REGRESIJSKA ANALIZA U STANDARDIZIRANOM ANTIIMAGE PROSTORU

KONSTANTIN MOMIROVIĆ
Sektor za istraživanja, nastavu i razvoj Sveučilišnog računskog centra
i
Zavod za sistematsku kineziologiju Fakulteta za fizičku kulturu
Sveučilišta u Zagrebu

SAŽETAK

Regresijska analiza u standardiziranom antiimage prostoru i Harrisovom, tj. inverznom antiimage prostoru omogućava poseban i koristan uvid u sklop i strukturu latentne dimenzije izvedene iz regresora i dopušta jednostavno testiranje hipoteza o konfiguraciji regresijskih koeficijenata.

KLJUČNE RIJEČI

regresijska analiza / antiimage prostor / Harrisov prostor / testiranje hipoteza

REGRESSION ANALYSIS IN STANDARDIZED ANTIIMAGE SPACE

SUMMARY

Regression analysis in Guttman standardized antiimage space and in Harris inverse antiimage space allows the very useful additional insight in the pattern and structure of latent dimension derived from set of regressors. In both Guttman and Harris space testing of hypotheses concerning the configuration of regression coefficients is usually simplest and more efficient than in usual unstandardized or in standardized I - metrics.

KEYWORDS

regression analysis / antiimage space / Harris space / hypotheses testing

0. UVOD

Neka je

$$Z = (z_{ij}) \qquad i = 1,\dots,n \quad j = 1,\dots,m \quad m < n$$

matrica podataka dobijena na nekom uzorku od n entiteta E_i pomoću m varijabli V_j, centriranih i standardiziranih tako da je $E(z_j) = 0$ i $E(z_j^2) = 1$, $j = 1,\dots,m$ i neka su varijable V_j normalno distribuirane sa funkcijom raspodjele N(0.,1.). Pridajmo varijablama V_j status prediktorskih, eksplanatornih ili, općenito regresorskih varijabli.

Neka je

$$Z_k = (z_{ik}) \qquad i = 1,\dots,n$$

vektor podataka, dobijen na istom uzorku, definiran varijablom V_k, $j \neq k$, koja je takodjer centrirana i standardizirana tako da je $E(z_k) = 0$ i $E(z_k^2) = 1$, i ima funkciju distribucije N(0.,1.). Neka V_k ima logički status kriterijske varijable.

Relacije izmedju svih varijabli V_j, i relacije izmedju varijabli V_j i varijable V_k biti će, u tom slučaju, sigurno linearne, pa matrica

$$R = Z^T Z \frac{1}{n}$$

sadrži produkt-moment koeficijente korelacije izmedju varijabli V_j, a vektor

$$Q = Z^T Z_k \frac{1}{n}$$

produkt-moment koeficijente korelacije izmedju varijabli V_j i varijable V_k.

Neka je

$$\beta = (\beta_j) \qquad j = 1,\dots,m$$

nepoznati vektor sa elementima β_j odredjen tako da vrijedi

$$Z\beta = Z_k + E$$

uz uvjet

$$E^T E = \min.$$

Radi se, naravno, o standardnom regresijskom problemu, pa je

$$\beta = R^{-1} Q$$

a procjena kriterijske varijable V_k

$$\Phi_k = Z\beta .$$

Varijanca procjene kriterijske varijable V_k je

$$\delta^2 = \Phi_k^T \Phi_k \frac{1}{n} = \beta^T R \beta = Q^T R^{-1} Q$$

a korelacija izmedju V_k i njene procjene

$$\rho = Z_k^T \Phi_k \delta^{-1} \frac{1}{n} = Q^T R^{-1} Q \delta^{-1} = \delta .$$

Vektor korelacija izmedju kriterijskih varijabli V_j i standardizirane procjene kriterijske varijable V_k je

$$F = Z^T \Phi_k \delta^{-1} \frac{1}{n} = Q\delta^{-1}$$

i može mu se pridati status jednog od kriterijski orijentiranih faktora matrice R (Seber, 1977; Cooley and Lohnes, 1971).

1. STANDARDIZIRANE ANTIIMAGE VARIJABLE

Neka je

$$U^{-2} = \operatorname{diag} R^{-1} .$$

Matrica

$$A^* = Z R^{-1} U^2$$

sadržavat će podatke o entitetima E_i definirane antiimage varijablama V_j^*. Kovarijance antiimage varijabli su u matrici

$$P^* = A^{*T} A^* \frac{1}{n} = U^2 R^{-1} U^2$$

a njihove su varijance

$$\text{diag } P^* = U^2 .$$

Standardizirane su antiimage varijable u matrici

$$A = A^* U^{-1} = Z R^{-1} U$$

pa su u matrici

$$P = A^T A \frac{1}{n} = U R^{-1} U$$

interkorelacije standardiziranih antiimage varijabli (Guttman, 1953; Harris, 1962; Horst, 1965).

Uočimo, da su kroskorelacije varijabli V_j i standardiziranih antiimage varijabli V_j^*

$$Z^T A \frac{1}{n} = U$$

i da su korelacije izmedju standardiziranih antiimage varijabli V_j^* i kriterijske varijable V_k elementi vektora

$$s = A^T z_k \frac{1}{n} = U R^{-1} q = U \beta .$$

2. REGRESIJSKI PROBLEM U ANTIIMAGE PROSTORU

Neka je

$$\gamma = (\gamma_j)$$

nepoznati vektor sa elementima γ_j odredjen tako da vrijedi

$$A\gamma = Z_k + E$$

uz uvjet

$$E^T E = \min.$$

Očito, u ovako definiranom regresijskom problemu

$$\gamma = P^{-1} S = U^{-1} R U^{-1} U R^{-1} Q = U^{-1} Q$$

pa su parcijalni regresijski koeficijenti $\gamma_j = u_j^{-1} q_{jk}$. Prema tome, parcijalni doprinos neke antiimage varijable predikciji ili eksplanaciji neke kriterijske varijable upravo je proporcionalan veličini korelacije neke prediktorske ili eksplanatorne varijable V_j sa kriterijskom varijablom V_k i obrnuto proporcionalan varijabilitetu te varijable transformirane u antiimage oblik. Ali, kako je za nesingularne prediktorske ili eksplanatorne sisteme $u_j^2 < 1$ za sve varijable V_j^*, parcijalni je doprinos neke antiimage varijable utoliko manji, ukoliko je, pri konstantnoj povezanosti neke realne varijable V_j sa kriterijskom varijablom V_k, manji njen varijabilitet u antiimage prostoru, dakle ukoliko je njen unikni varijabilitet manji.

Naravno, ukupni je prediktorski ili eksplanatorski potencijal sistema antiimage varijabli jednak potencijalu sistema realnih varijabli. Jer, ako je

$$\psi = A\gamma$$

očito je varijanca varijable ψ

$$\psi^T \psi \frac{1}{n} = \gamma^T P \gamma = S^T P^{-1} S = Q^T R^{-1} Q = \delta^2$$

i

$$\psi = A\gamma = Z R^{-1} U U^{-1} r = Z\beta = \Phi.$$

Vektor korelacija standardiziranih antiimage varijabli i standardizirane varijable ψ je

$$G = A^T \psi \delta^{-1} \frac{1}{n} = U Q \delta^{-1} = U F$$

pa kriterijski orijentiran faktor G matrice P predstavlja, u stvari, transformirani kriterijski faktor matrice R, projiciran u zajednički potprostor varijabli V_j i V_j^*.

Uočimo, osim toga, da je

$$\beta = U^{-1} s$$

tj. da su parcijalni regresijski koeficijenti upravo proporcionalni korelacijama standardiziranih antiimage varijabli i kriterijske varijable V_k i obrnuto proporcionalni varijancama antiimage varijabli.

3. NEKE IMPLIKACIJE REPREZENTACIJE REGRESIJSKOG PROBLEMA U ANTIIMAGE PROSTORU

Transformacija nekog skupa varijabli u standardizirane antiimage varijable metrički je invarijantna. Jer, ako je D neka dijagonalna, pozitivno definitna, proizvoljna matrica, operacija

$$B = Z D$$

prevodi varijable V_j u D-metriku, pa je matrica kovarijanci tako remetriziranih varijabli

$$C = B^T B \frac{1}{n} = D R D$$

a D^2 matrica varijanci remetriziranih varijabli.

Neka je

$$V^{-2} = \text{diag } C^{-1} = D^{-2} U^{-2} = U^{-2} D^{-2}$$

pa je antiimage transformacija varijabli iz B

$$B^* = B\,C^{-1}V^2 = Z\,R^{-1}D\,U^2 = Z\,R^{-1}U^2D = A^*D.$$

Standardizirane varijable iz B^* su, medjutim,

$$B^*V^{-1} = Z\,R^{-1}U^2D\,D^{-1}U^{-1} = Z\,R^{-1}U = A$$

i transformacija ma kog skupa varijabli u standardizirane antiimage varijable je zaista metrički invarijantna.

Razmotrimo sada transformaciju varijabli iz Z u Harrisove varijable koje su, naravno, takodjer metrički invarijantne.

Neka je

$$H = Z\,U^{-1}$$

matrica varijabli transformiranih u Harrisov oblik, i neka je

$$W = H^TH\frac{1}{n} = U^{-1}R\,U^{-1}$$

matrica kovarijanci tako transformiranih varijabli. Matrice P i W dijagonaliziraju se u istoj bazi; jer, ako je X matrica svojstvenih vektora od W,

$$X^TW\,X = \Lambda$$

a

$$X^TP\,X = \Lambda^{-1}$$

zbog toga što je, naravno, $W = P^{-1}$.

Kovarijance varijabli V_j transformiranih u Harrisov oblik i kriterijske varijable V_k su

$$H^Tz_k\frac{1}{n} = U^{-1}Q = \gamma$$

tj. jednake su parcijalnim regresijskim koeficijentima izvedenim na temelju standardiziranih antiimage varijabli.

Zbog simetričnosti Harrisovih i standardiziranih antiimage varijabli, regresijski su koeficijenti Harrisovih varijabli

$$W^{-1}\gamma = U R^{-1} Q = U \beta = S$$

tj. jednaki su korelacijama standardiziranih antiimage varijabli sa kriterijskom varijablom V_k. Naravno,

$$\gamma^T W \gamma = \beta^T R \beta = \delta^2$$

jer je

$$H U \beta = \Phi.$$

Kovarijance izmedju Harrisovih varijabli i varijable Φ su

$$H^T \Phi \frac{1}{n} = U^{-1} Q = \gamma$$

dakle jednake kovarijancama Harrisovih varijabli i standardizirane kriterijske varijable V_k, i dok su, naravno, korelacije izmedju standardiziranih varijabli iz H i Φ upravo F, kovarijance Harrisovih varijabli i standardizirane varijable Φ su, zbog toga što je $F = U^{-1} G$, $H^T \Phi \frac{1}{\delta} \frac{1}{n} = U^{-1} F = U^{-1} G$.

4. FUNKCIJE DISTRIBUCIJE REGRESIJSKIH KOEFICIJENATA U ANTIIMAGE I INVERZNOM ANTIIMAGE PROSTORU

Matrica kovarijanci regresijskih koeficijenata izvedenih iz ma kog prediktorskog sistema reprezentiranog matricom podataka Y je

$$K_Y = \sigma_e^2 (Y^T Y)^{-1}$$

gdje je σ_e^2 varijanca pogreške (Anderson, 1958). Otuda je matrica kovarijanci regresijskih koeficijenata $\gamma = (\gamma_j)$ izvedenih iz standardiziranih antiimage varijabli

$$K_A = (1-\delta^2)U^{-1}RU^{-1}g^{-1} = (1-\delta^2)P^{-1}g^{-1}(1-\delta^2)Wg^{-1}$$

gdje je $g = (n-m-1)$ ako je σ_e^2 procijenjena pod kriterijem najmanjih kvadrata, a n ako je σ_e^2 procijenjena pod kriterijem najveće vjerodostojnosti, a matrica kovarijanci regresijskih koeficijenata $s = (s_j)$ izvedenih iz Harrisovih varijabli

$$K_H = (1-\delta^2)UR^{-1}Ug^{-1} = (1-\delta^2)W^{-1}g^{-1} = (1-\delta^2)Pg^{-1}$$

što može biti veoma pogodno za testiranje hipoteza tipa

$$H_{o/\gamma} : c^T\gamma = 0$$

odnosno

$$H_{o/s} : c^T s = 0$$

gdje je c proizvoljni model vektor sa elementima c_j, $c^T c \neq 0$ (Morrison, 1967).

LITERATURA

Anderson, T.W.:
An introduction to multivariate statistical analysis.
Wiley, New York, 1958.

Cooley, W.W. and P.R. Lohnes:
Multivariate data analysis.
Wiley, New York, 1971.

Guttman, L.:
Image theory for the structure of quantitative variates.
Psychometrika, 8(1953), 277-296.

Harris, C.W.:
Some Rao-Guttman relationships.
Psychometrika, 27(1962), 247-263.

Morrison, D.F.:
Multivariate statistical methods.
McGraw-Hill, New York, 1967.

Seber, G.A.F.:
Linear regression analysis.
Wiley, New York, 1977.