschillen. Bartlett's toets is bij benadering een chi-kwadraat toets van de nulhypothese

$$H_0 : |P| = 1$$

De waarde van de determinant van de correlatiematrix van de populatie wordt uiteraard slechts één als $P = I$. Voor $|P|$ gebruiken we als schatting $|R|$.

Het toetskriterium is nu

$$- [(N-1) - \frac{1}{6} (2p+5)] e \log |R|$$

waarin N de steekproefgrootte, p het aantal variabelen en $|R|$ de determinant van de steekproef correlatiematrix is.

Voor grote N is de grootte bij benadering chi-kwadraat verdeeld met $\frac{1}{2}p(p-1)$ vrijheidsgraden.

Cooley en Lohnes (1971) geven, zonder volledigheid na te streven, enig inzicht in de 'redelijkheid' van Bartlett's toets:

1. $|R|$ ligt altijd tussen 1.00 en 0.00 als $|R| \rightarrow 1$ dan $R \rightarrow I$ als $R \rightarrow I$ dan wordt sphericiteit benaderd.
2. $e \log$ van decimale frakties (tussen 0.00 en 1.00) zijn negatief met als range $-\infty$ tot 0.00.
als $|R| \rightarrow 1.00$ dan $e \log |R| \rightarrow 0.00$ en $\chi^2 \rightarrow 0.00$
3. een t.o.v. p groter N doet voor een gegeven $|R|$ de χ^2 toenemen; een grote N/p ratio verhoogt de power van de test.

Een niet-signifikant χ^2 zou er toe moeten leiden de betreffende matrix niet te factoriseren, zoals bijvoorbeeld in onderstaand voorbeeld van Cooley en Lohnes

$$X = \begin{pmatrix} 7 & 4 & 3 \\ 4 & 1 & 8 \\ 6 & 3 & 5 \\ 8 & 6 & 1 \\ 8 & 5 & 7 \\ 7 & 2 & 9 \\ 5 & 3 & 3 \\ 9 & 5 & 8 \\ 7 & 4 & 5 \\ 8 & 2 & 2 \end{pmatrix} \rightarrow R = \begin{pmatrix} 1.00 & 0.67 & -0.10 \\ 0.67 & 1.00 & -0.29 \\ -0.10 & -0.29 & 1.00 \end{pmatrix}$$

en dus $\chi^2 = -(10-1-\frac{5}{6})e \log .499 = -(7.16)(-.695) = 4.99$ met $\frac{1}{2} \cdot 3(3-1) = 3$ vrijheidsgraden. Deze waarde is zelfs op 10% nivo niet significant.

Dziuban en Shirkey wijzen erop dat de toets op sfericiteit als een ondergrens voor de kwaliteit van de matrix kan worden gebruikt. D.w.z. bij aanvaarding van de nulhypothese kan men van verdere (factor-)analyse afzien. Bij verwerping van H_0 dient men echter bedacht te zijn op eventuele invloed van de steekproefomvang, die bij de chi-kwadraat voor grote steekproeven groot is.

Inspektie van de buiten-diagonale elementen van de anti-image covariantie matrix

Subjektiever is een tweede manier waarbij de buiten-diagonale elementen van de anti-image covariantie matrix $S^2 R^{-1} S^2$ worden bekeken. Hierbij is $S^2 = (\text{diag } R^{-1})^{-1}$ een diagonale matrix met als elementen de reciproken van de elementen uit de hoofddiagonaal van R^{-1} , d.i. de inverse van de correlatiematrix R . Uit de literatuur is bekend dat bij toepassing van het factor-analytisch model voor de data moet gelden dat R^{-1} ongeveer diagonaal moet zijn. Ook $S^2 R^{-1} S^2$, de matrix van covarianties van de unieke delen van de variabelen, zou bij benadering een diagonale matrix moeten zijn.

Met andere woorden, de buiten-diagonale elementen dienen zoveel mogelijk nul te zijn. Zoals Dziuban en Shirkey opmerken is hier de grote vraag: 'Hoe diagonaal is diagonaal'.

Hun antwoord kan niet veel vager: "Perhaps it seems reasonable that if say 25% of those elements are nonzero, one has reason to suspect the psychometric quality of his correlation matrix" (Dziuban en Shirkey, 1974, 360).

Measure of sampling adequacy

Een derde procedure wordt uitvoeriger beschreven door Kaiser (1970) en omvat de berekening van de 'Kaiser-Meyer-Olkin measure of sampling adequacy'. Deze index kan zowel voor het geheel van alle variabelen als voor een enkele variabele berekend worden en luidt in formulevorm:

$$\frac{\sum_{j \neq k} \sum_k r_{jk}^2}{\sum_{j \neq k} \sum_k r_{jk}^2 + \sum_{j \neq k} \sum_k q_{jk}^2} \quad \text{resp.} \quad \frac{\sum_k r_{jk}^2}{\sum_{k \neq j} r_{jk}^2 + \sum_{k \neq j} q_{jk}^2}$$

Hierin zijn de q^2 -en de gekwadrateerde buiten-diagonale elementen van de anti-image correlatie matrix $SR^{-1}S$ en de r^2 -en de gekwadrateerde buiten-diagonale elementen van de oorspronkelijke correlaties.

Als vuistregel voor de index, die tussen nul en een ligt, geeft Kaiser:

in the .90s - marvelous
 in the .80s - meritorious
 in the .70s - middling
 in the .60s - mediocre
 in the .50s - miserable
 below .50 - unacceptable

Conditie-toets

Indien de determinant van een matrix een waarde dichtbij nul heeft, zegt men dat die matrix 'slecht geconditioneerd' is. De conditie van een matrix A kan in een getal $C(A)$ uitgedrukt worden. Dit conditiegetal is dan 'een maat voor de onfatsoenlijkheid' van A , t.a.v. sommige rekenkundige bewerkingen met A ; de definitie luidt:

$$C(A) = \| A \| \| A^{-1} \|$$

Bijvoorbeeld voor het oplossen van het stelsel vergelijkingen $Ax = b$ geeft C aan hoe de relatieve fout in A en b door kan werken in de onbekende x .

De normen $\| \cdot \|$ zijn matrixnormen.

Matrixnormen

Om in te zien wat matrixnormen zijn, beschouwen we een afbeelding van een reële lineaire ruimte X in zichzelf, deze afbeelding heeft nl. een lineaire operator O . Bijvoorbeeld kan een O op X bij vast basis van n -dimensies worden gerepresenteerd door een $n \times n$ - matrix. Een norm op X is een afbeelding naar de positieve reële getallen $\| \cdot \|: X \rightarrow R_+$, met de eigenschappen:

$$\|x\| = 0 \Leftrightarrow x = 0 \text{ voor } x \in X \quad (1)$$

$$\|\lambda x\| = |\lambda| \|x\| \text{ voor } \lambda \in \mathbb{R} \text{ en } x \in X \quad (2)$$

$$\|x + y\| \leq \|x\| + \|y\| \text{ voor } x, y \in X \quad (3)$$

Bekende normen op X zijn de Höldernormen:

$$\|x\|_p = \left[\sum_{i=1}^n |x_i|^p \right]^{1/p} \text{ met } x = [x_1, x_2, \dots, x_n] \text{ en } p \geq 1.$$

We nemen hier X dus op een basis van n -dimensies. Deze Hölder-normen zijn een generalisatie van de aloude kwadratische of euklidische norm:

$$\text{Voor } p = 2 \text{ geldt } \|x\|_2 = \left[\sum_{i=1}^n |x_i|^2 \right]^{1/2}$$

De lineaire norm krijgen we met $p = 1$:

$$\|x\|_1 = \sum_{i=1}^n |x_i|$$

(Het zal duidelijk zijn dat de zg. 'kleinste kwadraten aanpassing' berust op de Höldernorm met $p = 2$ en dus een van de vele is uit de verzameling van Höldernormen).

Als de ruimte X een norm heeft en als X in zichzelf afgebeeld wordt, dan heeft de afbeelding (= deelruimte van X) dezelfde norm. Om nu een matrixnorm te induceren, nemen we een lineaire operator O op X met basis $e_1, e_2, \dots, e_n$, waarbij O gerepresenteerd wordt d.m.v. een matrix A :

A is een matrix (a_{ij}) , $i, j = 1, \dots, n$

Neem nu $Ax = b$, met $x \in X$, dan vinden we:

$$\|A\|_1 = \max_j \sum_i |a_{ij}|, \text{ dat is de maximale kolom absoluut som.}$$

Dit is als volgt te bewijzen. De afbeelding $A : x \rightarrow b$ induceert de $\| \cdot \|_1$ norm op A , doordat:

$$\|b\|_1 = \|Ax\|_1 \leq \max_j \sum_i |a_{ij}| \cdot \|x\|_1$$

Nemen we voor x de basisvector e_k , $1 \leq k \leq n$, waarvoor het maximum wordt aangenomen, dan geldt het gelijktteken voor $x = e_k$.

Omdat $\|e_k\|_1 = 1$ vinden we: $\|A\|_1 = \max_j \sum_i |a_{ij}|$. Hiermee hebben we geïllustreerd wat matrixnormen kunnen zijn.

Conditiegetallen

Nu we weten wat matrixnormen zijn, kunnen we conditiegetallen uitrekenen.

Voorbeeld 1.

$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$, dan is $A^{-1} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$, zodat

$\|A\|_1 = \max_j \sum_i |a_{ij}| = \max(2, 1) = 2$, evenzo $\|A^{-1}\|_1 = 2$ en

dus is $C_2^1(A) = \|A\|_1 \|A^{-1}\|_1 = 4$

Voorbeeld 2.

Met $A = \begin{pmatrix} 1 & 0 \\ 1 & a \end{pmatrix}$ zien we dat $A^{-1} = \begin{pmatrix} 1 & 0 \\ -1 & a^{-1} \end{pmatrix}$. Als nu $a \ll 1$ wordt,

bijv. $a = 10^{-2}$, dan is $\|A^{-1}\|_1 = 10^2$. In het algemeen geldt voor dit geval:

$$C_2^1(A) = \|A\|_1 \|A^{-1}\|_1 = \begin{cases} 2/a & \text{als } a < 1/2 \\ 4 & \text{als } 1/2 \leq a \leq 2 \\ 2a & \text{als } a > 2 \end{cases}$$

We zien dat naarmate a dichter bij nul komt het conditiegetal $C(A)$ groter wordt. Nu is voor dit geval: $\det(A) = a$, dus als $a \approx 0$, dan is $\det(A) \approx 0$, dus voor bijna afhankelijke matrices A is $C(A)$ groot. Het omgekeerde is alleen maar waar als de kolommen van A (bijna) gelijke norm hebben!

Hiermee hebben we geïllustreerd hoe de conditie $C(A)$ van een matrix A berekend wordt en dat indien $\det(A) \approx 0$ is, de conditie zeer slecht wordt (dus $C(A)$ zeer groot).

Voorbeeld 3.

De matrix A is tridiagonaal:

$$A = \begin{pmatrix} d_1 & b_1 & & & \\ c_1 & d_2 & & & \\ & & \ddots & & \\ \phi & & & b_{n-1} & \\ & & c_{n-1} & d_n & \end{pmatrix}$$

met $d_i = -2$, $b_i = c_i = 1$.

In dit voorbeeld is het 1-Conditiegetal voor $n = 1$ ($C_1^1(A)$) dus gelijk aan 1, terwijl het 1-Conditiegetal voor $n = 2$ $C_2^1(A) = 3$.

Voor $n \geq 3$ is $\|A\|_1 = 4$ en verder voor $n = \text{even}$ is $\|A^{-1}\|_1 = (n^2 + 2n)/8$ voor $n = \text{oneven}$ is $\|A^{-1}\|_1 = (n^2 + 2n + 1)/8$, dit is in te zien m.b.v.

$$\begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} -2/3 & -1/3 \\ -1/3 & -2/3 \end{pmatrix} = I$$

$$\begin{pmatrix} -2 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & -2 \end{pmatrix} \begin{pmatrix} -3/4 & -2/4 & -1/4 \\ -2/4 & -4/4 & -2/4 \\ -1/4 & -2/4 & -3/4 \end{pmatrix} = I$$

$$\text{Dus } C_n^1(A) = \begin{cases} (n^2 + 2n)/2 & \text{als } n = \text{even} \\ (n^2 + 2n + 1)/2 & \text{als } n = \text{oneven} \end{cases}$$

Het 2-Conditiegetal is $C_n^2(A) = \|A\|_2 \|A^{-1}\|_2$

Omdat A hier hermiets is, d.w.z. $A = A^*$ komt er met $\rho(A) = (A^* A)^{1/2} = \|A\|_2$:

$$C_n^2(A) = \rho(A) \rho(A^{-1})$$

De spectraalstraal $\rho(A)$ is de absoluut grootste eigenwaarde van A en $\rho(A^{-1})$ is de absoluut grootste eigenwaarde van A^{-1} .

Om de eigenwaarden te bepalen, berekenen we de karakteristieke functie van A:

$$\phi_A(\lambda) = \text{Det}[A - \lambda I] = 0$$

Tweemaal ontwikkelen van deze $n \times n$ -determinant D_n naar een kolom geeft een recurrentie-betrekking voor Chebyshev-polynomen van de tweede soort:

$$\phi_A(\lambda) = D_n(x) = x D_{n-1} - D_{n-2} = 0$$

$$\text{met } D_0 = 1 \text{ en } D_1 = x \text{ en } x = -2 - \lambda$$

$D_n(x)$ is een polynoom met leidende coëfficiënt 1.

$$\text{We vinden } D_n(x) = \sin[(n+1) \arccos(\frac{x}{2})] / \sin[\arccos(\frac{x}{2})] = 0$$

$$\text{met de wortels } \tau_k^n = 2 \cos \frac{k \pi}{n+1}, k = 1(1)n.$$

De eigenwaarden voor een $n \times n$ - matrix worden dus:

$$\lambda_k^n = -2 - \tau_k^n$$

Overzicht:

matrix A	$\phi_A(\lambda)$	λ_k^n	$-2(1 + \cos \frac{k \pi}{n+1})$
$\begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix}$	$\lambda^2 + 4\lambda + 3 = 0$	-3 -1	$-2(1 + \frac{1}{2})$ $k=1$ $-2(1 - \frac{1}{2})$ $k=2$
$\begin{bmatrix} -2 & 1 & 0 \\ -1 & -2 & 1 \\ 0 & 1 & -2 \end{bmatrix}$	$(\lambda+2)(\lambda^2 + 4\lambda + 2) = 0$	$-2-\sqrt{2}$ -2 $-2+\sqrt{2}$	$-2(1 + \frac{1}{2}\sqrt{2})$ $k=1$ $-2(1+0)$ $k=2$ $-2(1 - \frac{1}{2}\sqrt{2})$ $k=3$

$$\text{Nu is } |\lambda_k^n| \text{ maximaal als } k = 1 \rightarrow \rho(A) = 2(1 + \cos \frac{\pi}{n+1})$$

$$\text{Verder is } \rho(A^{-1}) = \frac{1}{\text{absoluut kleinste eigenwaarde van } A} = \frac{1}{\min |\lambda_k^n|},$$

$$|\lambda_k^n| \text{ is minimaal, treedt op als } k = n \rightarrow \rho(A^{-1}) = \frac{1}{2(1 + \cos \frac{n \pi}{n+1})}.$$

Dus het 2-Conditiëgetal is exact:

$$C_n^2(A) = \frac{1 + \cos \frac{\pi}{n+1}}{1 + \cos \frac{n\pi}{n+1}} = \frac{1 + \cos \frac{\pi}{n+1}}{1 - \cos \frac{\pi}{n+1}} = \cotg^2\left(\frac{\pi}{2(n+1)}\right)$$

Voor alle n geldt: $|\cotg(\frac{\pi}{2(n+1)})| \leq 2(n+1)/\pi$,

dus voor alle n : $C_n^2(A) < C_n^1(A)$

Toepassing op faktoranalyse

Laten er p variabelen $X_1, X_2, \dots, X_p$ aan n -proefpersonen (cases) gemeten zijn:

$X_{11}, X_{12}, \dots, X_{1p}$

$X_{21}, X_{22}, \dots, X_{2p}$

$\dots \dots \dots$

$X_{n1}, X_{n2}, \dots, X_{np}$

We bepalen de voorwaarde waaronder deze data "ellendig" zijn m.b.t. faktoranalyse als volgt:

- Eerst tonen we aan dat de produktmoment correlatie van Pearson, indien losgelaten op reële getallen een lineaire afbeelding is.
 - Vervolgens laten we zien dat lineaire afhankelijkheid tussen de vectoren $X_1, X_2, \dots, X_p$ invariant blijft onder een lineaire afbeelding van deze vectoren.
 - Uit a en b volgt dan dat voor "ellendige data", d.w.z. voor bijna lineair afhankelijke $X_1, \dots, X_p$ de correlatie-matrix slecht geconditioneerd is. (Voor uitleg van dit laatste begrip, zie het vorige hoofdstuk).
- ad a. Laten S en T lineaire ruimten over K zijn. We nemen uitsluitend $K = R$, $S = K^n$ en $T = K^p$. Nu heet de afbeelding $A: S \rightarrow T$ een homomorfisme of K -lineaire afbeelding als $A(x + y) = Ax + Ay$, voor alle $x, y \in S$ en $A(\alpha x) = \alpha(Ax)$, voor alle $\alpha \in K$ en $x \in S$. De bewerkingen links zijn in S , rechts in T . Het gewone reële inproduct in K is dus een lineaire afbeelding en dus is Pearson-correlatie er ook een. Merk op dat dit niet zo is, indien $K = \mathbb{C}$.

Er volgt met inductie:

$$A\left(\sum_{j=1}^p \alpha_j x_j\right) = \sum_{j=1}^p \alpha_j (Ax_j)$$

ad b. Laat nu gegeven zijn in totaal p vectoren $x_1, x_2, \dots, x_p$, waarbij iedere vector x_j , $1 \leq j \leq p$ een n -vector is, $n \geq 1$, terwijl er λ_j , $1 \leq j \leq p$ zijn, niet alle gelijk aan nul, zodanig dat,

$$\lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_p x_p = 0 \quad \dots (i)$$

waarbij de nul rechts uiteraard de nulvector is.

Indien de vectoren x_j , $1 \leq j \leq p$ aan de relatie (i) voldoen met de gestelde eis aan de λ 's noemt men ze lineair onafhankelijk.

We bewijzen nu dat iedere lineaire afbeelding A deze lineaire afhankelijkheid invariant laat, als volgt:

Neem $\sum_{j=1}^p \lambda_j x_j = 0$ dus lineair afhankelijk in S . Laat $A: S \rightarrow T$ een lineaire afbeelding zijn, dan geldt volgens a:

$$A \sum_{j=1}^p \lambda_j x_j = \sum_{j=1}^p \lambda_j (Ax_j) = 0$$

Dus de beeldvectoren Ax_j in T zijn weer lineair afhankelijk in T .

ad c. De correlatiematrix over de data x_{ij} , $1 \leq i \leq n$ en $1 \leq j \leq p$, wordt verkregen door allereerst aftrekken van het gemiddelde

$\mu_j = \sum_{i=1}^n x_{ij}/n$, waarin x_{ij} het i^e element van de j -vector is.

Er komt dan de datamatrix $[x_1 - \mu_1, x_2 - \mu_2, \dots, x_p - \mu_p] = [\hat{x}_1, \hat{x}_2, \dots, \hat{x}_p]$.

Aangezien translatie over μ_j een lineaire afbeelding is, blijft lineaire onafhankelijkheid tussen de x_j behouden voor de $\hat{x}_j$, $1 \leq j \leq p$ (volgens b).

De volgende stap na translatie van de data x_{ij} is het toepassen van een reëel inproduct, hetgeen alweer een lineaire afbeelding is (volgens a). Deze zal ook weer lineaire afhankelijkheid tonen

tussen de kolommen $\hat{X}_j$ in de data behouden.

De conclusie is:

Als de kolommen $X_1, X_2, \dots, X_p$ in de ruwe data (nagenoeg) afhankelijk zijn, dan is de correlatiematrix over deze data ill-conditioned.

Er kunnen zich 2 gevallen voordoen:

$n < p$, dus er zijn minder 'proefpersonen' dan 'variabelen', geeft altijd afhankelijkheid.

$n \geq p$ kan afhankelijkheid geven, oorzaak ligt in de opzet van het experiment. Er zijn dus afhankelijke metingen verricht, d.w.z. er zijn metingen die (combinaties van) andere metingen zijn. Er zal dus minstens een variabele buiten beschouwing gelaten moeten worden.

Bartlett, M.S. Tests of significance in factor analysis.
British Journal of Psychology, 1950, 3, 77-85.

Cooley, W.W. en Lohnes, P.R. Multivariate data analysis.
New York: John Wiley & Sons, 1971.

Dziuban, C.D. en Shirkey, E.C. When is a correlation matrix
appropriate for factor analysis? Psychological Bulletin, 1974,
81, 358-361.

Harman, H.H. Modern factor analysis. 2nd edition. Chicago:
University of Chicago Press, 1967.

Kaiser, H.F. A second generation little jiffy. Psychometrika,
1970, 35, 401-416.

Kaiser, H.F. en Rice, J. Little jiffy mark IV. Educational
and Psychological Measurement, 1974.

Kendall, M. Multivariate Analysis. London: Charles Griffin, 1975.

Timm, N.H. Multivariate Analysis: with applications in education
and psychology. Monterey: Brooks/Cole Publishing Comp., 1975.