Sveučilište Josipa Jurja Strossmayera u Osijeku - Odjel za matematiku
Diplomski studij Financijska matematika i statistika

Ela Kovačić

Kanoničke korelacije

Diplomski rad

Osijek,25. siječnja 2020.

Sveučilište Josipa Jurja Strossmayera u Osijeku - Odjel za matematiku

Diplomski studij Financijska matematika i statistika

Ela Kovačić

Kanoničke korelacije

Diplomski rad

Mentorica: prof. dr. sc. Mirta Benšić

Osijek, 25. siječnja 2020.

Sadržaj

1 Uvod

Kanonička korelacija je dodatna procedura za procjenu veze između varijabli. Prvi ju je predložio Harold Hotelling 1936. godine u časopisu Biometrika. Omogućava proučavanje odnosa između dva skupa varijabli, odnosno varijata. Traže se linearne kombinacije varijabli tako da je njihova korelacija maksimalna. Takvu korelaciju nazivamo kanonička korelacija između para kanoničkih varijata. Bitno je da par kanoničkih varijata koji odgovara kanoničkoj korelaciji nije u korelaciji s ostalim kanoničkim varijatama.

Ova procedura se koristi u istraživačkim studijima te se koristi u medicini, farmaciji, ekonometriji, psihologiji i drugim područjima. Na primjer, jedan skup varijabli može biti mjerenja fizičkih karakteristika kao što su duljina i širina lubanje, a druge varijable mogu biti mjerenja mentalnih karakteristika kao što su rezultati testa inteligencije. Između ova dva skupa traži se kanonička korelacija. Broj dobivenih kanoničkih korelacija, s time i parova kanoničkih varijata, jednak je manjem broju varijabli jednog od dva skupa varijata. U većini slučajeva veza između dva skupa varijata opisana je u potpunosti korelacijom prvih kanoničkih varijata s obzirom na to da je njihova korelacija najveća.

2 Kanoničke korelacije i varijate u populaciji

Neka je X slučajni vektor s p komponenata i matricom kovarijance Σ, za koju pretpostavimo da je pozitivno definitna. Pretpostavimo da je $EX = 0$. U razvoju koncepta i algebre nije potrebno pretpostaviti da je X normalno distribuiran. Rastavimo vektor X na dva vektora s p_1 i p_2 komponenti, respektivno,

$$X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}, \tag{1}$$

pri čemu je $p_1 \leq p_2$.
Slično, matricu kovarijance Σ rastavimo na p_1 i p_2 redaka i stupaca,

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}. \tag{2}$$

Neka je $U = \alpha^T X^{(1)}$ proizvoljna linearna kombinacija komponenata vektora $X^{(1)}$ i neka je $V = \gamma^T X^{(2)}$ proizvoljna linearna kombinacija komponenata vektora $X^{(2)}$. Tražimo linearnu funkciju od U i V koja će imati maksimalnu korelaciju između U i V. S obzirom da je korelacija višekratnika od U i višekratnika od V jednaka kao i korelacija od U i V, α i γ možemo proivoljno normalizirati. Tražimo α i γ takve da U i V imaju jediničnu varijancu, odnosno da vrijedi

$$VarU = EU^2 = E\alpha^T X^{(1)} X^{(1)T} \alpha = \alpha^T \Sigma_{11} \alpha = 1, \tag{3}$$

$$VarV = EV^2 = E\gamma^T X^{(2)} X^{(2)T} \gamma = \gamma^T \Sigma_{22} \gamma = 1. \tag{4}$$

Primijetimo da je $EU = E\alpha^T X^{(1)} = \alpha^T EX^{(1)} = 0$ i analogno tome je $EV = 0$. Korelacija između U i V jednaka je

$$\begin{aligned} Corr(U,V) &= \frac{Cov(U,V)}{\sqrt{VarU}\sqrt{VarV}} = Cov(U,V) \\ &= E\alpha^T X^{(1)} X^{(2)T} \gamma = \alpha^T \Sigma_{12} \gamma. \end{aligned} \tag{5}$$

Potrebno je pronaći α i γ takve da je korelaciju između U i V maksimalna, uz uvjete (3) i (4). Problem pronalaska uvjetnih ekstrema α i γ možemo svesti na problem pronalska maksimuma Lagrangeove funkcije $\psi(\alpha, \gamma)$.

Neka je

$$\psi(\alpha,\gamma) = \alpha^T\Sigma_{12}\gamma - \frac{1}{2}\lambda(\alpha^T\Sigma_{11}\alpha - 1) - \frac{1}{2}\mu(\gamma^T\Sigma_{22}\gamma - 1) \tag{6}$$

Lagrangeova funkcija, gdje su λ i μ Lagrangeovi multiplikatori. Diferenciramo funkciju ψ po α i γ te vektore derivacija izjednačimo s nulom,

$$\frac{\partial\psi(\alpha,\gamma)}{\partial\alpha} = \Sigma_{12}\gamma - \lambda\Sigma_{11}\alpha = 0, \tag{7}$$

$$\frac{\partial\psi(\alpha,\gamma)}{\partial\gamma} = \Sigma_{12}^T\alpha - \mu\Sigma_{22}\gamma = 0. \tag{8}$$

Množenjem (7) s lijeva s α^T i množenjem (8) s lijeva s γ^T dobivamo

$$\begin{aligned} \alpha^T\Sigma_{12}\gamma - \lambda\alpha^T\Sigma_{11}\alpha &= 0, \\ \gamma^T\Sigma_{12}^T\alpha - \mu\gamma^T\Sigma_{22}\gamma &= 0. \end{aligned}$$

Kako je $\alpha^T\Sigma_{11}\alpha = 1$ i $\gamma^T\Sigma_{22}\gamma = 1$, tada je

$$\alpha^T\Sigma_{12}\gamma - \lambda = 0, \tag{9}$$
$$\gamma^T\Sigma_{12}^T\alpha - \mu = 0, \tag{10}$$

odnosno $\lambda = \alpha^T\Sigma_{12}\gamma = \mu$. Uvrštavanjem λ umjesto μ u (8) te korištenjem jednakosti $\Sigma_{12}^T = \Sigma_{21}$, dobije se sljedeći sustav jednadžbi

$$\begin{aligned} -\lambda\Sigma_{11}\alpha + \Sigma_{12}\gamma &= 0, \\ \Sigma_{21}\alpha - \lambda\Sigma_{22}\gamma &= 0, \end{aligned}$$

Dobiveni sustav jednadžbi možemo prikazati u matričnom obliku

$$\begin{pmatrix} -\lambda\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\lambda\Sigma_{22} \end{pmatrix} \begin{pmatrix} \alpha \\ \gamma \end{pmatrix} = 0. \tag{11}$$

Kako bi postojalo netrivijalno rješenje ovog sustava (što je neophodno za rješenje koje zadovoljava (3) i (4)), lijeva matrica u (11) mora biti singularna, što bi značilo da determinanta matrice sustava mora biti jednaka nuli, odnosno da je

$$\begin{vmatrix} -\lambda\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\lambda\Sigma_{22} \end{vmatrix} = 0. \tag{12}$$

Determinanta u (12) je reda $(p_1+p_2)\times(p_1+p_2)$. U Laplaceovom raspisu determinante pomoću minora u jednom raspisu imamo $|-\lambda\Sigma_{11}|\cdot|-\lambda\Sigma_{22}| = (-\lambda)^{p_1+p_2}|\Sigma_{11}||\Sigma_{22}|$. U ostalim raspisima λ je manjeg stupnja jer jedan ili više redova svake minore u prvim p_1 stupcima ne sadrži λ. Kako je Σ pozitivno definitna matrica, to je $|\Sigma_{11}||\Sigma_{22}| \neq 0$. Ovo pokazuje da je determinanta matrice sustava polinomna jednadžba stupnja p_1+p_2 u varijabli λ. Pod pretpostavkom da je $p_1 < p_2$, od p_1+p_2 korijena jednadžbe, njih $p_2 - p_1$ iščezava. Za koeficijente $\lambda^{p_2-p_1-1}$ i manjeg stupnja, glavne minore koje sadrže $2p_1 + 1$ ili više redova u kojima je λ zamijenjena s nulom iščezavaju, što se može vidjeti iz Laplaceovog raspisa. Determinanta matrice sustava (12) ekvivalentna je determinanti

$$(-\lambda)^{p_2-p_1}\begin{vmatrix} \lambda^2\Sigma_{11} & -\lambda\Sigma_{12} \\ (-\lambda)^{-1}\Sigma_{21} & \Sigma_{22} \end{vmatrix} = 0,$$

gdje prvih p_1 redova množimo s $-\lambda$, a zadnjih p_2 redova množimo s $(-\lambda)^{-1}$. Množenjem s lijeva drugom determinantom reda $(p_1 + p_2) \times (p_1 + p_2)$, to će i dalje biti jednako nuli. Množimo s determinantom

$$\begin{vmatrix} I_{p_1} & \lambda\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & \Sigma_{22}^{-1} \end{vmatrix}.$$

Kako je determinanta produkta jednaka produktu determinanti, dobivamo

$$(-\lambda)^{p_2-p_1}\begin{vmatrix} \lambda^2\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & 0 \\ (-\lambda)^{-1}\Sigma_{22}^{-1}\Sigma_{21} & I_{p_2} \end{vmatrix} = 0,$$

odnosno

$$(-\lambda)^{p_2-p_1}\left|\lambda^2\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right| = 0.$$

Ovo je također polinom stupnja $p_1 + p_2$ u varijabli λ. $p_2 - p_1$ korijena je jednako nuli. Ostali korijeni nastaju računanjem determinante reda $p_1 \times p_1$. Korijeni koji nužno ne iščezavaju, $p_1 + p_2 - (p_2 - p_1) = 2p_1$, su korijeni koji nastaju u parovima i koje možemo pisati u obliku $\pm\rho_1, \pm\rho_2, ..., \pm\rho_{p_1}$, pri čemu je $\rho_1 \geq \rho_2 \geq \cdots \rho_{p_1}$.
Iz (9) vidimo da je $\lambda = \alpha^T\Sigma_{12}\gamma$ korelacija između $U = \alpha^T X^{(1)}$ i $V = \gamma^T X^{(2)}$ kad α i γ zadovoljavaju (11) za neke vrijednosti λ. Uvrštavanjem jednostrukih korijena umjesto λ i μ u jednadzbe (7) i (8), dobivaju se vrijednosti za $\alpha^{(1)}, ..., \alpha^{(p_1)}$ i $\gamma^{(1)}, ..., \gamma^{(p_1)}$. Tako su varijate U i V potpuno određene.

Promotrimo sljedeći sustav jednadžbi

$$\begin{aligned}
\Sigma_{12}\gamma^{(r)} &= -\rho_r\Sigma_{11}(-\alpha^{(r)}),\\
\Sigma_{21}(-\alpha^{(r)}) &= -\rho_r\Sigma_{22}(-\gamma^{(r)}),
\end{aligned}$$

za $1 \leq r \leq p_1$. Iz ovog sustava jednadžbi možemo vidjeti ako je ρ_r, $\alpha^{(r)}$, $\gamma^{(r)}$ rješenje, tada je rješenje i $-\rho_r$, $-\alpha^{(r)}$, $-\gamma^{(r)}$. Ako je ρ_r negativan, tada je $-\rho_r$ pozitivan broj i vrijedi $-\rho_r \geq \rho_r$. Ali kako je ρ_r maksimalan, tada je $\rho_r \geq -\rho_r$ te je $\rho_r \geq 0$. Tako dobivamo p_1 parova kanoničkih varijata

$$\begin{aligned}
U_1 &= \alpha^{(1)T}X^{(1)}, V_1 = \gamma^{(1)T}X^{(2)}, \quad Cov(U_1, V_1) = \rho_1\\
U_2 &= \alpha^{(2)T}X^{(1)}, V_2 = \gamma^{(2)T}X^{(2)}, \quad Cov(U_1, V_1) = \rho_2\\
U_3 &= \alpha^{(3)T}X^{(1)}, V_3 = \gamma^{(3)T}X^{(2)}, \quad Cov(U_1, V_1) = \rho_3\\
&\vdots\\
U_{p_1} &= \alpha^{(p_1)T}X^{(1)}, V_1 = \gamma^{(p_1)T}X^{(2)}, \quad Cov(U_1, V_1) = \rho_{p_1}.
\end{aligned}$$

Ovih p_1 nenegativnih korijena zovemo **kanoničke korelacije** između skupa varijata, a odgovarajuće linearne funkcije U i V čiji koeficijenti zadovoljavaju uvjete zovemo **kanoničke varijate**.
Treba još provjeriti da sve kanoničke varijate koje nisu u paru nisu korelirane. Neka kanonička korelacija ρ_r, za $1 \leq r \leq p_1$ odgovara paru varijata

$$U_r = \alpha^{(r)T}X^{(1)}, \quad V_r = \gamma^{(r)T}X^{(2)},$$

čiji koeficijenti zadovoljavaju (7) i (8). Za njih vrijedi

$$\Sigma_{12}\gamma^{(r)} = \rho_r\Sigma_{11}\alpha^{(r)}, \tag{13}$$
$$\Sigma_{12}^T\alpha^{(r)} = \rho_r\Sigma_{22}\gamma^{(r)}. \tag{14}$$

Također, neka su

$$U_s = \alpha^{(s)T}X^{(1)}, \quad V_s = \gamma^{(s)T}X^{(2)},$$

kanoničke varijate čija je kanonička korelacija ρ_s, gdje je $r \leq s \leq p_1$. Između ovih četiri varijata U_r, V_r, U_s, V_s imamo šest korelacija. Osim ρ_r i ρ_s, još imamo

$$\begin{aligned} EU_rU_s &= \alpha^{(r)T}\Sigma_{11}\alpha^{(s)}, \\ EU_rV_s &= \alpha^{(r)T}\Sigma_{12}\gamma^{(s)}, \\ EU_sV_r &= \alpha^{(s)T}\Sigma_{12}\gamma^{(r)}, \\ EV_rV_s &= \gamma^{(r)T}\Sigma_{22}\gamma^{(s)}. \end{aligned}$$

Želimo pokazati da su ove četiri korelacije jednake nuli. Množenjem (13) s lijeva sa $\alpha^{(s)T}$, dobivamo

$$\alpha^{(s)T}\Sigma_{12}\gamma^{(r)} = \rho_r\alpha^{(s)T}\Sigma_{11}\alpha^{(r)} = \rho_r\alpha^{(r)T}\Sigma_{11}\alpha^{(s)},$$

odnosno

$$EU_sV_r = \rho_r EU_rU_s.$$

Množenjem (14) s lijeva sa $\gamma^{(s)T}$ dobivamo

$$\gamma^{(s)T}\Sigma_{12}^T\alpha^{(r)} = \rho_r\gamma^{(s)T}\Sigma_{22}\gamma^{(r)} = \rho_r\gamma^{(r)T}\Sigma_{22}\gamma^{(s)},$$

odnosno

$$EU_rV_s = \rho_r EV_rV_s.$$

Međusobnom zamjenom r i s dobivamo $EU_sV_r = \rho_s EV_sV_r = \rho_s EV_rV_s$. Iz toga slijedi da je

$$\begin{aligned} \rho_r EU_rU_s = EU_sV_r &= \rho_s EV_rV_s, \\ \rho_s EU_rU_s &= \rho_r EV_rV_s. \end{aligned}$$

Ako vrijedi $\rho_r^2 \neq \rho_s^2$, iz zadnje dvije jednakosti sljedi da je $EU_rU_s = EV_rV_s = 0$. Prema tome, korelacije EU_sV_r i EU_rV_s iščezavaju. Tako su sve korelacije između kanoničkih varijata jednake nula osim odgovarajućih kanoničkih korelacija.

Definicija 1. *Neka je* $X = (X^{(1)T}\ X^{(2)T})^T$ *slučajni vektor, gdje* $X^{(1)}$ *ima* p_1 *komponenti, a* $X^{(2)}$ *ima* p_2 *komponenti.* r*-ti par kanoničkih varijata je par linearnih kombinacija* $U_r = \alpha^{(r)T} X^{(1)}$ *i* $V_r = \gamma^{(r)T} X^{(2)}$*, s jediničnom varijancom koji je nekoreliran s prvih* $r-1$ *parova kanoničkih varijata, te ima iznos korelacije* ρ_r *koji je* r*-ta po veličini među svim linearnim kombinacijama* $U = \alpha^T X^{(1)}$ *i* $V = \gamma^T X^{(2)}$*. Tu korelaciju zovemo* r*-ta kanonička korelacija.*

Definicija 2. *Neka je* $A = (\alpha^{(1)} \cdots \alpha^{(p_1)})$ *i* $\Gamma_1 = (\gamma^{(1)} \cdots \gamma^{(p_1)})$*. Elemente matrica* A *i* Γ_1 *zovemo kanonički koeficijenti. Oni osiguravaju maksimalnu korelaciju između parova kanoničkih varijata koje pripadaju različitim skupovima varijabli.*

Uvjeti za kanoničke koeficijente i kanoničke korelacije mogu se sažeti u sljedeće jednadžbe

$$A^T \Sigma_{11} A = I,$$
$$A^T \Sigma_{12} \Gamma_1 = \Lambda,$$
$$\Gamma_1^T \Sigma_{22} \Gamma_1 = I.$$

Redukcija skupa varijata u kanoničku formu može se izvršiti odabirom novih varijata $V_{p_1+1}, ..., V_{p_2}$ kao linearnih funkcija drugog i mnogobrojnijeg skupa (osim ako je $p_1 = p_2$), koje su međusobno nekorelirane. Definiramo sljedeću matricu $\Gamma_2 = (\gamma^{(p_1+1)} \cdots \gamma^{(p_2)})$ koja je dimenzije $p_2 \times (p_2 - p_1)$ i koja zadovoljava sljedeće uvjete

$$\Gamma_2^T \Sigma_{22} \Gamma_1 = 0,$$
$$\Gamma_2^T \Sigma_{22} \Gamma_2 = I.$$

Ova matrica se formira stupac po stupac: $\gamma^{(p_1+1)}$ je vektor ortogonalan na $\Sigma_{22}\Gamma_1$ i normaliziran pa je $\gamma^{(p_1+1)T} \Sigma_{22} \gamma^{(p_1+1)} = 1$; $\gamma^{(p_1+2)}$ je vektor ortogonalan na $\Sigma_{22}\Gamma_1 \gamma^{(p_1+1)}$ i normaliziran, te je $\gamma^{(p_1+2)T} \Sigma_{22} \gamma^{(p_1+2)} = 1$; i tako dalje. Tada su varijate oblika $V_k = \gamma^{(k)^T} X^{(2)}$, za $k = p_1 + 1, ..., p_2$. Korelacija između varijata V_k i U_r je

$$EU_r V_k = \alpha^{(r)T} \Sigma_{12} \gamma^{(k)} = \rho_r \gamma^{(r)^T} \Sigma_{22} \gamma^{(k)} = \rho_r E V_r V_k,$$

što je jednako nuli jer su varijate V_k i V_r nekorelirane.

Definicija 3. *Neka je*

$$U = \begin{pmatrix} U_1 \\ \vdots \\ U_{p_1} \end{pmatrix} = A^T X^{(1)}, \tag{15}$$

$$V^{(1)} = \begin{pmatrix} V_1 \\ \vdots \\ V_{p_1} \end{pmatrix} = \Gamma_1^T X^{(2)}, \tag{16}$$

$$V^{(2)} = \begin{pmatrix} V_{p_1+1} \\ \vdots \\ V_{p_2} \end{pmatrix} = \Gamma_2^T X^{(2)}. \tag{17}$$

Komponente od U su jedan skup kanoničkih varijata, a komponente od $V = (V^{(1)}$ $V^{(2)})$ su drugi skup.

Korelacija između kanoničkih varijata U i V je

$$\begin{aligned} & E \begin{pmatrix} U \\ V^{(1)} \\ V^{(2)} \end{pmatrix} \begin{pmatrix} U^T & V^{(1)T} & V^{(2)T} \end{pmatrix} \\ &= \begin{pmatrix} A^T & 0 \\ 0 & \Gamma_1^T \\ 0 & \Gamma_2^T \end{pmatrix} \cdot \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \cdot \begin{pmatrix} A & 0 & 0 \\ 0 & \Gamma_1 & \Gamma_2 \end{pmatrix} \\ &= \begin{pmatrix} I_{p_1} & \Lambda & 0 \\ \Lambda & I & 0 \\ 0 & 0 & I_{p_2-p_1} \end{pmatrix}, \end{aligned} \tag{18}$$

gdje je

$$\Lambda = diag(\rho_1, ..., \rho_{p_1}). \tag{19}$$

Determinanta korelacija između kanoničkih varijata je

$$\Delta = \begin{vmatrix} 1 & 0 & \cdots & 0 & \rho_1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \rho_2 & \cdots & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \\ 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & \rho_{p_1} & \cdots & 0 \\ \rho_1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \rho_2 & \cdots & 0 & 0 & 1 & \cdots & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \\ 0 & 0 & \cdots & \rho_{p_1} & 0 & 0 & \cdots & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 1 \end{vmatrix}$$
$$= (1-\rho_1^2)(1-\rho_2^2)\cdots(1-\rho_{p_1}^2)$$

2.1 Računanje kanoničkih korelacija

U ovom dijelu ćemo izvesti matrične jednadžbe koje se koriste za računanje koeficijenata α i γ te kanoničkih korelacija. Koristimo sljedeći sustav jednadžbi

$$\begin{aligned} -\lambda\Sigma_{11}\alpha + \Sigma_{12}\gamma &= 0, \\ \Sigma_{21}\alpha - \lambda\Sigma_{22}\gamma &= 0. \end{aligned}$$

Pomnožimo prvu jednadžbu sustava sa λ te drugu jednadžbu sa Σ_{22}^{-1}, dobivamo sljedeće

$$\begin{aligned} \lambda\Sigma_{12}\gamma &= \lambda^2\Sigma_{11}\alpha, \\ \Sigma_{22}^{-1}\Sigma_{21}\alpha &= \lambda\gamma. \end{aligned}$$

Uvrstimo li drugu jednadžbu u prvu jednadžbu sustava, dobivamo

$$\begin{aligned} \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\alpha &= \lambda^2\Sigma_{11}\alpha, \\ (\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \lambda^2\Sigma_{11})\alpha &= 0. \end{aligned} \tag{20}$$

Pomnožimo li (20) s lijeva sa Σ_{11}^{-1}, Dobivamo jednadžbu

$$(\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \lambda^2)\alpha = 0, \tag{21}$$

Veličine $\lambda_1^2, ..., \lambda_{p_1}^2$ su svojstvene vrijednosti koje zadovoljavaju jednadžbu

$$|\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \lambda^2| = 0, \tag{22}$$

a $\alpha^{(1)}, ..., \alpha^{(p_1)}$ svojstveni vektori koji zadovoljavaju (21) za $\lambda^2 = \lambda_1^2, ..., \lambda_{p_1}^2$, respektivno. Analogno dolazimo do koeficijenata $\gamma^{(1)}, ..., \gamma^{(p_2)}$. Pomnožimo prvu jednadžbu početnog sustava sa Σ_{11}^{-1} i drugu jednadžbu sa λ, dobivamo sljedeće

$$\begin{aligned} \Sigma_{11}^{-1}\Sigma_{12}\gamma &= \lambda\alpha, \\ \Sigma_{21}\lambda\alpha &= \lambda^2\Sigma_{22}\gamma. \end{aligned}$$

Uvrstimo li prvu jednadžbu sustava u drugu, dobivamo

$$\begin{aligned} \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\gamma &= \lambda^2\Sigma_{22}\gamma, \\ (\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} - \lambda^2\Sigma_{22})\gamma &= 0. \end{aligned} \tag{23}$$

Pomnožimo li s lijeva sa Σ_{22}^{-1}, dobivamo jednadžbu

$$(\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} - \lambda^2)\gamma = 0 \tag{24}$$

Veličine $\lambda_1^2, ..., \lambda_{p_1}^2$ su svojstvene vrijednosti koje zadovoljavaju

$$|\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} - \lambda^2| = 0, \tag{25}$$

a $\gamma^{(1)}, ..., \gamma^{(p_1)}$ su svojstveni vektori koji zadovoljavaju (24) za $\lambda^2 = \lambda_1^2, ..., \lambda_{p_1}^2$, respektivno.

Neka su $1 \geq \lambda_1^2 \geq \cdots \geq \lambda_{p_1}^2 \geq 0$ svojstvene vrijednosti matrica $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ i $\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$. Pozitivni korijeni svojstvenih vrijednosti, $\rho_i = \sqrt{\lambda_i^2}$, $i = 1, ..., p_1$ su **kanoničke korelacije**.

Mnogi algoritmi za računanje svojstvenih vrijednosti i svojstvenih vektora prihvaćaju samo simetrične matrice. Kako matrica $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ nije simetrična matrica, nego je produkt simetričnih matrica, matrice Σ_{11}^{-1} i matrice $\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, svojstvene vrijednosti i svojstvene vektore možemo dobiti pomoću Cholesky dekompozicije. Za matricu Σ_{11} napravimo Cholesky dekompoziciju ili dekompoziciju svojstvenim vrijednostima, tako da dobijemo matricu $\Sigma_{11}^{1/2}$, za koju je $\Sigma_{11}^{T/2}\Sigma_{11}^{1/2} = \Sigma_{11}$. Za tako dobivenu matricu vrijedi $\Sigma_{11}^{-1/2}\Sigma_{11}^{1/2} = I$. Tada se jednadžba (21) svodi na

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

Pomnožimo li s lijeva sa matricom $\Sigma_{11}^{-T/2}$, te zdesna matricom $I = \Sigma_{11}^{-1/2}\Sigma_{11}^{1/2}$, dobivamo

$$(\Sigma_{11}^{-T/2}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \lambda^2\Sigma_{11}^{1/2})\Sigma_{11}^{-1/2}\Sigma_{11}^{1/2}\alpha = 0,$$
$$(\Sigma_{11}^{-T/2}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1/2} - \lambda^2 I)\Sigma_{11}^{1/2}\alpha = 0.$$

Tako $\Sigma_{11}^{-T/2}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1/2}$ ima iste svojstvene vrijednosti $\lambda_1, ..., \lambda_{p_1}$ kao matrica $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, pri čemu je $1 \geq \lambda_1^2 \geq \lambda_2^2 \geq \cdots \lambda_{p_1}^2 \geq 0$. Svojstveni vektori su oblika $e = \Sigma_{11}^{1/2}\alpha$, gdje je α svojstveni vektor od $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. Pozitivni korijeni svojstvenih vrijednosti, $\rho_i = \sqrt{\lambda_i^2}$, $i = 1, ..., p_1$ su kanoničke korelacije.
Na analogan način nađemo Cholesky dekompoziciju matrice Σ_{22}, $\Sigma_{22}^{T/2}\Sigma_{22}^{1/2} = \Sigma_{22}$, tako da je $\Sigma_{22}^{-1/2}\Sigma_{22}^{1/2} = I$. Jednadžba (24) svodi se na

$$(\Sigma_{22}^{-T/2}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1/2} - \lambda^2 I)\Sigma_{22}^{1/2}\gamma = 0.$$

Dobivene svojstvene vrijednosti jednake su svojstvenim vrijednostima $\lambda_1, ..., \lambda_{p_1}$, a svojstveni vektori su oblika $f = \Sigma_{22}^{1/2}\gamma$.
Prema tome, k-ti par kanoničkih varijata, za $k = 1, ..., p_1$, je dan s

$$U_k = e_k^T\Sigma_{11}^{-1/2}X^{(1)} \quad V_k = f_k^T\Sigma_{22}^{-1/2}X^{(2)}. \tag{26}$$

Vrijedi da je $Var(U_k) = Var(V_k) = 1$ i

$$Corr(U_k, V_k) = \rho_k. \tag{27}$$

2.2 Korelacija između originalnog vektora i kanoničkih varijata. Indeks redundancije

Kovarijanca između originalnih varijabli i odgovarajućih kanoničkih varijata je

$$Cov(U, X^{(1)}) = Cov(\alpha^T X^{(1)}, X^{(1)}) = \alpha^T\Sigma_{11},$$
$$Cov(U, X^{(2)}) = Cov(\alpha^T X^{(1)}, X^{(2)}) = \alpha^T\Sigma_{12},$$
$$Cov(V, X^{(1)}) = Cov(\gamma^T X^{(2)}, X^{(1)}) = \gamma^T\Sigma_{21},$$
$$Cov(V, X^{(2)}) = Cov(\gamma^T X^{(2)}, X^{(2)}) = \gamma^T\Sigma_{22}.$$

Korelacija između originalnih varijabli i odgovarajuće kanoničke varijate naziva se **kanoničko opterećenje** [1]. Kanonička opterećenja daju uvid u to koliki utjecaj ima

[1] engl. canonical loadings

pojedina varijabla na kanoničku varijatu. Korelacija između originalnih varijabli i suprotne kanoničke varijate naziva se **kanonička kros-opterećenja** [2]. Tako je

$$\begin{aligned}
Corr(U, X^{(1)}) &= Cov(\alpha^T X^{(1)}, \tilde{\Sigma_{11}}^{-1/2} X^{(1)}) = \alpha^T \Sigma_{11} \tilde{\Sigma_{11}}^{-1/2}, \\
Corr(U, X^{(2)}) &= Cov(\alpha^T X^{(1)}, \tilde{\Sigma_{22}}^{-1/2} X^{(2)}) = \alpha^T \Sigma_{12} \tilde{\Sigma_{22}}^{-1/2}, \\
Corr(V, X^{(1)}) &= Cov(\gamma^T X^{(2)}, \tilde{\Sigma_{11}}^{-1/2} X^{(1)}) = \gamma^T \Sigma_{21} \tilde{\Sigma_{11}}^{-1/2}, \\
Corr(V, X^{(2)}) &= Cov(\gamma^T X^{(2)}, \tilde{\Sigma_{22}}^{-1/2} X^{(2)}) = \gamma^T \Sigma_{22} \tilde{\Sigma_{22}}^{-1/2},
\end{aligned}$$

gdje je $Var(U_k) = Var(V_l) = 1$, za svaki $k, l = 1, ..., p_1$. $\tilde{\Sigma_{11}} = diag(\Sigma_{11})$ je dijagonalna matrica koja sadrži varijance od $X^{(1)}$, a $\tilde{\Sigma_{22}}$ dijagonalna matrica koja sadrži varijance od $X^{(2)}$.

Kanonička opterećenja predstavljaju korelacije između kanoničke varijate i originalne varijable odgovorajućeg skupa. Kvadriranjem kanoničkih opterećenja dobivamo vrijednosti koje predstavljaju količinu varijabiliteta za svaku varijablu. Može se gledati i njihov prosjek kako bi dobili naznaku koliko je varijabilnosti u prosjeku objašnjeno odgovarajućom kanoničkom varijatom u tom skupu varijabli.

Indeks redundancije je količina varijance jedne kanoničke varijate koja se može objasniti drugom kanoničkom varijatom. Predstavlja ukupan varijabilitet jednog skupa varijabli koji se može objasniti na temelju drugog skupa varijabli. Za njegovo računanje potrebno je izračunati kanonička opterećenja i kanoničku korelaciju. Indeks redundancije se računa kao umnožak prosjeka kvadriranih kanoničkih opterećenja i kvadriranih kanoničkih korelacija,

$$\begin{aligned}
RI_U &= \frac{\sum_{k=1}^{p_2} Corr(V, X_k^{(2)})^2}{p_2} \cdot \rho^2 \\
RI_V &= \frac{\sum_{k=1}^{p_1} Corr(U, X_k^{(1)})^2}{p_1} \cdot \rho^2.
\end{aligned}$$

RI_U označava indeks redundancije kanoničke varijate U objašnjene kanoničkom varijatom V. $Corr(V, X_k^{(2)})$ je kanoničko opterećenje kanoničke varijate V i k-te originalne varijable iz podskupa $X^{(2)}$, a ρ je odgovarajuća kanonička korelacija. Analogno tome, RI_V označava indeks redundancije kanoničke varijate V objašnjene kanoničkom varijatom U. $Corr(U, X_k^{(1)})$ je kanoničko opterećenje za kanoničku varijatu U i k-te varijable iz podskupa $X^{(1)}$. Indeksi redundancija nisu simetrični.

[2]engl. canonical cross loadings

3 Procjena kanoničkih korelacija i varijata

3.1 Procjenitelj maksimalne vjerodostojnosti kanoničkih korelacija. Nepristran procjenitelj kanoničkih korelacija

Neka je vektor x vektor opservacija slučajnog vektora X s p komponenata. Neka su $x_1, ..., x_N$ neovisne opservacije iz $N(\mu, \Sigma)$. Podijelimo x_α na p_1 i p_2 komponenata ($p_1 \leq p_2$), respektivno,

$$x_\alpha = \begin{pmatrix} x_\alpha^{(1)} \\ x_\alpha^{(2)} \end{pmatrix}, \quad \alpha = 1, ..., N. \tag{28}$$

Vrijedi sljedeći teorem.

Teorem 1. *Ako $x_1, ..., x_N$ tvore jednostavni slučajni uzorak iz $N(\mu, \Sigma)$, pri čemu je $p < N$, tada su procjenitelji maksimalne vjerodostojnosti* [3] *od μ i Σ, redom,*

$$\hat{\mu} = \overline{x} = \frac{1}{N} \sum_{\alpha=1}^{N} x_\alpha,$$

$$\hat{\Sigma} = \frac{1}{N} \sum_{\alpha=1}^{N} (x_{\alpha - \hat{x}})(x_{\alpha - \hat{x}})^T.$$

Prema ovom teoremu, procjenitelj maksimalne vjerodostojnosti od matrice kovarijanci $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ je

$$\begin{aligned} \hat{\Sigma} &= \begin{pmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{pmatrix} \\ &= \frac{1}{N} \sum_{\alpha=1}^{N} (x_{\alpha - \hat{x}})(x_{\alpha - \hat{x}})^T \\ &= \frac{1}{N} \begin{pmatrix} \sum (x_\alpha^{(1)} - \hat{x}^{(1)})(x_\alpha^{(1)} - \hat{x}^{(1)})^T & \sum (x_\alpha^{(1)} - \hat{x}^{(1)})(x_\alpha^{(2)} - \hat{x}^{(2)})^T \\ \sum (x_\alpha^{(2)} - \hat{x}^{(2)})(x_\alpha^{(1)} - \hat{x}^{(1)})^T & \sum (x_\alpha^{(2)} - \hat{x}^{(2)})(x_\alpha^{(2)} - \hat{x}^{(2)})^T \end{pmatrix}. \end{aligned} \tag{29}$$

Procjenitelji maksimalne vjerodostojnosti kanoničkih korelacija Λ i kanoničkih varijabli definiranih s A i Γ uključuju korištenje algebre iz prethodnog poglavlja na $\hat{\Sigma}$. Prema tome, procjenitelji maksimalne vjerodostojnosti od $\rho_1, ..., \rho_p$ su korijeni od

$$\begin{vmatrix} -\hat{\rho}\hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & -\hat{\rho}\hat{\Sigma}_{22} \end{vmatrix} = 0, \tag{30}$$

[3] engl. maximum likelihood estimator (MLE)

i j-ti stupac od $\hat{A}$ i $\hat{\Gamma}_1$ zadovoljavaju

$$\begin{pmatrix} -\hat{\rho}_j\hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & -\hat{\rho}_j\hat{\Sigma}_{22} \end{pmatrix} \begin{pmatrix} \hat{\alpha}^{(j)} \\ \hat{\gamma}^{(j)} \end{pmatrix} = 0, \tag{31}$$

$$\hat{\alpha}^{(j)T}\hat{\Sigma}_{11}\hat{\alpha}^{(j)} = 1, \quad \hat{\gamma}^{(j)T}\hat{\Sigma}_{22}\hat{\gamma}^{(j)} = 1. \tag{32}$$

U populaciji, kanoničke korelacije i kanoničke varijable nalazimo u obliku maksimiziranih korelacija linearnih kombinacija dva skupa varijabli. Cijeli postupak iz prethodnog poglavlja može biti proveden u terminima uzorka. Tako $\hat{\alpha}^{(1)T}x_\alpha^{(1)}$ i $\hat{\gamma}^{(1)T}x_\alpha^{(2)}$ imaju maksimalnu uzoračku korelaciju između bilo koje linearne kombinacije od $x_\alpha^{(1)}$ i $x_\alpha^{(2)}$, i ta korelacija je $\hat{\rho}_1$. Slično, $\hat{\alpha}^{(2)T}x_\alpha^{(1)}$ i $\hat{\gamma}^{(2)T}x_\alpha^{(2)}$ imaju drugu po veličini uzoračku korelaciju $\hat{\rho}_2$, i tako dalje.
Možemo odrediti kanoničke uzoračke varijable i korelacije u terminu nepristranog procjenitelja. Kako je

$$E\hat{X} = \frac{1}{N}E\sum_{\alpha=1}^{N} X_\alpha = \mu,$$

to znači da je aritmetička sredina uzorka nepristran procjenitelj očekivanja populacije. Međutim, $\hat{\Sigma}$ nije nepristran procjenitelj za Σ, jer je

$$E\hat{\Sigma} = \frac{N-1}{N}\Sigma.$$

Stoga definiramo matricu S,

$$S = \frac{1}{N-1}\sum_{\alpha=1}^{N}(x_{\alpha-\hat{x}})(x_{\alpha-\hat{x}})^T = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$$

koju zovemo **uzoračku matricu kovarijance**. S je nepristran procjenitelj od Σ. Neka je $a^{(j)} = \sqrt{\frac{N-1}{N}}\hat{\alpha}^{(j)}$, $b^{(j)} = \sqrt{\frac{N-1}{N}}\hat{\gamma}^{(j)}$. Tada $\hat{\rho}$ zadovoljava

$$\begin{pmatrix} -\hat{\rho}S_{11} & S_{12} \\ S_{21} & -\hat{\rho}S_{22} \end{pmatrix} \begin{pmatrix} a^{(j)} \\ b^{(j)} \end{pmatrix} = 0,$$

odnosno, $\hat{\rho}$ zadovoljava

$$S_{12}b^{(j)} = \hat{\rho}_j S_{11}a^{(j)}, \tag{33}$$

$$S_{21}a^{(j)} = \hat{\rho}_j S_{22}b^{(j)}, \tag{34}$$

pri čemu je

$$a^{(j)T}S_{11}a^{(j)}=1, \quad b^{(j)T}S_{22}b^{(j)}=1. \tag{35}$$

Linearne kombinacije $a^{(j)T}x_\alpha^{(1)}$ i $b^{(j)T}x_\alpha^{(2)}$ zovemo **kanoničke uzoračke varijate**. Također, kanoničke uzoračke varijate možemo dobiti iz **uzoračke matrice korelacije** R koju definiramo na sljedeći način

$$R=\left(\frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}}\right)=(r_{ij})=\begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}. \tag{36}$$

Definirajmo sljedeće matrice

$$S_1=\begin{pmatrix} \sqrt{s_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{s_{22}} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sqrt{s_{p_1p_1}} \end{pmatrix}, \tag{37}$$

$$S_2=\begin{pmatrix} \sqrt{s_{p_1+1,p_1+1}} & 0 & \cdots & 0 \\ 0 & \sqrt{s_{p_1+2,p_1+2}} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sqrt{s_{pp}} \end{pmatrix}. \tag{38}$$

Tada jednakosti (34) do (36) možemo zapisati u obliku

$$R_{12}(S_2b^{(j)})=\hat{\rho}_jR_{11}(S_1a^{(j)}), \tag{39}$$

$$R_{21}(S_1a^{(j)})=\hat{\rho}_jR_{22}(S_2b^{(j)}), \tag{40}$$

pri čemu je

$$(S_1a^{(j)})^TR_{11}(S_1a^{(j)})=1, \quad (S_2b^{(j)})^TR_{22}(S_2b^{(j)})=1. \tag{41}$$

Prilikom računanja je prikladno koristiti matrice R_{ij} (zbog $-1<r_{ij}<1$) kako bismo dobili $S_1a^{(j)}$ i $S_2b^{(j)}$, iz čega se mogu izračunati $a^{(j)}$ i $b^{(j)}$.

3.2 Standardizirani koeficijenti

Koeficijenti u kanoničkim varijatama $U_i = \alpha_i^T X^{(1)}$ i $V_i = \gamma_i^T X^{(2)}$ odražavaju razlike u raspršenosti varijabla kao i razlike u doprinosu varijabla kanoničkoj korelaciji te su jedinstveni do na skaliranje. Kako bismo uklonili efekt skaliranja, α_i i γ_i se mogu standardizirati množenjem standardnom devijacijom odgovarajućih varijabli,

$$c_i = S_1\alpha_i, \quad d_i = S_2\gamma_i,$$

gdje su $S_1 = diag(\sqrt{s_{11}}, ..., \sqrt{s_{p_1p_1}})$ i $S_2 = diag(\sqrt{s_{p_1+1p_1+1}}, ..., \sqrt{s_{pp}})$. Vektore c_i i d_i možemo direktno dobiti računanjem svojstvenih vektora matrica $R_{11}^{-1}R_{12}R_{22}^{-1}R_{21}$ i $R_{22}^{-1}R_{21}R_{11}^{-1}R_{12}$, respektivno. Vektore c_i i d_i nazivamo **vektori standardiziranih koeficijenata**. Primjenjuju se na standardizirane varijable. U terminima centriranih varijabli $x^{(1)} - \overline{x}^{(1)}$ imamo

$$\begin{aligned} U &= \alpha^T(x^{(1)} - \overline{x}^{(1)}) \\ &= \alpha^T S_1 S_1^{-1}(x^{(1)} - \overline{x}^{(1)}) \\ &= c^T S_1^{-1}(x^{(1)} - \overline{x}^{(1)}) \\ &= c_1 \frac{x_1^{(1)} - \overline{x_1}^{(1)}}{\sqrt{s_{11}}} + \cdots + c_{p_1} \frac{x_{p_1}^{(1)} - \overline{x_{p_1}}^{(1)}}{\sqrt{s_{p_1p_1}}}. \end{aligned}$$

Time su razlike u varijabilnosti uklonjene, a koeficijenti u c_i odražavaju relativni doprinos svake $x^{(1)}, ..., x_{p_1}^{(1)}$ u U_i. Analogno zaključujemo za vektore d_i. Standardizirani koeficijenti prikazuju doprinos varijabla u prisutnosti jednih od drugih. Ako su neke od varijabla uklonjene ili dodane, koeficijenti će se promijeniti.

4 Statistička značajnost kanoničkih korelacija

Neka je X slučajni vektor iz $N(\mu, \Sigma)$. Podijelimo vektor X na dva skupa, $X^{(1)}$ i $X^{(2)}$. Hipoteza o nezavisnosti $X^{(1)}$ i $X^{(2)}$ može se izraziti kao

$$H_0 : \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix} \quad ili \quad H_0 : \Sigma_{12} = 0.$$

Statistička značajnost kanoničkih korelacija određuje se potvrđivanjem postojanja korelacije između kanoničkih varijata, odnosno linearnih kombinacija podskupova od $X^{(1)}$ i $X^{(2)}$ oblika $U = \alpha^T X^{(1)}$ i $V = \gamma^T X^{(2)}$, za sve U i V. U tom postupku postavlja se nul-hipoteza

$$H_0 : \rho(U, V) = 0, \tag{42}$$

gdje je

$$\rho(U, V) = \frac{Cov(U, V)}{\sqrt{Var(U)Var(V)}} = \frac{\alpha^T \Sigma_{12} \gamma}{\sqrt{\alpha^T \Sigma_{11} \alpha \gamma^T \Sigma_{22} \gamma}}.$$

Hipotezi H_0 suprotstavljamo alternativnu hipotezu H_1, da postoji barem jedna linearna kombinacija takva da je korelacija $\rho(U, V)$ različita od nula. Kako vrijede ograničenja $Var(U) = EU^2 = 1$ i $Var(V) = EV^2 = 1$, tada je

$$\rho(U, V) = \alpha^T \Sigma_{12} \gamma. \tag{43}$$

S obzirom da je $\rho(U, V) = Cov(U, V) = (\Lambda \quad 0)$, nul-hipoteza je ekvivalentna $\Lambda = 0$, odnosno $\rho_1 = \rho_2 = \cdots = \rho_{p_1} = 0$. Stoga je hipoteza o nepostojanju veze između $X^{(1)}$ i $X^{(2)}$ ekvivalentna hipotezi o nepostojanju korelacije između kanoničkih varijata.
Za testiranje značajnosti kanoničke korelacije može se koristiti kvocijent vjerodostojnosti. Kvocijent vjerodostojnosti za testiranje prve kanoničke korelacije je oblika

$$\Lambda_1 = \frac{|S|}{|S_{11}||S_{22}|} = \frac{|R|}{|R_{11}||R_{22}|}, \tag{44}$$

gdje je S uzoračka matrica kovarijanci, a R uzoračka matrica korelacija. Λ_1 može se izraziti u terminima kvadrata kanoničkih korelacija. Tada je

$$\Lambda_1 = \prod_{i=1}^{p_1} (1 - \hat{\rho}_i^2), \tag{45}$$

gdje su $\hat{\rho}_1 \geq \hat{\rho}_2 \geq \cdots \geq \hat{\rho}_{p_1} \geq 0$ uzoračke kanoničke korelacije. Ukoliko je jedan ili više $\hat{\rho}_i^2$ velik, tada je Λ_1 mali. Statističku značajnost kanoničkih korelacija možemo testirati Bartlettovim χ^2 - testom (Bartlett, 1941). Za testiranje značajnosti prve kanoničke korelacije koristimo χ^2 - aproksimaciju

$$\chi^2 = -[(N - \frac{1}{2}(p_1 + p_2 + 3)) \ln \Lambda_1, \tag{46}$$

koja ima χ^2 - distribuciju sa $p_1 p_2$ stupnjeva slobode ako je H_0 istinita hipoteza. Nul-hipotezu odbacujemo ako je $\chi^2 > \chi^2_\alpha$, gdje je α razina značajnosti (npr. $\alpha = 0.05$). Alternativno, može se koristiti F - aproskimacija dana s

$$F = \frac{1 - \Lambda_1^{\frac{1}{t}}}{\Lambda_1^{\frac{1}{t}}} \frac{df_2}{df_1},$$

koja ima aproksimacijsku F - distribuciju s df_1 i df_2 stupnjeva slobode ako je H_0 istinita hipoteza, pri čemu je

$$\begin{aligned} df_1 &= p_1 p_2, \\ df_2 &= wt - \frac{p_1 p_2}{2} + 1, \\ w &= N - \frac{p_1 + p_2 + 3}{2}, \\ t &= \sqrt{\frac{p_1^2 p_2^2 - 4}{p_1^2 p_2^2 - 5}}. \end{aligned}$$

Nul- hipoteza H_0 se odbacuje ako $F > F_\alpha$. Sljede još tri test-statistike koje se koriste za testiranje postojanja korelacije.
Hotelling - Lawly statistika za kanoničke korelacije je

$$\Lambda_{HL} = \sum_{i=1}^{p_1} \frac{\hat{\rho}_i^2}{1 - \hat{\rho}_i^2}.$$

Pillai test - statistika:

$$\Lambda_{PB} = \sum_{i=1}^{p_2} \hat{\rho}_i^2.$$

Roy (1953) je predložio da kriterij bude maksimalan karakterističan korijen, a test-statistika je oblika

$$\Lambda_R = \hat{\rho}^2.$$

Nul-hipoteza se odbacuje ako Λ_R premašuje vrijednosti koje se mogu naći u tablici navedenoj u knjizi [7]. Za testiranje značajnosti kanoničkih korelacija $\rho_2, ..., \rho_{p_1}$, izbacimo $\hat{\rho}_1^2$ iz Λ_1 kako bismo dobili

$$\Lambda_2 = \prod_{i=2}^{p_1} (1 - \hat{\rho}_i^2).$$

Općenito, u k-tom koraku testiramo značajnost kanoničkih korelacija $\rho_k, ..., \rho_{p_1}$, ako je poznato da je prvih $k-1$ kanoničkih korelacija različito od nula. Test - statistika je oblika

$$\Lambda_k = \prod_{i=k}^{p_1} (1 - \hat{\rho}_i^2).$$

Korištenjem Λ_k u χ^2 - aproksimaciji dobivamo

$$\chi^2 = -[(N - \frac{1}{2}(p_1 + p_2 + 3)) \ln \Lambda_k,$$

koja ima χ^2 - distribuciju sa $(p_1 - k + 1)(p_2 - k + 1)$ stupnjeva slobode. Analogno, primjenom Λ_k u F - aproksimaciji dobijemo

$$F = \frac{1 - \Lambda_k^{\frac{1}{t}}}{\Lambda_k^{\frac{1}{t}}} \frac{df_2}{df_1},$$

gdje je

$$\begin{aligned}
df_1 &= (p_1 - k + 1)(p_2 - k + 1), \\
df_2 &= wt - \frac{(p_1 - k + 1)(p_2 - k + 1)}{2} + 1, \\
w &= N - \frac{(p_1 - k + 1) + (p_2 - k + 1) + 3}{2}, \\
t &= \sqrt{\frac{(p_1 - k + 1)^2 (p_2 - k + 1)^2 - 4}{(p_1 - k + 1)^2 (p_2 - k + 1)^2 - 5}}.
\end{aligned}$$

Ostale modifikacije testova mogu se pronaći u knjizi [7]. Niti jedan od ovih testova se ne preferira. Jačina različitih testova, koja predstavlja vjerojatnost odbacivanja H_0 kad je H_0 pogrešna, ovisi o alternativnoj hipotezi H_1. Uspoređujući ove testove pokazano je da u slučaju kad su kanoničke korelacije različite od nula te se u velikoj mjeri razlikuju, kriterij Hotteling - Lawely testa je najjači. S druge strane, u slučaju kad su kanoničke korelacije približno jednake, najbolje se pokazuje Pillai test. U oba slučaja, Wilks test se pokazuje kao drugi najbolji test. Royov test je najjači samo ako se testira jednodimenzionalna linearna veza, a to je ekvivalentno hipotezi

H_1 gdje su sve kanoničke korelacije jednake nuli osim jedne. Testiranje statističke značajnosti provodi se redom od prve do posljednje kanoničke korelacije. Ukoliko jedna od kanoničkih korelacija nije statistički značajna, tada niti jedna sljedeća kanonička korelacija ne može biti statistički značajna.

5 Primjer

Baza podataka koja se koristi u ovom primjeru preuzeta je s internetske stranice `https://stats.idre.ucla.edu/`. Sadrži osam varijabli, *control, concept, motivation, read, write, math, science, sex* , a svaka varijabla sadrži 600 opservacija. Ove varijable podijelimo u dva skupa. Prvi skup sadrži psihološke karakteristike osobe i sastoji se od 3 varijable, a drugi skup sadrži akademske karakteristike i indikator varijablu koja određuje spol osobe (0 za muškarce, 1 za žene) te se sastoji od ukupno pet varijabli.
Zanima nas kako je skup psiholoških karakteristika povezan sa skupom akademskih karakteristika osobe zajedno sa spolom. Zanima nas koliko je kanoničkih korelacija potrebno za razumjeti povezanost između ova dva skupa varijabli.
Osim računanja svojstvenih vrijednosti i svojstvenih vektora kako bismo došli do kanoničkih korelacija i svojstvenih vrijednosti, postoje već gotove funkcije u različitim programima koje računaju tražene vrijednosti. Tako će se u ovom primjeru koristiti i funkcija *cancor* iz programa *R*.
a. Nestandardizirani podaci
Dobivene su sljedeće uzoračke matrice kovarijanci :

$$S_{11} = \begin{pmatrix} 0.449275 & 0.080953 & 0.056313 \\ 0.080953 & 0.497748 & 0.069776 \\ 0.056313 & 0.069776 & 0.117463 \end{pmatrix},$$

$$S_{12} = \begin{pmatrix} 2.529718 & 2.339679 & 2.128339 & 2.111976 & 0.037886 \\ 0.432342 & 0.133459 & 0.356007 & 0.478159 & -0.044287 \\ 0.729256 & 0.847547 & 0.629251 & 0.384785 & 0.016757 \end{pmatrix},$$

$$S_{22} = \begin{pmatrix} 102.070264 & 61.769243 & 64.610611 & 67.730284 & -0.210183 \\ 61.769243 & 94.603927 & 57.934534 & 53.731563 & 1.184407 \\ 64.610611 & 57.934534 & 88.637261 & 59.354421 & -0.226249 \\ 67.730284 & 53.731563 & 59.354421 & 94.209905 & -0.668464 \\ -0.210183 & 1.184407 & -0.226249 & -0.668464 & 0.248389 \end{pmatrix}.$$

Napravimo Cholesky dekompoziciju matrica S_{11} i S_{22}, a zatim računamo svojstvene vrijednosti i pripadajuće svojstvene vektore sljedećih matrica :

$$\Sigma_{11}^{-T/2}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1/2} = \begin{pmatrix} 0.180621 & -0.020758 & 0.070570 \\ -0.020758 & 0.021917 & -0.018581 \\ 0.070570 & -0.018581 & 0.051712 \end{pmatrix},$$

$$\Sigma_{22}^{-T/2}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1/2} = \begin{pmatrix} 0.155942 & 0.077611 & 0.021980 & 0.000241 & 0.045214 \\ 0.077611 & 0.047607 & 0.008167 & -0.011290 & 0.030522 \\ 0.021980 & 0.008167 & 0.003961 & 0.003461 & 0.003628 \\ 0.000241 & -0.011290 & 0.003461 & 0.015821 & -0.006364 \\ 0.045214 & 0.030522 & 0.003628 & -0.006364 & 0.030918 \end{pmatrix}$$

Svojstvene vrijednosti su 0.21537589, 0.02805932 i 0.01081416. Kanoničke korelacije su pozitivni korijeni svojstvenih vrijednosti, stoga su kanoničke korelacije $\rho_1 = 0.4641$, $\rho_2 = 0.1675$ i $\rho_3 = 0.1040$. Pripadajući svojstveni vektori su stupci sljedećih matrica:

$$\hat{A} = \begin{pmatrix} 1.253834 & -0.621478 & -0.661690 \\ -0.351350 & -1.187687 & 0.826721 \\ 1.262420 & 2.027264 & 2.000228 \end{pmatrix},$$

$$\hat{\Gamma} = \begin{pmatrix} 0.044621 & -0.004910 & 0.021381 & 0.148425 & -0.007857 \\ 0.035877 & 0.042071 & 0.091307 & -0.071612 & -0.090126 \\ 0.023417 & 0.004229 & 0.009398 & -0.050766 & 0.152340 \\ 0.005025 & -0.085162 & -0.109835 & -0.052016 & -0.050252 \\ 0.632119 & 1.084642 & -1.794647 & 0.263460 & 0.333744 \end{pmatrix},$$

gdje je $\hat{\Gamma} = (\hat{\Gamma}_1 \quad \hat{\Gamma}_2)$. Korištenjem funkcije *cancor* dobivene kanoničke korelacije su jednake $\rho_1 = 0.4641$, $\rho_2 = 0.1675$ i $\rho_3 = 0.1040$, a koeficijenti su

$$\hat{A}_{cancorc} = \begin{pmatrix} -0.051230 & -0.025393 & -0.027036 \\ 0.014356 & -0.048528 & 0.033779 \\ -0.051581 & 0.082832 & 0.081727 \end{pmatrix},$$

$$\hat{\Gamma}_{cancor} = \begin{pmatrix} -0.001823 & -0.000201 & 0.000874 & 0.000613 & -0.006042 \\ -0.001466 & 0.001719 & 0.003731 & -0.004088 & 0.002327 \\ -0.000957 & 0.000173 & 0.000384 & 0.005833 & 0.003004 \\ -0.000205 & -0.003480 & -0.004488 & -0.002355 & 0.001785 \\ -0.025828 & 0.044317 & -0.073327 & 0.015126 & -0.008546 \end{pmatrix}.$$

Usporedimo li dobivene kanoničke koeficijente, vidimo da su različiti. S obzirom da su koeficijenti jedinstveni do na skaliranje, matricu $\hat{A}$ pomnožimo s dijagonalnom matricom $diag(-1,1,1)$ i podijelimo s $\sqrt{599}$, a matricu $\hat{\Gamma}_1$ pomnožimo dijagonalnom matricom $diag(-1,1,1)$ i podijelimo s $\sqrt{599}$. Usporedbom korigiranih koeficijenta

$$\sum(\hat{A}_{cancor} - \hat{A}\cdot diag(-1,1,1)\cdot 1/\sqrt{599})^2 = 1.836856\cdot 10^{-32},$$
$$\sum(\hat{\Gamma}_{1cancor} - \hat{\Gamma}_1\cdot diag(-1,1,1)\cdot 1/\sqrt{599})^2 = 2.876669\cdot 10^{-32},$$

zaključujemo da se radi o ekvivalentim procedurama. Definiramo kanoničke varijate

$$U = X^{(1)}\hat{A}, \quad V = X^{(2)}\hat{\Gamma}_1$$

pri čemu vrijedi

$$\hat{A}^T S_{11}\hat{A} = I,$$
$$\hat{\Gamma}^T S_{22}\hat{\Gamma} = I,$$
$$\hat{A}^T S_{12}\hat{\Gamma} = \begin{pmatrix} 0.4641 & 0 & 0 & 0 & 0 \\ 0 & 0.1675 & 0 & 0 & 0 \\ 0 & 0 & 0.1040 & 0 & 0 \end{pmatrix}$$

b. Standardizirani podaci

Koristi se isti postupak kao i kod nestandardiziranih podataka za računanje kanoničkih korelacija i kanoničkih koeficijenata. Dobivene svojstvene vrijednosti su 0.21537589, 0.02805932 i 0.01081416, a kanoničke korelacije su , $\rho_1 = 0.4641$, $\rho_2 = 0.1675$ i $\rho_3 = 0.1040$. Standardizacijom podataka svojstvene vrijednosti ostaju ne promijenjene, pa tako i kanoničke korelacije. Ali mijenjaju se svojstveni vektori.

$$\hat{A} = \begin{pmatrix} 0.840420 & -0.416564 & 0.443517 \\ -0.247882 & -0.837928 & -0.583262 \\ 0.432669 & 0.694803 & -0.685537 \end{pmatrix},$$
$$\hat{\Gamma} = \begin{pmatrix} 0.450801 & -0.049606 & 0.216008 & -0.079380 & 1.499536 \\ 0.348957 & 0.409206 & 0.888097 & -0.876611 & -0.696534 \\ 0.22047 & 0.039819 & 0.088481 & -1.434238 & -0.477950 \\ 0.048775 & -0.826599 & -1.066078 & -0.487757 & -0.504881 \\ 0.315040 & 0.540571 & -0.894428 & 0.166333 & 0.131305 \end{pmatrix},$$

gdje je $\hat{\Gamma} = (\hat{\Gamma}_1 \quad \hat{\Gamma}_2)$. Korištenjem funkcije *cancor* dobiveni su sljedeći koeficijenti

$$\hat{A}_{cancorc} = \begin{pmatrix} -0.034339 & -0.017020 & -0.018122 \\ 0.010128 & -0.034237 & 0.023831 \\ -0.017678 & 0.028389 & 0.028010 \end{pmatrix},$$

$$\hat{\Gamma}_{cancor} = \begin{pmatrix} -0.018419 & -0.002027 & 0.008826 & 0.006190 & -0.061042 \\ -0.014258 & 0.016720 & 0.036287 & -0.039758 & 0.022631 \\ -0.009008 & 0.001627 & 0.003615 & 0.054914 & 0.028283 \\ -0.001993 & -0.033774 & -0.043559 & -0.022857 & 0.017329 \\ -0.012872 & 0.022087 & -0.036545 & 0.007538 & -0.004259 \end{pmatrix}.$$

Dobiveni koeficijenti nisu jednaki. Množenjem odgovarajućom dijagonalnom matricom i skalarom $1/\sqrt{599}$ dobivamo

$$\sum (\hat{A}_{cancor} - \hat{A} \cdot diag(-1, 1, -1) \cdot 1/\sqrt{599})^2 = 9.942613 \cdot 10^{-33},$$
$$\sum (\hat{\Gamma}_{1cancor} - \hat{\Gamma}_1 \cdot diag(-1, 1, 1) \cdot 1/\sqrt{599})^2 = 4.057976 \cdot 10^{-33},$$

što potvrđuje da su procedure ekvivalentne. Definiramo kanoničke varijate

$$U = X^{(1)}\hat{A}, \quad V = X^{(2)}\hat{\Gamma}_1,$$

pri čemu vrijedi

$$\hat{A}^T S_{11} \hat{A} = I,$$
$$\hat{\Gamma}^T S_{22} \hat{\Gamma} = I,$$
$$\hat{A}^T S_{12} \hat{\Gamma} = \begin{pmatrix} 0.4641 & 0 & 0 & 0 & 0 \\ 0 & 0.1675 & 0 & 0 & 0 \\ 0 & 0 & 0.1040 & 0 & 0 \end{pmatrix}$$

Potrebno je provjeriti koje od dobivnih kanoničkih korelacija su statistički značajne. Koristimo funkciju *p.asym()* pri čemu određujemo koju metodu odnosno test želimo koristiti, na primjer *tstat='Wilks'*. Dobivene vrijednosti prikazane su u Tablici 1.

Za testiranje nul-hipoteze o nepostojanju korelacije između kanoničkih varijata koristili smo četiri testa. Prva tri testa ukazuju da su prva i druga kanonička korelacija statistički značajne (p-vrijednosti su manje od 0.05), dok treća kanonička korelacija nije statistički značajna (p-vrijednosti veći od 0.05). Royov test potvrđuje značajnost prve kanoničke korelacije. Prema ovim rezultatima, interpretiramo vezu između prvog para kanoničkih varijata (U_1 i V_1), koje imaju maksimalnu korelaciju

tstat	kanonička korelacija	stat	approx F	df1	df2	p-value
Wilks	0.4641	0.75436	11.7157	15	1634.7	0.00000
	0.1675	0.96143	2.9445	8	1186.0	0.002905
	0.01081	0.98919	2.1646	3	594.0	0.091092
Hotelling	0.4641	0.3143	12.3763	15	1772	0.00000
	0.1675	0.0398	2.9486	8	1778	0.002807
	0.01081	0.0109	2.1670	3	1784	0.090013
Pillai	0.4641	0.2542	11.0006	15	1782	0.00000
	0.1675	0.03887	2.9341	8	1788	0.0029326
	0.01081	0.0108	2.1634	3	1794	0.0904404
Roy	0.4641	0.21538	32.61008	5	594	0

Tablica 1: Rezultati testiranja značajnosti kanoničkih korelacija

0.4641. Standardizirani kanonički koeficijenti dani su za prvi par kanoničkih varijata u sljedećoj tablici.

	1
Control	0.8404
Concept	-0.2479
Motivation	0.4327
Read	0.4508
Write	0.3490
Math	0.2205
Science	0.0488
Sex (Female=1)	0.3150

Tablica 2: Standardizirani kanonički koeficijenti

Među psihološkim varijablama za prvu kanoničku korelaciju najveći utjecaj ima varijabla *control*. Među akademskim varijablama plus spol, za prvu kanoničku korelaciju najveći utjecaj imaju varijable *read* i *write*.

Kaničke varijate (U_1 i V_1) koje predstavljaju optimalne linearne kombinacije, mogu biti određene korištenjem standardiziranih kanoničkih koeficijentata kao,

$$U_1 = 0.8404Control - 0.2479Concept + 0.4327Motivation$$

$$V_1 = 0.4508Read + 0.349Write + 0.2205Math + 0.0488Science + 0.315Sex$$

Prema tome, ukoliko se vrijednost psiholoških varijabli povećava (osim varijable *concept*), varijable u akademskoj varijati V_1 će se povećati. Sljdeće tablice sadrže vrijednosti kanoničkih opterećenja i kanoničkih kros-opterećenja.
Kanonička opterećenja za akademske komponente ukazuju da najveći utjecaj ima varijabla *write* u formiranju kanoničke varijate V_1, dok za formiranje kanoničke vari-

	Control	Concept	Motivation		Read	Write	Math	Science	Sex
U_1	-0.9040	-0.0208	-0.5672	V_1	-0.8404	-0.8765	-0.7639	-0.6584	-0.3641

Tablica 3: Kanonička opterećenja između originalne varijate i kanoničke varijate

	Read	Write	Math	Science	Sex		Control	Concept	Motivation
U_1	-0.3900	-0.4068	-0.3545	-0.3056	-0.1690	V_1	-0.4196	-0.0097	-0.2632

Tablica 4: Kanonička kros-opterećenja između originalne varijate i suprotne kanoničke varijate

jate U_1 najveći utjecaj ima varijabla *control*. Kros-opterećenja ukazuju da najmanji utjecaj prilikom formiranja kanoničke varijate U_1 ima varijabla *sex*, a pri formiranju V_1 najmanji utjecaj ima varijabla *concept*.

RI_{U_1}	RI_{U_2}	RI_{U_3}	Ukupno
0.081799	0.007270	0.003905	0.092974

Tablica 5: Redundacija za psihološke varijable i ukupna kanonička redundacija

RI_{V_1}	RI_{V_2}	RI_{V_3}	Ukupno
0.1130458	0.0070132	0.0009804	0.1210394

Tablica 6: Redundacija za akademske varijable i ukupna kanonička redundacija

Prva kanonička varijata za psihološke karakteristike, U_1, objašnjava 8.18% varijance u akademskoj kanoničkoj varijati, dok ukupno objašnjavaju 9.3% varijance. Prva kanonička varijata za akademske karakteristike objašnjava 11.3% varijance u psihološkoj kanoničkoj varijati, dok ukupno objašnjavaju 12.1% varijance.

Sažetak

Kanonička korelacija jedna je od procedura za objašnjavanje povezanosti između dva skupa varijata. Potrebno je pronaći linearne kombinacije varijata tako da je njihova korelacija maksimalna. Parovi čija je korelacija maksimalna, te nisu u korelaciji s ostalim linearnim kombinacijama nazivamo kanoničke varijate, a njihovu korelaciju nazivamo kanonička korelacija. Koeficijenti kojim se postiže maksimalna korelacija nazivaju se kanonički koeficijenti. Može se proučavati i povezanost između varijabli originalnog vektora i kanoničke varijate, te dobivene vrijednosti nazivamo kanonička opterećenja. Također, proučava se i veza između originalne varijate i suprotne kanoničke varijate, što se naziva kanonička kros-opterećenja. Kanoničke korelacije i kanoničke koeficijente dobivamo računanjem svojstvenih vrijednosti i svojstvenih vektora. Tako su pozitivni korijeni svojstvenih vrijednosti tražene kanoničke korelacije, a svojstveni vektori traženi kanonički koeficijenti. S obzirom da većina podataka nije iste varijabilnosti, potrebno je standardizirati podatke. Algebra korištena za računanje kanoničkih koeficijenata i kanoničke korelacije koristi se i za računanje njihovih procjena iz podataka. U primjeru smo proučavali vezu između dvije varijate, od kojih prvi varijat sadrži psihološke karakteristike osobe, a drugi akademske karakteristike osobe te interpretiramo dobivene rezultate. Za interpretaciju koristimo prvi par kanoničkih varijata s obzirom da je njihova korelacija najveća, i statistički značajna.

Ključne riječi

kanonička varijata, kanonička korelacija, kanoničke funkcije, kanonički koeficijenti, kanonička opterećenja, kanonička kros-opterećenja, indeks redundancije

Abstract

Canonical correlation is one of the procedures for explaining the relationship between the two sets of variates. The procedure is consisted of finding the linear combinations of variables that have the maximal correlation. The pairs that have the maximal correlation, but are not correlated with other linear combinations, are called canonical variates, while the correlation between the canonical variates is called the canonical correlation. Coefficients that maximize the correlation are the canonical coefficients. It is possible to study the relationship between the variables of the original vector and the canonical variate, where the obtained values are called canonical loadings. The correlation between variables of the original vector and the opposite canonical variate, which is called the canonical cross-loadings, can also be studied. As for the canonical correlations and canonical coefficients, they are obtained by the computing of eigenvalues and eigenvectors, so that the positive square roots of eigenvalue are canonical correlation, and eigenvectors are canonical coefficients. Given that the majority of data are not of the same variability, it is necessary to standardize them. Algebra used to calculate canonical coefficients and canonical correlations was also applied in the calculation of their estimates. The example included in this thesis studies the relationship between two variates. The first variate contains the psychological characteristics of the person, and the second is consisted of the academic characteristics. The first pair of canonical variates is used for the interpretation of the obtained results, since their correlation is the largest and statistically significant.

Key words

canonical variate, canonical correlation, canonical functions, canonical coefficients, canonical loadings, canonical cross-loadings, the index of redundancy

Životopis

Zovem se Ela Kovačić. Rođena sam 10. veljače 1994. godine u Varaždinu. Pohađala sam Osnovnu školu Sračinec te sam nakon toga upisala Drugu gimnaziju u Varaždinu, opći smjer. Godine 2012. upisujem preddiplomski studij matematike na Odjelu za matematiku u Rijeci. Preddiplomski studij sam završila 2017. godine uz završni rad na temu Neriješeni problemi u teoriji brojeva pod mentorstvom dr.sc. Ane Jurasić. Iste godine upisujem diplomski studij Financijske matematike i statistike na Odjelu za matematiku u Osijeku. Stručnu praksu sam odradila u Raiffeisen banci u Zagrebu.

Literatura

[1] Anderson, T. W. : *An introduction to multivariate statistical analysis*, Third edition, Standford University, Department of Statistics, Standford, CA, (2003)

[2] Dizdar, Dražen:*Kvantitativne metode*, Kineziološki fakultet Sveučilišta u Zagrebu, Zagreb, (2006.)

[3] Ewerbring, L. M., and F. T. Luk : *Canonical corelations and generalized SVD*, J. of Computation adn Applied Mathematics, Vol.27, Nizozemska (1989),

[4] Friederichs, P., and A. Hense : *Statistical Inference in Canonical Correlation Analyses Exemplified by the Influence of North Atlantic SST on European Climate*, J. Climate, Vol. 16

[5] Hotelling, Harold : *Relations Between Two Sets of Variates*, Biometrika, vol. 28, no. 3/4, (1936)

[6] Kendal, Sir Maurice G., Alan. F. Stuart : *The Advanced Theory Of Statistics*, Vol. 3, Charles Griffin and Co. Ltd., (1966)

[7] Rencher, A.C., *Method of Multivariate Analysis*, Second edition, Brigham Young University, John Wiley and Sons, Inc. Publication, (2003)