ANDREWSOVE KRIVULJE V KORESPONDENČNI ANALIZI

Jože ROVAN - Ekonomska fakulteta, Univerza v Ljubljani

Povzetek

V korespondenčni analizi lahko predstavimo vrstice in stolpce matrike nenegativnih podatkov, npr. kontingenčne tabele, kot točke večrazsežnega prostora. Koordinate teh točk so vrednosti posplošenih glavnih komponent (PGK), določene na podlagi matrike podatkov. Grafične metode lahko pomembno prispevajo k boljšemu poznavanju narave proučevanega pojava (glej npr. Benzécri, Greenacre, Lebart, Morineau).

Vektorji relativnih frekvenc vrstic in stolpcev matrike podatkov, imenovani profili, so običajno predstavljeni kot točke v dvorazsežnem razsevnem grafikonu. Položaj točk je v tem primeru določen z vrednostmi prvih dveh PGK . Če je delež variacije prvih dveh PGK v skupni variaciji relativno nizek, potem je položaj profilov slabo predstavljen z njihovimi projekcijami na optimalni dvorazsežni podprostor, prikazanimi v dvorazsežnem razsevnem grafikonu. Posledica tega so lahko pomankljivi sklepi o proučevanem pojavu.

V tem članku je položaj profilov prikazan z dvema alternativnima grafičnima metodama:

- s trirazsežnim razsevnim grafikonom in
- z Andrewsovim grafikonom.

V trirazsežnem razsevnem grafikonu je položaj profilov določen z vrednostmi prvih treh PGK . Posledica tega je večji delež pojasnjene variance in boljši prikaz položaja profilov kot v primeru dvorazsežnega razsevnega grafikona.

Pogosto pa se zgodi, da so nekateri profili slabo predstavljeni tudi v trirazsežnem podprostoru. V tem primeru priporočamo uporabo Andrewsovega grafikona v katerem lahko na podlagi poljubnega števila PGK s posebnimi krivuljami dobro prikažemo položaj profilov. Profili-točke, ki so predstavljeni s podobnimi krivuljami, imajo podoben položaj v večrazsežnem prostoru in obratno.

Ključne besede:
korespondenčna analiza, posplošene glavne komponente, profili, dvo- in trirazsežni razsevni grafikon, Andrewsov grafikon.

Abstract

In correspondence analysis, rows and columns of a nonnegative data matrix (e.g., two-way contingency table) are displayed as points in multidimensional space. The coordinates of these points are the values of generalized principal components (GPC) determined by the data matrix. The graphical representation can lead to new insights of the explored phenomena (e.g., Benzécri, Greenacre, Lebart, Morineau).

In the standard graphical approach, profiles, e.g., vectors of relative frequencies of rows and columns of a data matrix, are represented as points in a two-dimensional scatterplot. The position of points is determined by the values of the first two GPC. When the proportion of the variation of the first two GPC in total variation is relatively low, the position of profiles is poorly represented by its projections in a two-dimensional subspace, presented by the scatterplot. This can, of course, be at least partly misleading when making conclusions about the explored phenomena.

In this paper two alternative ways of graphical representation of the position of profiles are presented:

- three-dimensional scatterplot
- Andrews' plot.

The position of points in the three-dimensional scatterplot is determined by the values of the first three GPC. Thus the proportion of the explained variation is increased and the representation of profiles is improved as compared to its representation in a two-dimensional scatterplot.

However, some profiles are frequently poorly represented even in a three-dimensional subspace. In such cases it is suggested to use Andrews' plot, by which the position of profiles is well defined, since a large number of GPC can be taken into account. Profile-points, represented by similar curves, are close together in multidimensional space and vice versa.

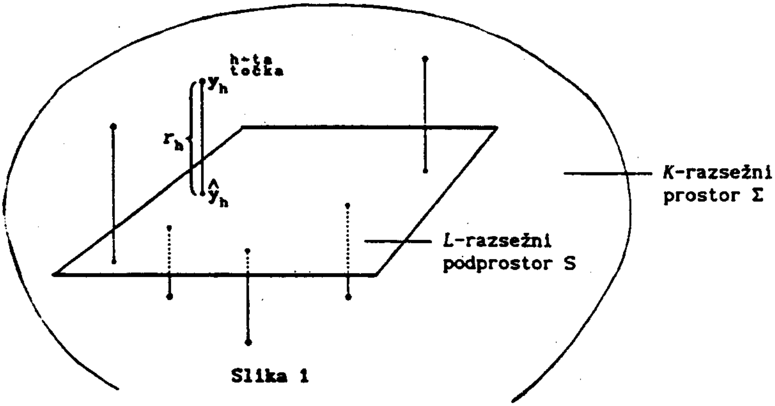
Key words

correspondence analysis, generalized principal components, profiles, two- and three dimensional scatterplot, Andrews' plot.

1. Osnove korespondenčne analize

1.1 Bivariatna korespondenčna analiza

Korespondenčna analiza (KA) je z geometrijskega vidika zelo podobna analizi glavnih komponent (AGK) (K. Pearson, 1901), zato lahko AGK vzamemo kot izhodišče za razlago KA. Vzemimo, da je množica enot predstavljena z vektorji-točkami y_h $(h = 1,2,\ldots,H)$ v večrazsežnem evklidskem prostoru Σ (slika 1). Koordinate teh točk so vrednosti proučevanih spremenljivk.



Slika 1

Cilj AGK je najti tisti podprostor S^* izbrane razsežnosti, ki se najbolje prilega množici točk večrazsežnega evklidskega prostora Σ. Oddaljenost množice točk od podprostora S je izražena z vsoto kvadratov razdalj točk y_h od njihovih pravokotnih projekcij $\hat{y}_h$ $(h = 1,2,\ldots,H)$ na podprostor S

$$\Phi(S; y_1, y_2, \ldots, y_H) = \sum_{h=1}^{H} r_h^2 = \sum_{h=1}^{H} \| y_h - \hat{y}_h \|^2 \quad . \tag{1}$$

Funkcija (1) ima minimum za podprostor S^*, ki ga zato imenujemo optimalni podprostor.

Ker je KA poseben primer posplošene AGK, jo je mogoče opisati s termini poznanimi iz AGK (glej npr. Greenacre, 1987). Izhodišče za bivariatno KA je matrika nenegativnih podatkov, ki se nanašajo na dve spremenljivki, npr. kontingenčna tabela. V KA obravnavamo dve množici točk, prva predstavlja vrstice in druga stolpce matrike podatkov. Če delimo elemente vrstic (stolpcev) matrike podatkov z vrstično (stolpično) vsoto, dobimo strukturne vrste, vsota elementov strukturne vrste pa je enaka 1 . Takšne strukturne vrste se v KA imenujejo profili. Podatki namenjeni KA so najpogosteje prikazani v obliki kontingenčne tabele, profili pa so v tem primeru vektorji relativnih frekvenc.

Če seštejemo elemente vrstic oz. stolpcev prvotne matrike podatkov dobimo zbirno vrstico oz. zbirni stolpec. Elemente teh dveh stolpcev lahko delimo z vsoto vseh elementov matrike podatkov. Tako dobimo zbirno strukturo po vrsticah in zbirno strukturo po stolpcih, katerih vsoti elementov sta tudi enaki 1 . Mogoče je pokazati, da smo tako dobili povprečni vrstični in povprečni stolpični profil (centroid). Omenjene profile je mogoče izraziti kot tehtani povprečji profilov vrstic oz. stolpcev, pri čemer so uteži proporcionalne elementom zbirne vrstice oz. zbirnega stolpca prvotne matrike podatkov.

Posebnost KA je uporaba posplošene evklidske razdalje (metrike), kjer je vsaka kvadrirana razlika koordinat deljena z ustreznim elementom povprečnega profila. Tako je kvadrat razdalje med vrstičnima profiloma p_i in $p_{i''}$ $(i = 1,2,\dots,I)$ enak

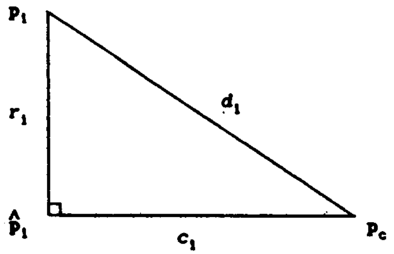
$$d^2(p_i,p_{i''}) = (p_i - p_{i''})' D_c^{-1} (p_i - p_{i''}) \quad , \tag{2}$$

pri čemer je D_c diagonalna matrika, njeni diagonalni elementi pa so elementi povprečnega vrstičnega vektorja p_c . Oblika izraza (2) je analogna statistiki χ^2, kjer so kvadrati razlik frekvenc tehtani z inverznimi vrednostmi teoretičnih frekvenc. Zato se ta specifični tip metrike imenuje χ^2-metrika.

Nadalje moramo definirati mero prileganja podprostora S množici profilov-točk prvotnega prostora Σ . V prostoru profilov-vrstic bo takšna mera tehtana vsota kvadratov razdalj vstičnih profilov-točk p_i $(i = 1,2,\dots,I)$ od njihovih pravokotnih projekcij $\hat{p}_i$ na podprostor S

$$\Phi(S;p_1,p_2,\dots,p_I) = \sum_{i=1}^{I} p_{i.} r_i^2 = \sum_{i=1}^{I} p_{i.} (p_i - \hat{p}_i)' D_c^{-1} (p_i - \hat{p}_i) \quad , \tag{3}$$

kjer so vrednosti $p_{i.}$ $(i = 1,2,\dots,I)$ elementi povprečnega stolpičnega profila p_r , r_i pa je razdalja med vrstičnim profilom-točko p_i in njegovo pravokotno projekcijo $\hat{p}_i$, izražena v χ^2-metriki. Mogoče je pokazati, da mora optimalni podprostor S^* vključevati povprečni profil (centroid). Profil-točka p_i , njegova pravokotna projekcija $\hat{p}_i$ na podprostor S^* in centroid p_c določajo oglišča pravokotnega trikotnika (slika 2). Razdalje med oglišči so izražene v χ^2-metriki. Ob tem, ko je dolžina hipotenuze d_i dana, želimo določiti tisti podprostor S^*, ki minimizira izraz $\Sigma_i p_{i.} r_i^2$ oz. ekvivalentno, ki maksimizira izraz $\Sigma_i p_{i.} c_i^2$, pri čemer je c_i razdalja med projekcijo i-tega profila-točke $\hat{\mathbf{p}}_i$ in centroidom $\mathbf{p}_c$.



Slika 2

Če je podprostor S enorazsežen želimo določiti optimalno premico. Vsota kvadratov razdalj med projekcijami profilov-točk na optimalno premico $\hat{\mathbf{p}}_i$ $(i = 1,2,\ldots,I)$ in centroidom $\mathbf{p}_c$ je tedaj enaka

$$\lambda_1 = \sum_{i=1}^{I} p_{i.} c_i^2 \quad , \tag{4}$$

in se imenuje inercija 1.PGK . Premica, vzdolž katere je maksimizirana inercija 1.PGK se imenuje os 1.PGK . Razsežnost množice profilov K je enaka rangu matrike podatkov, je torej največ

$$K = \min(I,J)-1 \quad . \tag{5}$$

V primeru K-razsežnega prostora profilov-točk obstaja K osi PGK in ustreznih inercij PGK . Vsota vseh inercij PGK se imenuje skupna inercija In in je enaka

$$\mathrm{In} = \sum_{k=1}^{K} \lambda_k = \sum_{i=1}^{I} p_{i.} d_i^2 \quad . \tag{6}$$

Skupna inercija meri skupno variabilnost matrike podatkov.

Lastne vrednosti in pripadajoče lastne vektorje lahko, podobno kot v AGK , izračunamo s spektralno dekompozicijo nesimetrične matrike

$$\mathbf{D}_r(\mathbf{P}_r-\mathbf{1}\mathbf{p}_c')\mathbf{D}_c^{-1}(\mathbf{P}_r-\mathbf{1}\mathbf{p}_c')' \tag{7}$$

katere lastne vrednosti (diagonalni elementi matrike $\mathbf{D}_\lambda$) so inercije PGK

λ_k $(k = 1,2,\ldots,K)$. Ustrezni lastni vektorji $U = [u_1, u_2, \ldots, u_K]$ morajo biti normalizirani v D_r-metriki $(U'D_rU)$, določajo položaj osi PGK ter predstavljajo novo bazo prostora profilov. Sled matrike (7) je v tem primeru enaka skupni inerciji. Profili točke so predstavljeni v prostoru, ki ga opredeljuje nekaj prvih osi PGK, njihove koordinate pa so vrednosti PGK. Matrika vrednosti PGK T je enaka

$$T = [c_{ik}] = UD_\lambda^{1/2} \quad . \tag{8}$$

Skupno inercijo In lahko razčlenimo glede na osi PGK (podobno kot skupno varianco množice spremenljivk v AGK) in glede na profile-točke

		Osi PGK 1	2	...	K	Total
V r s t i c e	1	$p_{1.}c_{11}^2$	$p_{1.}c_{12}^2$	...	$p_{1.}c_{1K}^2$	$p_{1.}d_1^2$
	2	$p_{2.}c_{21}^2$	$p_{2.}c_{22}^2$	...	$p_{2.}c_{2K}^2$	$p_{2.}d_2^2$
	:	:	:		:	
	I	$p_{I.}c_{I1}^2$	$p_{I.}c_{I2}^2$	...	$p_{I.}c_{IK}^2$	$p_{I.}d_I^2$
Total		λ_1	λ_2	...	λ_K	In

oz.

$$\begin{array}{c|c} D_r(T\cdot T) & D_r(T\cdot T)1 \\ \hline 1'D_r(T\cdot T) & 1'D_r(T\cdot T)1 \end{array} \tag{9}$$

Kjer operator $\cdot$ izraža množenje istoležnih elementov matrik.

Kvaliteto aproksimacije položaja profilov-točk lahko izrazimo z deležem vsote inercij nekaj prvih PGK v skupni inerciji In

$$\tau = \sum_{l=1}^{L} \lambda_l / \text{ In} \quad . \tag{10}$$

Prav tako pa je pomembno posvetiti pozornost kvaliteti aproksimacije položaja posameznih profilov v optimalnem podprostoru izbrane razsežnosti. Zato moramo definirati matriko kumulativnih deležev inercij profilov v skupni inerciji profilov

		Osi PGK				
		1	2	...	$K-1$	K
V	1	c_{11}^2/d_1^2	$\sum_{k=1}^{2} c_{1k}^2/d_1^2$	...	$\sum_{k=1}^{K-1} c_{1k}^2/d_1^2$	1
r s	2	c_{21}^2/d_2^2	$\sum_{k=1}^{2} c_{2k}^2/d_2^2$	...	$\sum_{k=1}^{K-1} c_{2k}^2/d_2^2$	1
t i	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
c e	I	c_{I1}^2/d_I^2	$\sum_{k=1}^{2} c_{Ik}^2/d_I^2$	...	$\sum_{k=1}^{K-1} c_{Ik}^2/d_I^2$	1

oz.

$$(\mathrm{diag}((\mathbf{T}\cdot\mathbf{T})\mathbf{1}))^{-1}(\mathbf{T}\cdot\mathbf{T})\mathbf{T} \tag{11}$$

Kumulativni deleži inercij profilov-točk so izračunani na podlagi vrednosti prve PGK, prvih dveh PGK, prvih treh PGK itd.. Matrika $\mathbf{T}$ je zgornja trikotna matrika, njeni neničelni elementi pa so vsi enaki številu 1.

Analogni zaključki, kot so bili navedeni za profile vrstic, veljajo tudi za profile stolpcev matrike podatkov.

1.2 Multipla korespondenčna analiza

Za razliko od bivariatne KA je multipla KA namenjena analizi podatkov, ki se nanašajo na več kot dve spremenljivki. Najpogostejše izhodišče za multiplo KA je Burtova kontingenčna tabela $\mathbf{B}$. To je simetrična bločna matrika, ki jo sestavljajo vse mogoče dvorazsežne kontingenčne tabele $\mathbf{F}_{qq^*}$ $(q, q^* = 1,2,\ldots,Q)$ za proučevano množico Q spremenljivk

$$\mathbf{B} = \begin{bmatrix} \mathbf{F}_{11} & \mathbf{F}_{12} & \cdots & \mathbf{F}_{1Q} \\ \mathbf{F}_{21} & \mathbf{F}_{22} & \cdots & \mathbf{F}_{2Q} \\ \vdots & \vdots & & \vdots \\ \mathbf{F}_{Q1} & \mathbf{F}_{Q2} & \cdots & \mathbf{F}_{QQ} \end{bmatrix} . \tag{12}$$

Na podlagi Burtove kontingenčne tabele je definirana matrika profilov-vrstic

$$\frac{1}{Q}\mathbf{D}_f^{-1}\mathbf{B} , \tag{13}$$

kjer je Q število proučevanih spremenljivk, $\mathbf{D}_f$ pa diagonalna matrika s frekvencami kategorij na glavni diagonali (25 kategorij). S spektralno dekompozicijo nesimetrične matrike (13) izračunamo diagonalno matriko inercij PGK $D_{\lambda,Z}$ in matriko standardiziranih koordinat PGK Γ_Z ($\Gamma_Z' D_c \Gamma_Z = I$). Skupno število kategorij vseh Q spremenljivk je J, število netrivialnih PGK pa $J-Q$.

Na podlagi spektralne dekompozicije matrike (13) ugotavljamo, da je aproksimacija prvotnega prostora z optimalnim podprostorom odvisna od vseh vrednosti Burtove kontingenčne tabele, torej tudi od diagonalnih podmatrik F_{qq} ($q = 1,2,\ldots,Q$). Podmatrike F_{qq} ne izražajo povezanosti med po dvema različnima spremenljivkama, ob tem pa močno povečajo vrednost skupne inercije in tako zmanjšajo deleže inercij PGK $\lambda_{j,Z}$ v skupni inerciji In. Zato J.P. Benzécri (1979) predlaga, da se kot analitično pomembne obravnava le tiste PGK, katerih inercije so višje od $1/Q$. Nove, bolj realistične inercije PGK, izračunamo takole

$$\lambda_j = [\frac{Q}{Q-1}(\lambda_{j,Z} - \frac{1}{Q})]^2 \quad . \qquad j = 1,2,\ldots,X \tag{14}$$

Na podlagi X novih inercij PGK lahko definiramo množico nestandardiziranih PGK, izraženih z matriko Υ_F, osnova zanjo pa so standardizirane PGK iz matrike Γ_Z

$$\Upsilon_F = \Gamma_Z \frac{1}{Q-1}(QD_{\lambda,Z} - I) \quad . \tag{15}$$

Dekompozicija skupne inercije in matrika kumulativnih deležev inercij profilov so definirane na enak način kot v bivariatni KA. Zaradi simetričnosti Burtove kontingenčne tabele med profili-vrsticami in profili-stolpci ni nobenih razlik.

2. Grafična predstavitev položaja profilov

Profili so geometrijsko predstavljeni kot točke v večrazsežnem prostoru, z vrednostmi PGK kot koordinatami. Glede na to, da je razsežnost prostora profilov lahko relativno visoka, smo soočeni z resnim problemom praktičnega grafičnega prikaza položaja profilov. Ta problem se običajno reši tako, da se profile-točke prvotnega prostora profilov projecira na podprostor majhnih dimenzij, ki ga je mogoče prikazati v grafikonu. V skladu z naravo KA prvih nekaj PGK običajno predstavlja večino variabilnosti prvotne matrike podatkov. Položaj projekcij profilov-točk na podprostoru majhnih dimenzij je v takšnem primeru dobra aproksimacija stvarnega položaja profilov v prvotnem večrazsežnem prostoru.

2.1 Dvorazsežni razsevni grafikon

Tradicionalno se položaj profilov (pravzaprav njihovih projekcij na dvorazsežni podprostor) prikazuje s točkami v dvorazsežnem razsevnem grafikonu (slika 9). Položaj teh točk je določen z vrednostmi prvih dveh PGK . Če je delež inercij prvih dveh PGK v skupni inerciji relativno visok, potem je položaj večine profilov dobro predstavljen z njihovimi projekcijami na dvorazsežnem podprostoru. Pogosto pa se zgodi, da je ta delež relativno nizek. Tedaj je položaj večine profilov slabo predstavljen z njihovimi projekcijami, posledica tega pa so lahko nepopolni sklepi o proučevanem pojavu.

Osnovne značilnosti dvorazsežnega razsevnega grafikona so:

- položaj projekcij vseh profilov-točk je jasno prikazan v eni sami sliki, risanje je enostavno, mogoče je hkrati prikazati položaj velikega števila profilov-točk,
- ocena razdalj med točkami je enostavna in natančna
- za izdelavo tega tipa grafikona potrebujemo zgolj splošno razširjeno programsko in strojno opremo
- delež inercije pojasnjen s prvima dvema PGK je pogosto relativno nizek, posledica tega pa so lahko nepopolni sklepi.

2.2 Trirazsežni razsevni grafikon

Projekcije profilov-točk prvotnega večrazsežnega prostora na optimalni trirazsežni podprostor lahko prikažemo s točkami v trirazsežnem razsevnem grafikonu (slika 10). Položaj teh točk je določen z vrednostmi prvih treh PGK . Z vključitvijo tretje PGK se je povečal delež inercij PGK v skupni inerciji, s tem pa tudi kvaliteta aproksimacije položaja profilov glede na dvorazsežni razsevni grafikon.

Osnovne značilnosti trirazsežnega razsevnega grafikona so:

- položaj projekcij profilov-točk je zadosti dobro predstavljen, če je prikazan z nekaj slikami iz različnih zornih kotov in če število profilov ni preveliko,
- ocena razdalj med točkami je zahtevna in zgolj približna, saj je vtis tretje dimenzije dosežen s pomočjo perspektive,
- za izdelavo tega tipa grafikona je potrebna zgolj splošno razširjena programska in strojna oprema, priporočljiva pa je uporaba barvne grafike,
- uporaba zmogljive grafične postaje in ustrezne programske opreme omogoča tudi zvezno rotiranje množice profilov-točk v trirazsežnem razsevnem grafikonu. Tako je znatno izboljšana predstavitev položaja profilov-točk, saj je mnogo lažje oceniti razdalje med njimi,
- delež inercije pojasnjen s prvimi tremi PGK je včasih še vedno relativno nizek in zato je položaj vsaj nekaterih profilov slabo predstavljen; posledica tega so lahko nepopolni sklepi.

2.3 Andrewsov grafikon

V zadnjih tridesetih letih so bile razvite številne metode, ki omogočajo grafično predstavitev podatkov multivariatne narave (točk večrazsežnega prostora) v ravnini npr.: stolpičasti diagrami, zvezde, glifi, obrazi Chernoffa, Andrewsove krivulje itd. (glej npr. S.H.C. du Toit et al., 1986, str. 54-72). Osnovni cilj vsake izmed teh metod je prikazati razlike med multivariatnimi observacijami. Empirične raziskave so pokazale (P.C.C. Wang, 1978, str. 123-141), da so za takšno predstavitev običajno najprimernejše Andrewsove krivulje. Kot bomo v kratkem pokazali, pa so Andrewsove krivulje še posebej primerne za KA. Izbrali smo jih za prikaz položaja profilov v primerih, ko je potrebno upoštevati več kot tri PGK .

D.F. Andrews (1972) je predlagal uporabo trigonometričnih funkcij za grafično predstavitev podatkov multivariatne narave v ravnini. Andrewsova funkcija

$$f_{\mathbf{x}}(t) = x_1/\sqrt{2} + x_2\sin t + x_3\cos t + x_4\sin 2t + x_5\cos 2t + \dots , \qquad (16)$$

je definirana za vrednosti spremenljivke t na intervalu $-\pi<t<\pi$. Vektor opazovanih vrednosti $\mathbf{x} = (x_1,x_2,\dots,x_K)$ je grafično predstavljen s krivuljo v ravnini. Vzemimo, da je neka množica K-variatnih vektorjev opazovanih vrednosti predstavljena z množico krivulj vrisanih v isti grafikon. V tem primeru lahko na podlagi primerjave oblike Andrewsovih krivulj primerjamo položaj množice vektorjev opazovanih vrednosti oz. enot, ki jih ti vektorji predstavljajo. Izkaže se, da bodo točkam, katerih oddaljenost v K-razsežnem vektorskem prostoru je relativno majhna, ustrezale podobne Andrewsove krivulje in obratno. V primeru večjega števila enot (več kot 10 - 15) je težko ločiti posamezne krivulje v Andrewsovem grafikonu in primerjati njihovo obliko. Zato je potrebno narisati vsako krivuljo v poseben grafikon in jih nato ponovno primerjati.

Andrewsova funkcija $f_{\mathbf{x}}(t)$ ima nekatere lastnosti, zaradi katerih je zelo primerna za grafično predstavitev podatkov multivariatne narave. Med njimi sta najpomembnejši naslednji:

- funkcija $f_{\mathbf{x}}(t)$ ohranja povprečja. Če je $\bar{\mathbf{x}}$ povprečni vektor (centroid) množice vektorjev opazovanih vrednosti $\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_n$, potem je funkcija $f_{\bar{\mathbf{x}}}(t)$, ki ustreza povprečnemu vektorju $\bar{\mathbf{x}}$, enaka povprečju funkcij $f_{\mathbf{x}_1}(t)$, $f_{\mathbf{x}_2}(t),\dots,$ $f_{\mathbf{x}_n}(t)$, ki ustrezajo posameznim vektorjem opazovanih vrednosti,
- funkcija $f_{\mathbf{x}}(t)$ ohranja razdalje. Če vzamemo kot mero oddaljenosti kvadrat razdalje med funkcijama $f_{\mathbf{x}}(t)$ in $f_{\mathbf{y}}(t)$

$$\| f_{\mathbf{x}}(t) - f_{\mathbf{y}}(t) \|_{L_2} = \int_{-\pi}^{\pi} [f_{\mathbf{x}}(t) - f_{\mathbf{y}}(t)]^2 dt \ , \qquad (17)$$

ugotovimo, da je zanjo značilno, da se njen kvadratni koren dobro ujema s človekovim vizualnim zaznavanjem razdalj med točkami prostora. Gornja mera oddaljenosti je namreč proporcionalna kvadratu evklidske razdalje med odgovarjajočimi točkami, kajti

$$\| f_x(t) - f_y(t) \|_{L_2} = \pi \| x - y \|^2 = \pi \sum_{k=1}^{K} (x_k - y_k)^2 \quad . \qquad (18)$$

Na podlagi omenjenih lastnosti sklepamo, da so Andrewsove krivulje dobra osnova za razvrščanje vektorjev opazovanih vrednosti v homogene skupine in za primerjavo položaja posameznih vektorjev opazovanih vrednosti s povprečnim vektorjem (centroidom).

Za Andrewsovo funkcijo je značilno, da imajo spremenljivke v njenih prvih členih večjo težo pri oblikovanju Andrewsove krivulje kot pri kasnejših. Zato je iz vsebinskih razlogov priporočljivo razvrstiti pomembnejše spremenljivke v prve člene Andrewsove funkcije. Pomembnost spremenljivk je v splošnem stvar subjektivne presoje, kar ima lahko tudi neugodne posledice z vidika objektivnosti analize. V KA te nevarnosti ni, saj se tu v grafičnih prikazih v vlogi spremenljivk pojavijo PGK , katerih vrstni red je objektivno določen. Prva PGK ima vlogo prve spremenljivke, druga PGK vlogo druge spremenljivke itd. v enačbi Andrewsove funkcije.

Oblika Andrewsovih krivulj je odvisna od:

- števila opazovanih vrednosti (števila koordinat) vektorja opazovanih vrednosti
- absolutne vrednosti opazovanih vrednosti (koordinat)
- predznaka opazovanih vrednosti (koordinat).

Za pravilno razlago in primerjavo oblike Andrewsovih krivulj je bistveno, da se zavedamo vloge vsakega posameznega člena v Andrewsovi funkciji. Prvi člen funkcije (16) določa položaj horizontalne osi Andrewsove krivulje. Le-ta bo pretežno potekala nad abscisno osjo, če je prvi koeficient funkcije (prva koordinata) x_{i1} pozitiven in pod njo, če je negativen. Velikost absolutne vrednosti koeficienta x_{i1} določa razdaljo med abscisno osjo in osjo Andrewsove krivulje.

Sodi členi Andrewsove funkcije imajo obliko $x_{2j}\sin jt$ (j=1,2,...) in določajo sinusoide, s koeficienti x_{2j} kot amplitudami. Sprememba predznaka koeficienta x_{2j} pomeni zrcaljenje sinusoide glede na abscisno os. Sinusoida seka abscisno os za vsak večkratnik števila π/j .

Lihi členi Andrewsove funkcije (izvzemši prvega), ki imajo obliko $x_{2j+1}\cos jt$ (j=1,2,...) prav tako definirajo sinusoide, ki pa so premaknjene proti levi za $\pi/2j$. Ta tip sinusoid seka abscisno os za vsak lihi večkratnik števila $\pi/2j$.

Oblika Andrewsove krivulje je seveda rezultat skupnega učinka vseh členov. Če je Andrewsova funkcija sestavljena iz številnih členov je oblika krivulje relativno zapletena.

Prikažimo gornje značilnosti Andrewsovih krivulj na treh primerih vektorjev opazovanih vrednosti:

vektor	Andrewsova funkcija
(1,1,1,1,1)	$1/\sqrt{2} + \sin t + \cos t + \sin 2t + \cos 2t$
(-10,-8,5,3,-2,1)	$-10/\sqrt{2} - 8\sin t + 5\cos t + 3\sin 2t - 2\cos 2t + \sin 3t$
(8,-2,4,20,-3,-1)	$8/\sqrt{2} - 2\sin t + 4\cos t + 20\sin 2t - 3\cos 2t - \sin 3t$

Ker so koordinate prvega vektorja medsebojno enake imajo vsi členi Andrewsove funkcije pomemben vpliv na obliko Andrewsove krivulje. V sliki 3 krivulje z oznakami 1 do 5 predstavljajo ustrezne člene Andrewsove funkcije, celotna funkcija pa je prikazana z neprekinjeno črto. Prvi členi močneje vplivajo na obliko Andrewsove krivulje kot zadnji, kar je v skladu z naravo Andrewsove funkcije. To potrjuje tudi slika 4 , kjer smo prikazali Andrewsove krivulje za prvi člen Andrewsove funkcije, za prva dva člena, za prve tri člene itd.. Očitno je, da vsak dodani člen vpliva na obliko Andrewsove krivulje, toda intenziteta spreminjanja oblike krivulj se z vsakim naslednjim členom zmanjšuje.

Absolutne vrednosti koordinat drugega vektorja tvorijo padajoče zaporedje. Posledica tega je, da je oblika Andrewsove krivulje skoraj v celoti določena z nekaj prvimi členi (slika 5 in 6). Takšni primeri so v KA zelo pogosti.

Za tretji vektor je značilno, da je absolutna vrednost četrte koordinate mnogo višja od ostalih. Zato je v tem primeru oblika Andrewsove krivulje predvsem izraz vpliva četrtega člena Andrewsove funkcije (slika 7 in 8).

Povzemimo osnovne značilnosti Andrewsovega grafikona:

- Andrewsove krivulje omogočajo popolno predstavitev položaja profilov,
- ocena podobnosti Andrewsovih krivulj je do neke mere subjektivna, kljub naštetim ugodnim lastnostim Andrewsove funkcije,
- pri kompleksnejših primerih analiz moramo narisati veliko število grafikonov.

3. Primer multiple korespondenčne analize

Prikažimo uporabo omenjenih treh tipov grafikonov. Naše izhodišče bodo podatki zbrani v okviru socialno-ekonomske raziskave "Pogoji življenja in težnje Francozov" (L. Lebart in Y. Houzel van Effentere, 1980). Podatki se nanašajo na reprezentativni delni vzorec 1000 anketirancev in na izbranih sedem spremenljivk (Q=7) s skupno 25 vrednostmi (J=25):

- spol (SPOM - moški ; SPOŽ - ženski)
- najvišja stopnja dosežene šolske izobrazbe (IZO1 - brez dokončane osnovne šole ; IZO2 - dokončana osnovna šola ; IZO3 - nedokončana srednja šola ; IZO4 - dokončana srednja šola ; IZO5 - vsaj nedokončana fakulteta)
- stanovanjske razmere (STN1 - hipotečni dolžnik ; STN2 - lastnik ; STN3 - najemnik ; STN4 - brezplačno bivanje)
- lastništvo delnic ali obveznic (DEL1 - da ; DEL2 - ne)
- lastništvo nepremičnin, brez stanovanja, v katerem biva (NEP1 - da ; NEP2 - ne)
- starost (STR1 - 19 do 24 let ; STR2 - 25 do 34 let ; STR3 - 35 do 49 let ; STR4 - 50 do 64 let ; STR5 - 65 let in več)
- velikost kraja bivanja (število prebivalcev) (VEL1 - pod 2.000 ; VEL2 - 2.000 do pod 20.000 ; VEL3 - 20.000 do pod 100.000 ; VEL4 - 100.000 do pod 500.000 ; VEL5 - nad 500.000).

Izhodišče analize je Burtova kontingenčna tabela (12), prikazana v tabeli 1 . Na podlagi Burtove tabele B lahko formiramo matriko profilov-vrstic $\frac{1}{Q}\,D_f^{-1}B$ (13) . S spektralno dekompozicijo te matrike pa dobimo diagonalno matriko inercij PGK $D_{\lambda,Z}$ in matriko standardiziranih koordinat PGK Γ_Z . Število netrivialnih inercij kot tudi število PGK je $J-Q$ = 18 . Vrednosti inercij PGK $\lambda_{Z,j}$, odstotnih deležev inercij PGK v skupni inerciji $\lambda_{Z,j}\%$ in kumulativnih odstotnih deležev inercij PGK v skupni inerciji $\mathrm{kum}\lambda_{Z,j}\%$ so prikazane v tabeli 2 .

Tabela 1: Burtova kontingenčna tabela (Vir: L.Lebart et al., stran 104).

	SPO M	SPO F	IZO 1	IZO 2	IZO 3	IZO 4	IZO 5	STN 1	STN 2	STN 3	STN 4	DEL 1	DEL 2	NEP 1	NEP 2	STR 1	STR 2	STR 3	STR 4	STR 5	VEL 1	VEL 2	VEL 3	VEL 4	VEL 5
SPOM	469	0	102	164	69	65	69	62	151	224	32	54	415	35	434	18	111	169	84	87	42	40	81	161	145
SPOŽ	0	531	94	163	91	102	81	58	142	302	29	67	464	47	484	22	136	187	104	82	41	47	94	168	181
IZO1	102	94	196	0	0	0	0	17	59	105	15	12	184	11	185	9	27	54	47	59	19	17	35	70	55
IZO2	164	163	0	327	0	0	0	45	116	151	15	26	301	21	306	0	55	125	84	63	43	34	67	110	73
IZO3	69	91	0	0	160	0	0	16	37	98	9	19	141	14	146	14	47	60	27	12	12	13	34	52	49
IZO4	65	102	0	0	0	167	0	27	39	87	14	27	140	19	148	16	62	63	16	10	8	15	21	57	66
IZO5	69	81	0	0	0	0	150	15	42	85	8	37	113	17	133	1	56	54	14	25	1	8	18	40	83
STN1	62	58	17	45	16	27	15	120	0	0	0	11	109	7	113	3	23	68	20	6	7	20	27	48	18
STN2	151	142	59	116	37	39	42	0	293	0	0	60	233	48	245	9	24	91	80	89	59	33	62	72	67
STN3	224	302	105	151	98	87	85	0	0	526	0	45	481	23	503	22	180	182	79	63	11	29	80	191	215
STN4	32	29	15	15	9	14	8	0	0	0	61	5	56	4	57	6	20	15	9	11	6	5	6	18	26
DEL1	54	67	12	26	19	27	37	11	60	45	5	121	0	39	82	2	18	35	27	39	4	9	22	36	50
DEL2	415	464	184	301	141	140	113	109	233	481	56	0	879	43	836	38	229	321	161	130	79	78	153	293	276
NEP1	35	47	11	21	14	19	17	7	48	23	4	39	43	82	0	3	11	27	20	21	7	12	12	25	26
NEP2	434	484	185	306	146	148	133	113	245	503	57	82	836	0	918	37	236	329	168	148	76	75	163	304	300
STR1	18	22	9	0	14	16	1	3	9	22	6	2	38	3	37	40	0	0	0	0	2	4	8	15	11
STR2	111	136	27	55	47	62	56	23	24	180	20	18	229	11	236	0	247	0	0	0	13	16	34	102	82
STR3	169	187	54	125	60	63	54	68	91	182	15	35	321	27	329	0	0	356	0	0	31	34	71	100	120
STR4	84	104	47	84	27	16	14	20	80	79	9	27	161	20	168	0	0	0	188	0	25	20	26	60	57
STR5	87	82	59	63	12	10	25	6	89	63	11	39	130	21	148	0	0	0	0	169	12	13	36	52	56
VEL1	42	41	19	43	12	8	1	7	59	11	6	4	79	7	76	2	13	31	25	12	83	0	0	0	0
VEL2	40	47	17	34	13	15	8	20	33	29	5	9	78	12	75	4	16	34	20	13	0	87	0	0	0
VEL3	81	94	35	67	34	21	18	27	62	80	6	22	153	12	163	8	34	71	26	36	0	0	175	0	0
VEL4	161	168	70	110	52	57	40	48	72	191	18	36	293	25	304	15	102	100	60	52	0	0	0	329	0
VEL5	145	181	55	73	49	66	83	18	67	215	26	50	276	26	300	11	82	120	57	56	0	0	0	0	326

Tabela 2: Prikaz inercij PGK $\lambda_{z,j}$, odstotnih deležev inercij PGK v skupni inerciji $\lambda_{z,j}\%$ (prikazani so tudi v histogramu) in kumulativnih odstotnih deležev inercij PGK v skupni inerciji $kum\lambda_{z,j}\%$ (Vir: Tabela 1)

$\lambda_{z,j}$	$\lambda_{z,j}\%$	$kum\lambda_{z,j}\%$	2 4 6 8 10
0.2500	9.72	9.72	••••••••••••••••••••••••
0.2184	8.49	18.21	•••••••••••••••••••••
0.1847	7.18	25.40	••••••••••••••••••
0.1746	6.79	32.19	•••••••••••••••••
0.1610	6.26	38.45	••••••••••••••••
0.1580	6.15	44.59	•••••••••••••••
0.1542	6.00	50.59	•••••••••••••••
0.1457	5.67	56.26	••••••••••••••
0.1409	5.48	61.74	••••••••••••••
0.1344	5.23	66.96	•••••••••••••
0.1290	5.02	71.98	•••••••••••••
0.1253	4.87	76.85	••••••••••••
0.1223	4.76	81.61	••••••••••••
0.1177	4.58	86.18	•••••••••••
0.1060	4.12	90.31	••••••••••
0.0957	3.72	94.03	•••••••••
0.0869	3.38	97.41	••••••••
0.0666	2.59	100.00	••••••
2.5714			

$$\frac{1}{Q} = \frac{1}{7} = 0.1429$$

Deleži inercij nekaj prvih PGK so v primeru Z-analize praviloma relativno nizki. Kot pa smo že omenili v razdelku 1.2 , Benzécri (1979) meni, da so analitično pomembne le tiste PGK, katerih inercije $\lambda_{Z,j}$ so večje od $\frac{1}{Q}$, torej v našem primeru od $\frac{1}{7} = 0.1429$. Izračunajmo ponovno inercije PGK, upoštevaje Benzécrijev izraz (14), in prikažimo oba tipa inercij PGK $\lambda_{Z,j}$ in λ_j , odstotne deleže inercij PGK v skupni inerciji $\lambda_j\%$ in kumulativne odstotne deleže inercij PGK v skupni inerciji kum$\lambda_j\%$ v tabeli 3 .

Tabela 3: Prikaz inercij PGK $\lambda_{Z,j}$ in λ_j , odstotnih deležev inercij PGK v skupni inerciji $\lambda_j\%$ (prikazani so tudi v histogramu) in kumulativnih odstotnih deležev inercij PGK v skupni inerciji kum$\lambda_j\%$ (Vir: Tabela 2)

$\lambda_{Z,j}$	λ_j	$\lambda_j\%$	kum$\lambda_j\%$	10 20 30 40 50
0.2500	0.015631	55.65	55.65	••••••••••••••••••••••••••••
0.2187	0.007760	27.63	83.27	••••••••••••••
0.1847	0.002378	8.47	91.74	••••
0.1746	0.001372	4.88	96.62	••
0.1610	0.000450	1.60	98.22	•
0.1580	0.000313	1.12	99.34	•
0.1542	0.000174	0.62	99.96	
0.1457	0.000011	0.04	100.00	
	0.028089			

Kot lahko vidimo iz tabele 3 , se je število inercij PGK sedaj skrčilo od 18 na vsega 8 , pri čemer pa že s prvimi štirimi PGK predstavimo kar 96.62% skupne inercije. Na podlagi slednjega bi lahko sklepali, da prikaz položaja profilov na podlagi prvih treh ali štirih PGK nudi dobro osnovo za analizo proučevanega pojava. Menimo, da je takšno sklepanje pomanjkljivo. Na podlagi izkušenj ugotavljamo, da je lahko osnova za popolno analizo le takšen prikaz, v katerem so vsi profili dobro predstavljeni. Za to pa je nujno potrebno pri nekaterih profilih upoštevati večje število PGK . K temu problemu se bomo vrnili kasneje.

Izračunali smo že matriko standardiziranih PGK Γ_Z . Sedaj moramo transformirati prvih osem stolpcev matrike Γ_Z v ustrezne stolpce vrednosti PGK matrike Υ_F (15) (v "nestandardizirane" PGK). Matrika vrednosti PGK Υ_F je prikazana v tabeli 4 .

Tabela 4: Matrika koordinat PGK T_F (Vir: Tabela 3).

		1.PGK	2.PGK	3.PGK	4.PGK	5.PGK	6.PGK	7.PGK	8.PGK
1	SPOM	0.0354	-0.0308	-0.0163	-0.0077	0.0196	0.0144	0.0031	0.0041
2	SPOŽ	-0.0313	0.0272	0.0144	0.0068	-0.0173	-0.0127	-0.0028	-0.0036
3	IZO1	0.0662	-0.0643	-0.1156	0.0348	0.0387	0.0046	-0.0043	-0.0072
4	IZO2	0.1084	-0.1130	0.0132	-0.0449	-0.0154	0.0058	-0.0040	0.0014
5	IZO3	-0.0994	-0.0002	0.0340	0.0523	-0.0154	-0.0581	-0.0069	0.0041
6	IZO4	-0.1365	0.0860	0.0912	0.0763	-0.0023	0.0274	0.0069	0.0008
7	IZO5	-0.0649	0.2350	-0.0157	-0.0883	0.0020	0.0128	0.0140	0.0012
8	STN1	-0.0276	-0.1335	0.1796	-0.0326	0.0444	0.0272	-0.0043	-0.0009
9	STN2	0.2891	0.0127	-0.0071	0.0124	-0.0098	-0.0049	0.0037	0.0019
10	STN3	-0.1450	0.0212	-0.0283	-0.0148	-0.0058	-0.0109	-0.0068	-0.0014
11	STN4	-0.0839	0.0189	-0.0753	0.1323	0.0098	0.0640	0.0491	0.0046
12	DEL1	0.2452	0.3338	0.0545	-0.0019	0.0074	-0.0020	-0.0140	0.0021
13	DEL2	-0.0338	-0.0460	-0.0075	0.0003	-0.0010	0.0003	0.0019	-0.0003
14	NEP1	0.3442	0.3153	0.1066	0.0503	-0.0077	0.0043	-0.0207	0.0009
15	NEP2	-0.0307	-0.0282	-0.0095	-0.0045	0.0007	-0.0004	0.0018	-0.0001
16	STR1	-0.1802	0.0028	0.0238	0.2997	0.0437	-0.0406	0.0156	0.0014
17	STR2	-0.2308	0.0497	-0.0178	-0.0051	-0.0180	0.0167	-0.0139	0.0047
18	STR3	-0.0184	-0.0384	0.0710	-0.0347	0.0125	-0.0085	0.0191	-0.0002
19	STR4	0.1624	-0.0646	-0.0165	0.0142	-0.0532	0.0036	-0.0171	-0.0058
20	STR5	0.2379	0.0795	-0.1109	-0.0061	0.0487	-0.0008	-0.0047	-0.0003
21	VEL1	0.3198	-0.1962	-0.0247	0.0395	-0.0798	0.0096	0.0324	0.0089
22	VEL2	0.1220	-0.0638	0.1289	0.0307	0.0144	0.0325	0.0001	-0.0148
23	VEL3	0.0664	-0.0548	0.0263	-0.0183	0.0430	-0.0547	0.0030	0.0033

Vrednosti PGK bodo rabile kot koordinate projekcij profilov-točk na optimalni podprostor izbrane razsežnosti.

Posebej pa je pomembno, ali je položaj vsakega posameznega profila dobro predstavljen v optimalnem podprostoru. Zato moramo izračunati matriko kumulativnih deležev inercij profilov (11), prikazano v tabeli 5 . Glede na dosedanje izkušnje menimo, da mora za dobro aproksimacijo kumulativni delež inercije, ki je pojasnjen z izbranim številom PGK, presegati 90% skupne inercije profila.

Prikažimo sedaj položaj profilov na podlagi vrednosti prvih dveh PGK v dvorazsežnem razsevnem grafikonu (slika 9). Na podlagi položaja točk v grafikonu lahko oblikujemo naslednje potencialne skupine profilov-točk:

a) IZO4 - osebe z dokončano srednjo šolo
   STN3 - najemniki stanovanj
   STR1 - osebe, stare od 19-24 let
   STR2 - osebe, stare od 25-34 let

b) STN2 - lastniki stanovanj
   STR5 - osebe, starejše od 65 let

c) DEL1 - lastniki delnic ali obveznic
   NEP1 - lastniki nepremičnin

Tabela 5: Matrika kumulativnih deležev inercij profilov-točk, ki so pojasnjene s posamezno PGK v inercijah profilov-točk (Vir: Tabela 1 in 4).

		1.PGK	2.PGK	3.PGK	4.PGK	5.PGK	6.PGK	7.PGK	8.PGK
1	SPOM	0.3985	0.7002	0.7848	0.8037	0.9262	0.9916	0.9948	1.0000
2	SPOŽ	0.3985	0.7002	0.7848	0.8037	0.9262	0.9916	0.9948	1.0000
3	IZO1	0.1777	0.3454	0.8865	0.9355	0.9963	0.9972	0.9979	1.0000
4	IZO2	0.4352	0.9081	0.9146	0.9893	0.9981	0.9993	0.9999	1.0000
5	IZO3	0.5661	0.5661	0.6325	0.7893	0.8029	0.9963	0.9990	1.0000
6	IZO4	0.4550	0.6354	0.8383	0.9804	0.9805	0.9988	1.0000	1.0000
7	IZO5	0.0621	0.8761	0.8797	0.9946	0.9947	0.9971	1.0000	1.0000
8	STN1	0.0140	0.3402	0.9306	0.9500	0.9861	0.9997	1.0000	1.0000
9	STN2	0.9940	0.9959	0.9965	0.9984	0.9995	0.9998	1.0000	1.0000
10	STN3	0.9265	0.9463	0.9815	0.9912	0.9927	0.9979	0.9999	1.0000
11	STN4	0.1893	0.1989	0.3512	0.8220	0.8246	0.9348	0.9994	1.0000
12	DEL1	0.3439	0.9815	0.9985	0.9985	0.9988	0.9988	1.0000	1.0000
13	DEL2	0.3439	0.9815	0.9985	0.9985	0.9988	0.9988	1.0000	1.0000
14	NEP1	0.5099	0.9379	0.9869	0.9978	0.9981	0.9982	1.0000	1.0000
15	NEP2	0.5099	0.9379	0.9869	0.9978	0.9981	0.9982	1.0000	1.0000
16	STR1	0.2564	0.2565	0.2609	0.9700	0.9850	0.9981	1.0000	1.0000
17	STR2	0.9362	0.9795	0.9851	0.9856	0.9913	0.9962	0.9996	1.0000
18	STR3	0.0390	0.2093	0.7919	0.9311	0.9493	0.9577	1.0000	1.0000
19	STR4	0.7717	0.8936	0.9016	0.9075	0.9901	0.9905	0.9990	1.0000
20	STR5	0.7289	0.8103	0.9686	0.9691	0.9997	0.9997	1.0000	1.0000
21	VEL1	0.6794	0.9352	0.9392	0.9496	0.9919	0.9925	0.9995	1.0000
22	VEL2	0.3921	0.4991	0.9362	0.9610	0.9665	0.9943	0.9943	1.0000
23	VEL3	0.3315	0.5574	0.6093	0.6344	0.7732	0.9985	0.9992	1.0000
24	VEL4	0.6312	0.7426	0.7512	0.7670	0.7749	0.8240	0.9987	1.0000
25	VEL5	0.2856	0.9050	0.9532	0.9675	0.9753	0.9753	0.9996	1.0000
Skupaj		0.5565	0.8327	0.9174	0.9662	0.9822	0.9934	0.9996	1.0000

d) SPOM - moški
SPOŽ - ženske
DEL2 - osebe, ki niso lastniki niti delnic niti obveznic
NEP2 - osebe, ki niso lastniki nepremičnin
STR3 - osebe, stare od 35-49 let
VEL4 - osebe, ki živijo v mestih z med 100000 in 500000 prebivalci

Ko smo oblikovali gornje potencialne skupine profilov-točk (v sliki 9 so obkrožene s črtkano črto) nismo upoštevali deležev inercij profilov-točk pojasnjenih s prvima dvema PGK v posameznih inercijah profilov-točk. V skladu z vrednostmi kumulativnih deležev inercij v skupni inerciji posameznih profilov-točk (tabela 5) so nekateri profili slabo predstavljeni v dvorazsežnem razsevnem grafikonu, kajti njihovi deleži inercij so manjši od 90% . Zato menimo, da v tem primeru lahko brez zadržkov obravnavamo kot skupine le tiste profile iz potencialnih skupin, pri katerih omenjeni delež presega 90% (v sliki 9 so obkroženi z neprekinjeno črto).

V primeru, ko želimo predstaviti položaj profilov na podlagi vrednosti prvih treh PGK, je najprimerneje uporabiti trirazsežni razsevni grafikon (slika 10). Vpliv upoštevanja dodatne, v tem primeru tretje, PGK na že omenjene skupine profilov-točk je nasledji:

- Profil IZ04 je potrebno izločiti iz potencialne skupine "a" , kajti njegova vrednost tretje PGK se močno razlikuje od ustreznih vrednosti pri ostalih profilih;
- oba profila potecialne skupine "b" sta dobro predstavljena z vrednostmi prvih treh PGK, vendar pa se njuni vrednosti tretje PGK tako močno razlikujeta, da očitno ne oblikujeta skupine;
- Ker je bila skupina "c" dobro predstavljena že v dvorazsežnem razsevnem grafikonu, se relativni položaj njenih profilov-točk tudi v trirazsežnem razsevnem grafikonu ne more pomembneje spremeniti;
- Profil STR3 je potrebno izločiti iz potencialne skupine "d" , kajti njegova vrednost tretje PGK se močno razlikuje od ustreznih vrednosti pri ostalih profilih.

Ko smo oblikovali gornje potencialne skupine profilov-točk (v sliki 10 so obkrožene s črtkano črto) nismo upoštevali deležev inercij profilov-točk pojasnjenih s prvimi tremi PGK v posameznih inercijah profilov-točk. V našem primeru so nekateri profili slabo predstavljeni tudi v trirazsežnem razsevnem grafikonu. Zato lahko brez zadržkov obravnavamo kot skupine le tiste profile, ki so v sliki 10 obkroženi z neprekinjeno črto.

Izkazalo se je, da je kar 8 izmed skupno 14 profilov iz proučevanih potencialnih skupin slabo predstavljenih tudi na podlagi vrednosti treh PGK. Zato menimo, da je v našem primeru smiselno uporabiti za prikaz profilov Andrewsov grafikon, ki omogoča prikaz položaja profilov na podlagi vrednosti poljubnega števila PGK. Če položaj vseh 25 profilov prikažemo v enem samem Andrewsovem grafikonu, kar pomeni, da narišemo v isti sliki vseh 25 Andrewsovih krivulj, je zelo težko slediti poteku posameznih krivulj in primerjati njihove oblike. Bolje je narisati vsako krivuljo v poseben grafikon, nato pa paroma primerjati njihove oblike in postopoma oblikovati skupine. Če pri tem upoštevamo vrednosti vseh osmih PGK je položaj profilov tudi popolno predstavljen. V našem primeru smo oblikovali naslednje skupine profilov (slike 11 in 12):

1) STR2 - osebe, stare od 25-34 let
STN3 - najemniki

2) DEL1 - lastniki delnic ali obveznic
NEP1 - lastniki nepremičnin

3) STR1 - osebe, stare od 19-24 let
STN4 - osebe, ki bivajo brezplačno

4) SPOM - moški
SPOŽ - ženske
DEL2 - osebe, ki niso lastniki niti delnic niti obveznic
NEP2 - osebe, ki niso lastniki nepremičnin

5) IZO5 - osebe z vsaj nedokončano fakulteto
VEL5 - osebe, ki živijo v mestih z nad 500000 prebivalci

Preostali profili ne oblikujejo skupin (slika 13).

V zvezi z oblikovanjem skupin omenimo še naslednje:

- profila, ki predstavljata osebe obeh spolov, pripadata isti skupini. Njuni Andrewsovi krivulji se le zelo malo razlikujeta od ravne črte, ki predstavlja položaj centroida in ki leži na abscisni osi. To pomeni, da se anketiranci obeh spolov ne razlikujejo med seboj z vidika preostalih šestih spremenljivk. To potrjujejo tudi vrednosti χ^2-preizkusov, na podlagi katerih smo ugotavljali povezanost med spremenljivko spol in vsako izmed ostalih šestih spremenljivk, kajti le eden izmed šestih χ^2-preizkusov je odkril značilne razlike ob 5% tveganju.
- na podlagi dvo- oz. trirazsežnega razsevnega grafikona ni mogoče sklepati, da profila STR1 in STN4 oblikujeta skupino. Ker pri obeh profilih izrazito izstopa vrednost četrte PGK, njuni aproksimaciji v razsevnih grafikonih ne odražata stvarnega položaja profilov. Šele na podlagi Andrewsovega grafikona, ki predstavlja položaj profilov na podlagi vseh PGK, lahko realno ovrednotimo položlaj omenjenih dveh profilov in se prepričamo, da res oblikujeta skupino.

Kot kriterij za dobro aproksimacijo položaja profila smo izbrali 90% delež inercije profila, pojasnjen z vrednostmi PGK. Utemeljitev za izbor tega empiričnega kriterija izvira iz Andrewsovega grafikona. Prikažimo zato položaj enega in istega profila z množico Andrewsovih krivulj, takšnih, da prva predstavlja položaj profila le na podlagi prvega člena Andrewsove funkcije, druga na podlagi prvih dveh členov iste funkcije, itd.. Opazili bomo, da se oblike tistih krivulj, ki ustrezajo 90% in večjim deležem inercije profila, pojasnjene z vrednostmi PGK, le malo razlikujejo (tabela 5 in slika 14).

4. Zaključek

Ob koncu želimo ovrednotiti omenjene tri tipe grafikonov z vidika uporabe v KA. V primerih, ko že na podlagi vrednosti prvih dveh oz. treh PGK dosežemo dovolj dobro aproksimacijo položaja profilov, bomo zaradi njihove preprostosti dali prednost razsevnim grafikonom. Pogosto pa se zgodi, da vrednosti prvih treh PGK ne zadostujejo za dobro aproksimacijo položaja profilov. Tedaj smo prisiljeni uporabiti zahtevnejše grafične metode, s katerimi lahko prikažemo položaj profilov na podlagi večjega števila PGK. Menimo, da je v takšnih primerih položaj profilov še posebej dobro prikazan v Andrewsovem grafikonu.

5. Programska oprema

Vsi izračuni in grafikoni so izvedeni s sistemom SAS. Matrika standardiziranih PGK Γ_z in inercije PGK $\lambda_{j,z}$ so izračunane s proceduro CORRESP modula SAS/STAT, vsi nadaljni izračuni pa so izvedeni z matričnim jezikom SAS/IML. Grafikoni so narisani s procedurami PLOT in G3D modula SAS/GRAPH. V programih so pogosto uporabljene macro spremenljivke.

6. Literatura

Andrews D.F. 1972. "Plots of high dimensional data", Biometrics 28:125-136.

Benzécri J.P. 1979. "Sur le calcul des taux d'inertie dans l'analyse d'un questionnaire. Addendum et erratum à [BIN.MULT]". Cahiers de L'analyse des Données 4:377-378.

Benzécri J.-P et al. 1980. L'Analyse des Données, L'Analyse des Correspondances. Dunod, Paris.

Greenacre M.J. 1981. "Practical correspondence analysis." P.119-146, v Barnett V. (ed.): Interpreting Multivariate Data. Chichester, U.K., J.Wiley.

Greenacre M.J. 1984. Theory and Application of Correspondence Analysis. Academic Press.

Greenacre M.J., Hastie T. 1987. "The geometric interpretation of correspondence analysis." JASA 82:437-447.

Lebart L., Houzel van Effenterre Y. 1980. "Le Systeme d'Enquetêtes sur les Aspirations des Français: Une Breve Presentation." Consommation 1:3-25.

Lebart L., Morineau A., Warwick K.M. 1984. Multivariate Descriptive Statistical Analysis, Correspondance Analysis and Related Techniques for Large Matrices. J.Wiley.

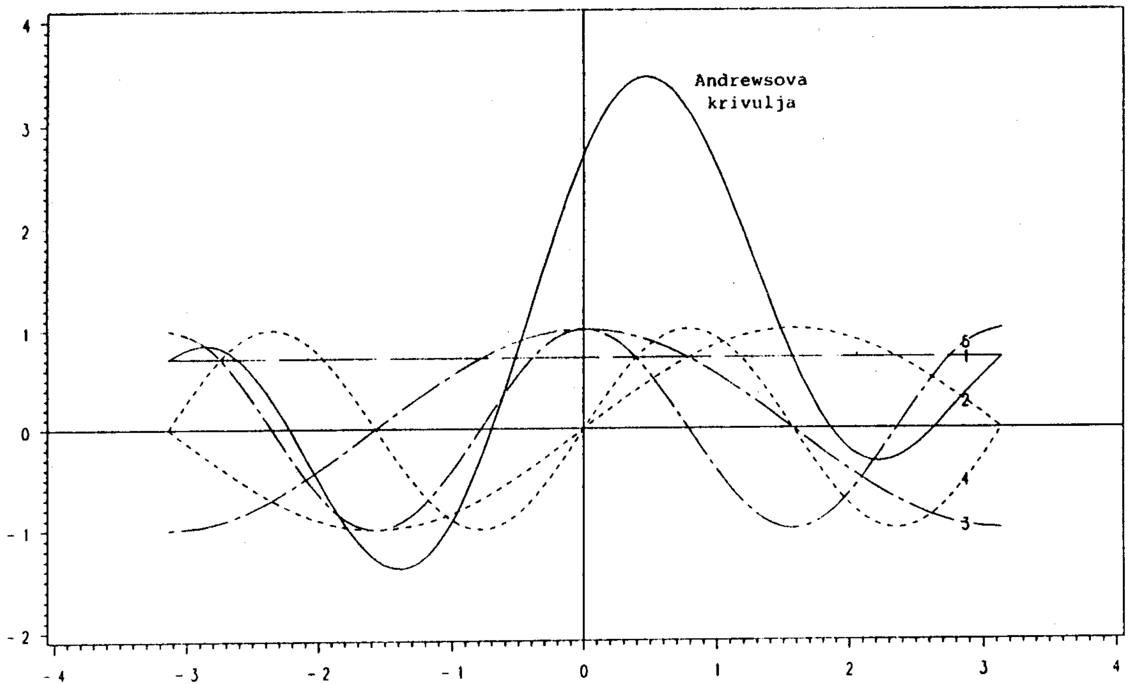
Pearson K. 1901. "On lines and planes of closest fit to systems of points in space." Philosophical Magazine 2:559-572.

Rovan J. 1991. Andrewsove krivulje v korespondenčni analizi (doktorsko delo, nepublicirano), Ljubljana.

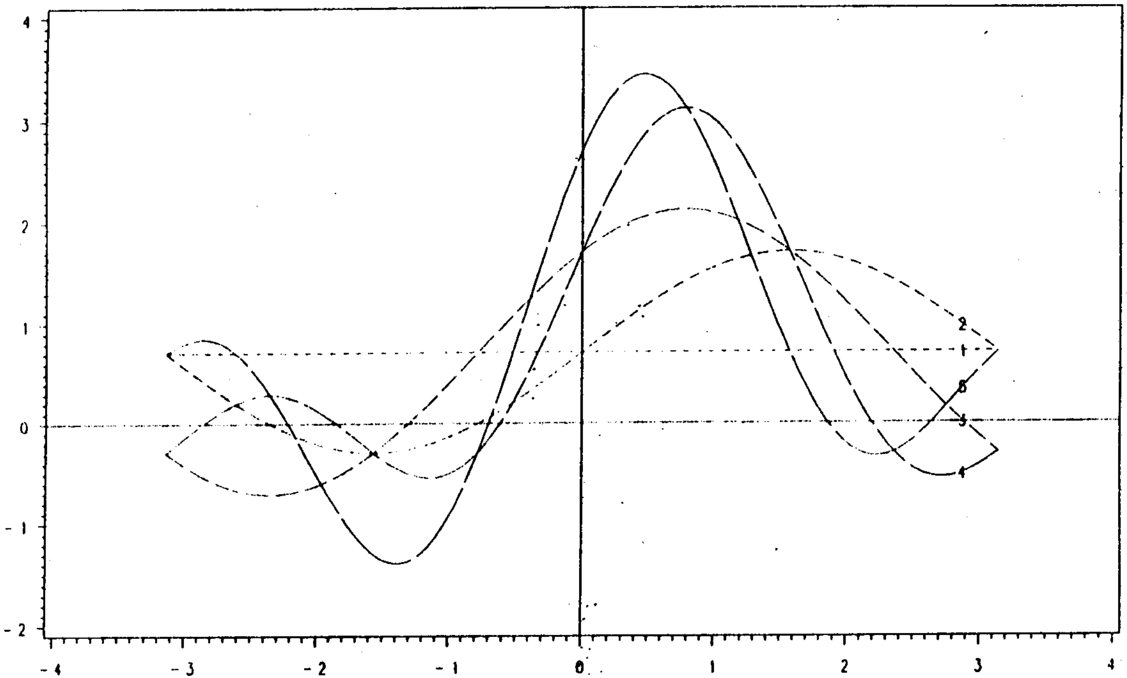
du Toit S.H.C., Steyn A.G.W., Stumpf R.H. 1986. Graphical Exploratory Data Analysis. Springer-Verlag.

Wang P.C.C.(ed.). 1978. Graphical Representation of Multivariate Data. Academic Press, New York.

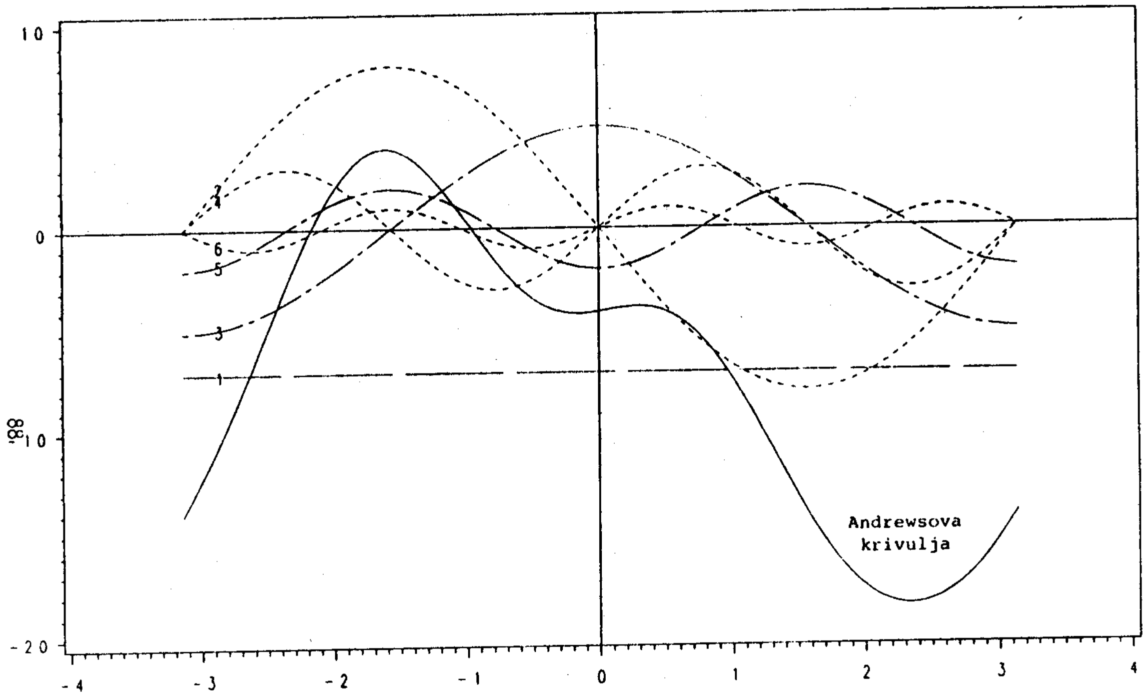
SAS, SAS/STAT, SAS/IML in SAS/GRAPH so zaščitene blagovne znamke SAS Institute Inc., Cary, NC, USA



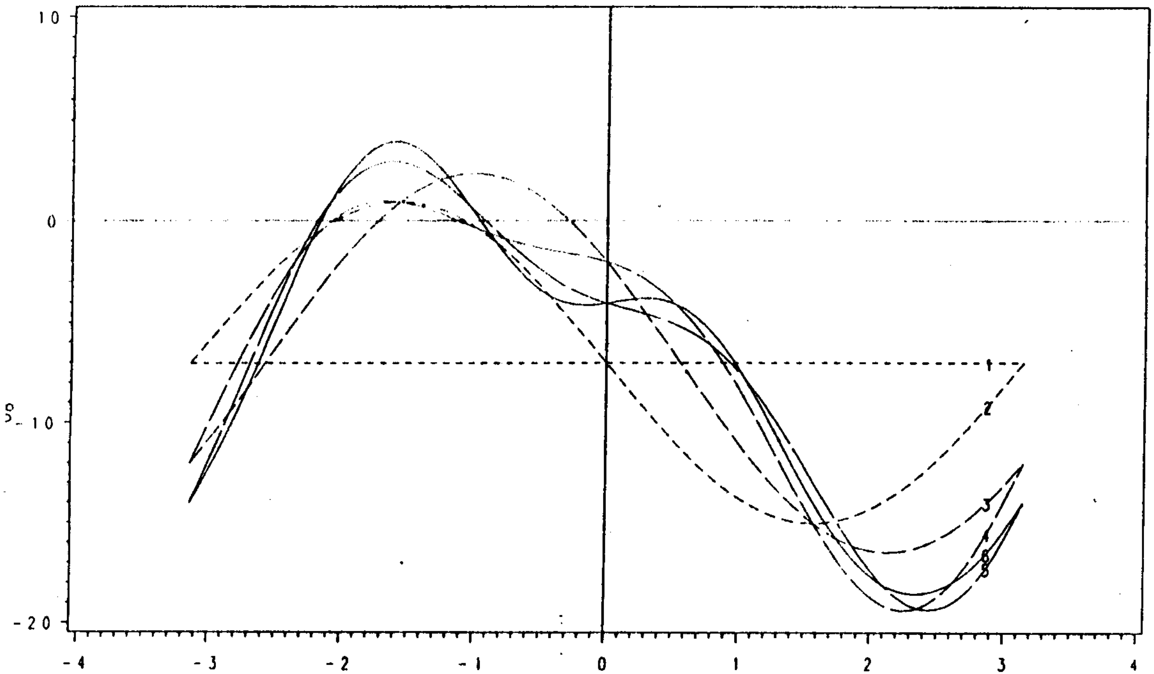
Slika 3: Prikaz krivulj, ki ustrezajo posameznim členom Andrewsove funkcije in celotni funkciji: $1/\sqrt{2}$ + sint + cost + sin2t + cos2t



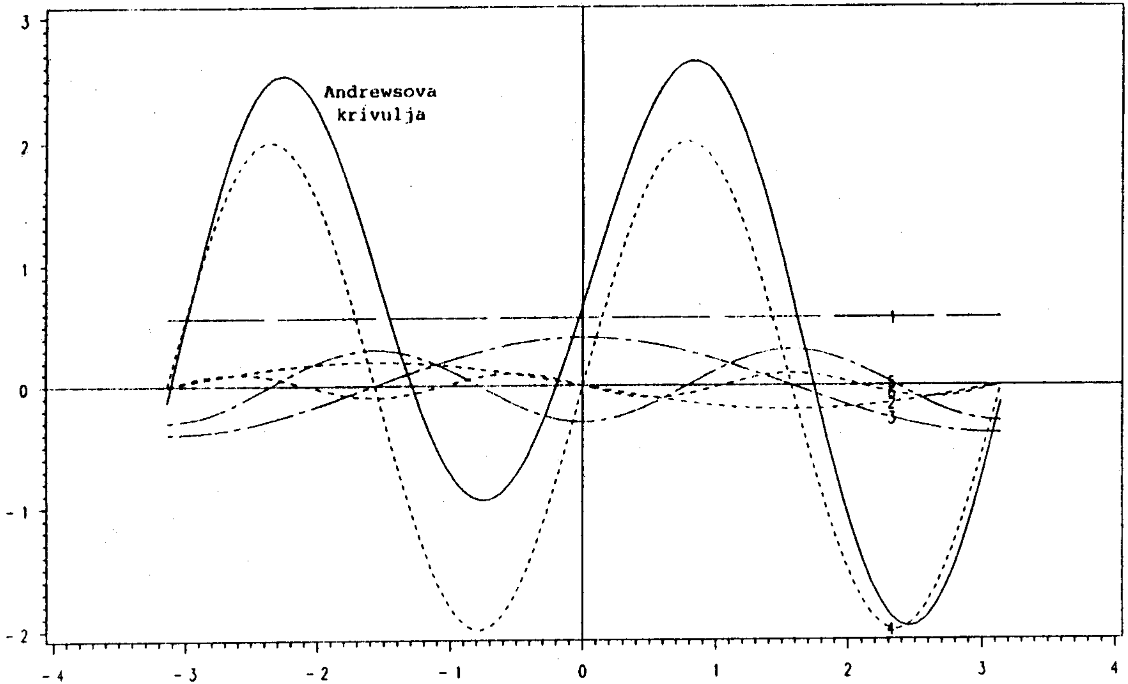
Slika 4: Prikaz Andrewsovih krivulj, ki ustrezajo prvemu (1), prvima dvema (2), prvim trem (3) itd. členom funkcije $1/\sqrt{2} + \sin t + \cos t + \sin 2t + \cos 2t$



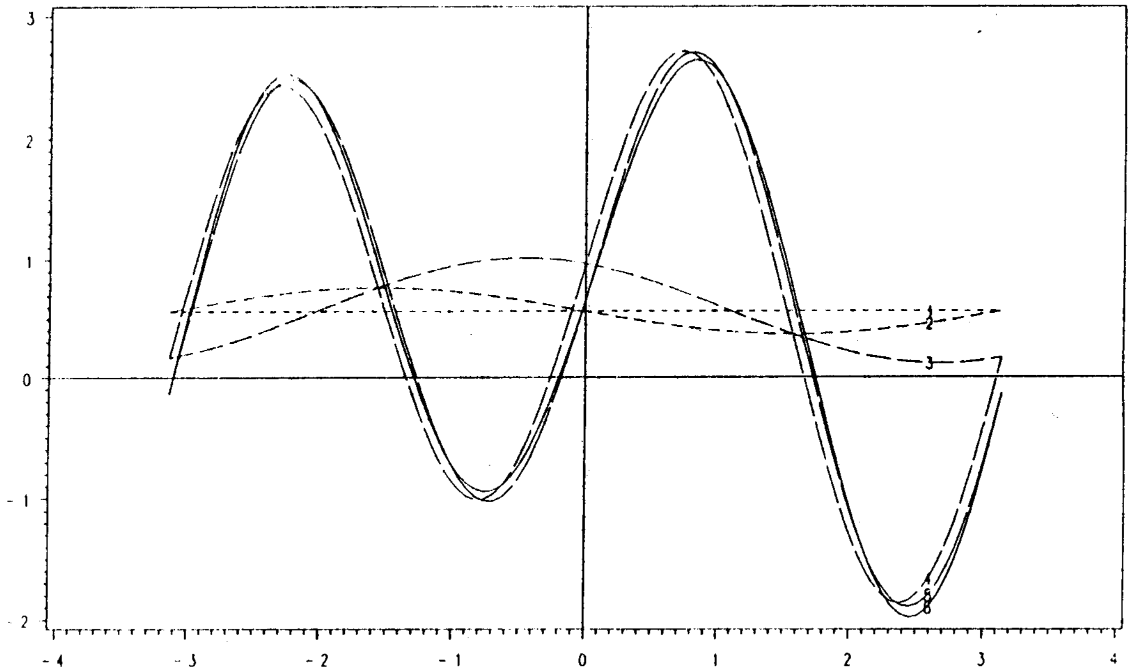
Slika 5: Prikaz krivulj, ki ustrezajo posameznim členom Andrevsove funkcije in celotni funkciji: $-10\sqrt{2}-8\sin t + 5\cos t + 3\sin 2t + 2\cos 2t + \sin 3t$



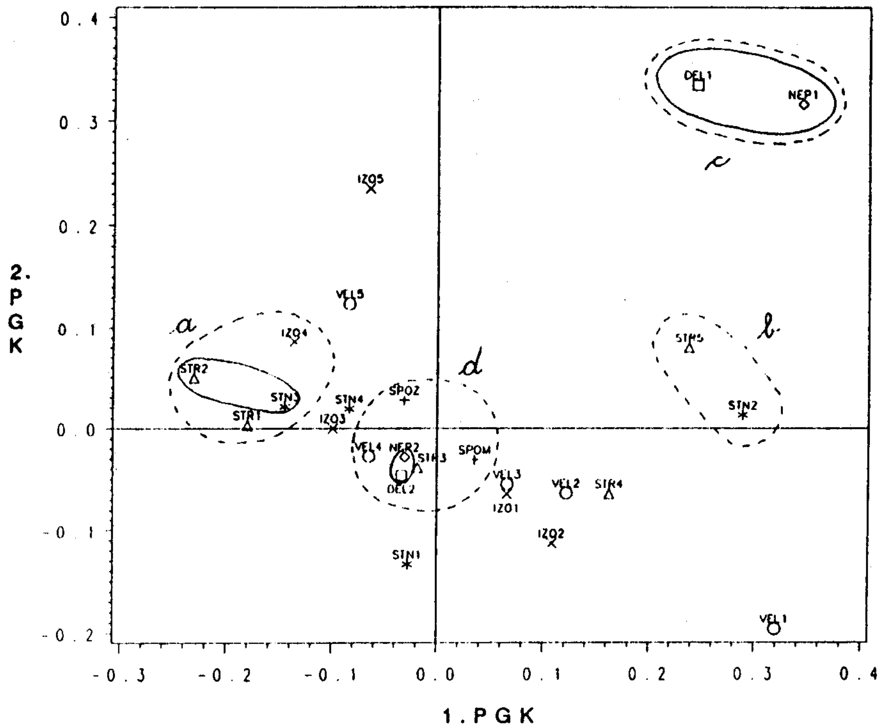
Slika 6: Prikaz Andrewsovih krivulj, ki ustrezajo prvemu (1), prvima dvema (2), prvim trem (3) itd. členom funkcije $-10/\sqrt{2} - 8\sin t + 5\cos t + 3\sin 2t + 2\cos 2t + \sin 3t$



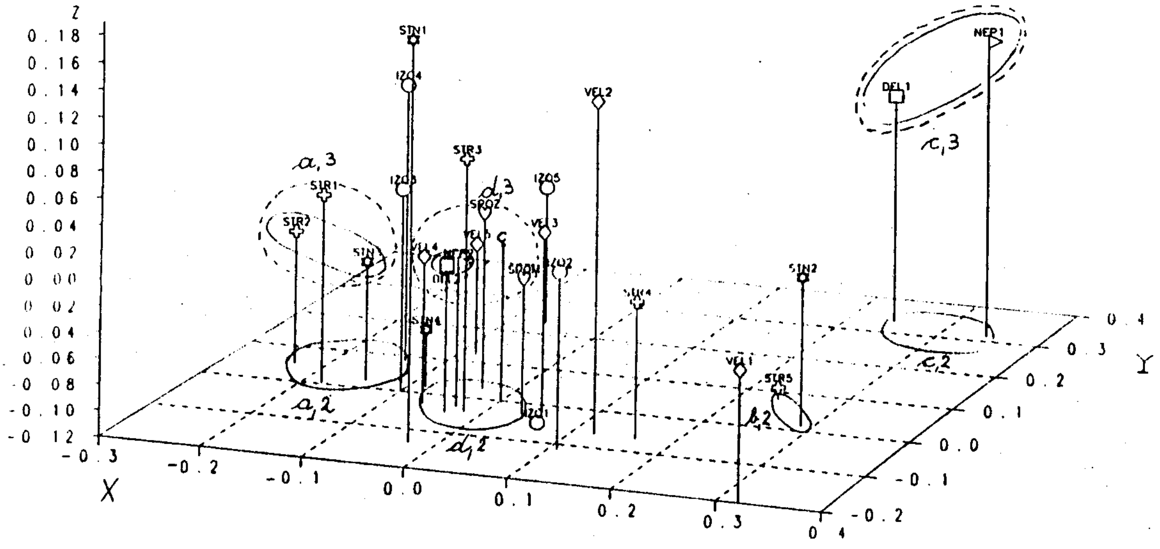
Slika 7: Prikaz krivulj, ki ustrezajo posameznim členom Andrevsove funkcije in celotni funkciji: $8/\sqrt{2}-2\sin t+4\cos t+20\sin 2t-3\cos 2t-\sin 3t$



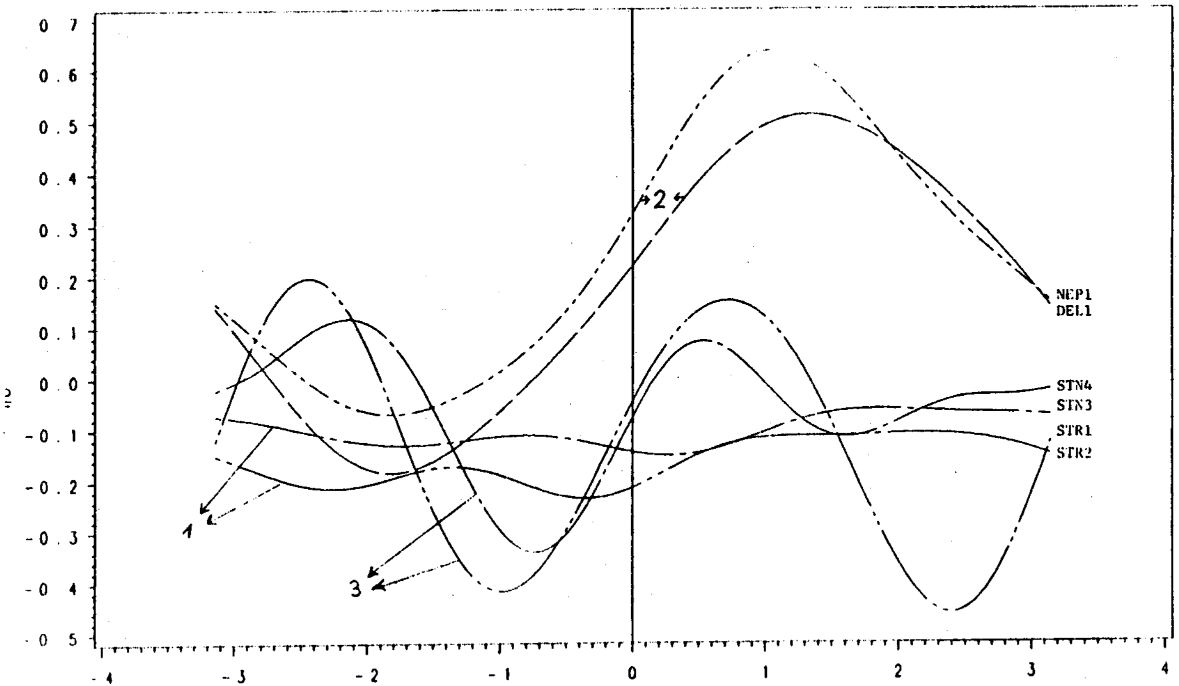
Slika 8: Prikaz Andrewsovih krivulj, ki ustrezajo prvemu (1), prvima dvema (2), prvim trem (3) itd. členom funkcije $8/\sqrt{2} - 2\sin t + 4\cos t + 20\sin 2t - 3\cos 2t - \sin 3t$



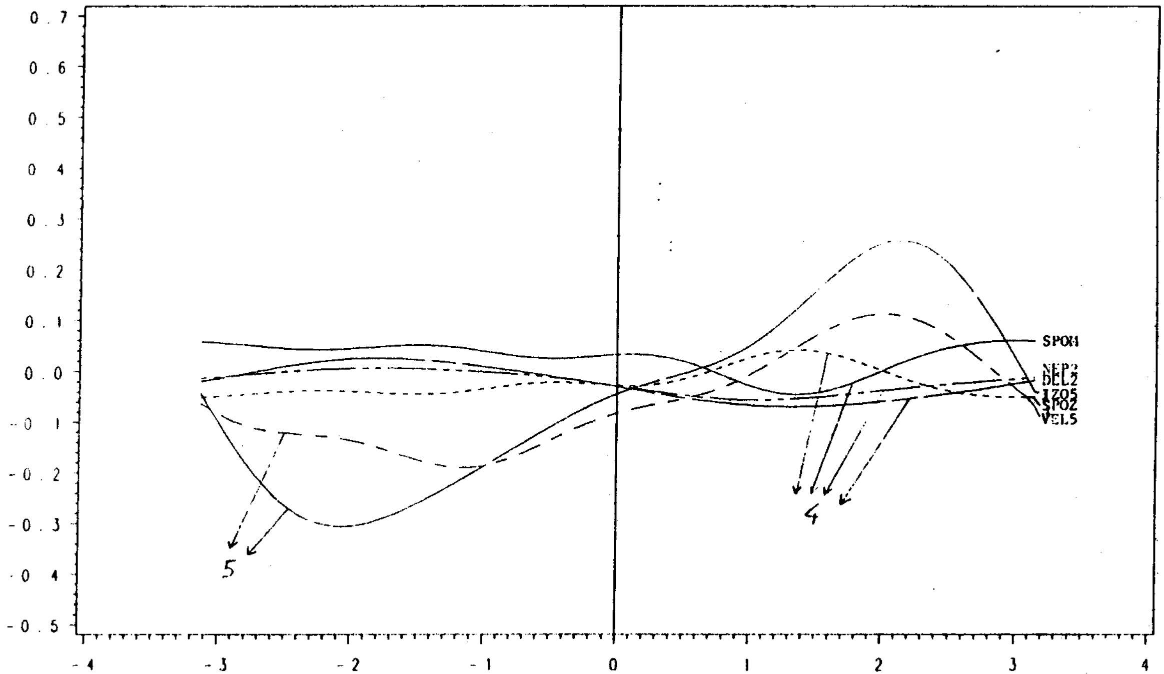
Slika 9: Prikaz položaja profilov - točk v optimalni ravnini



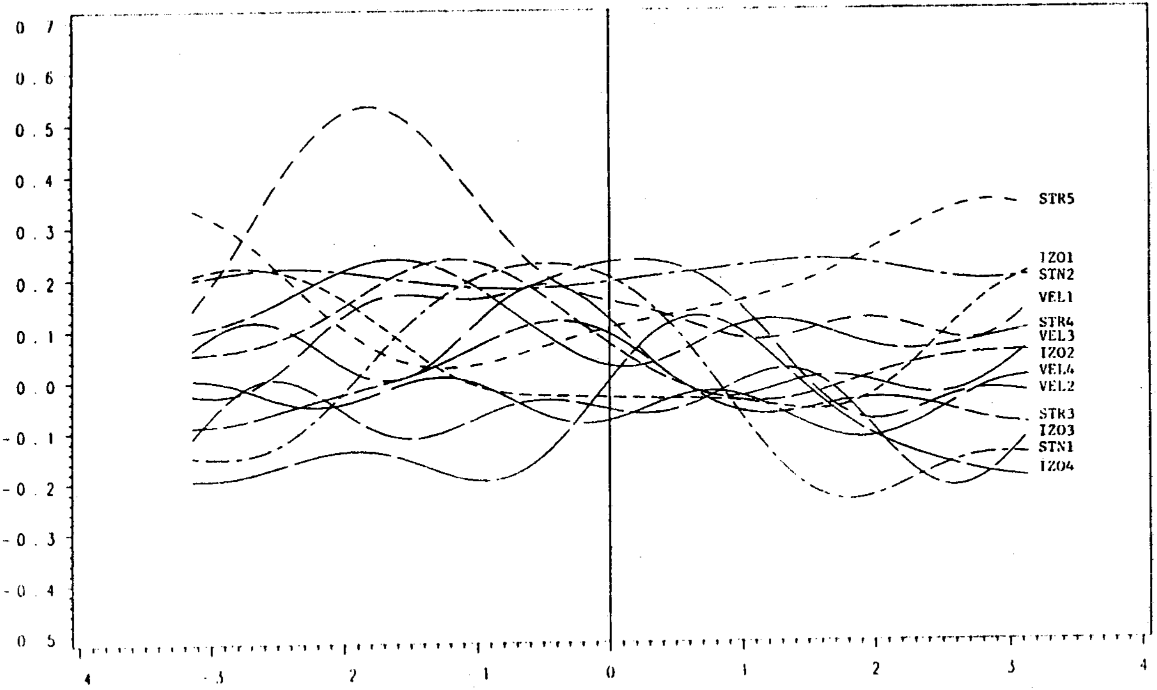
Slika 10: Prikaz položaja profilov - točk v optimalnem trirazsežnem podprostoru



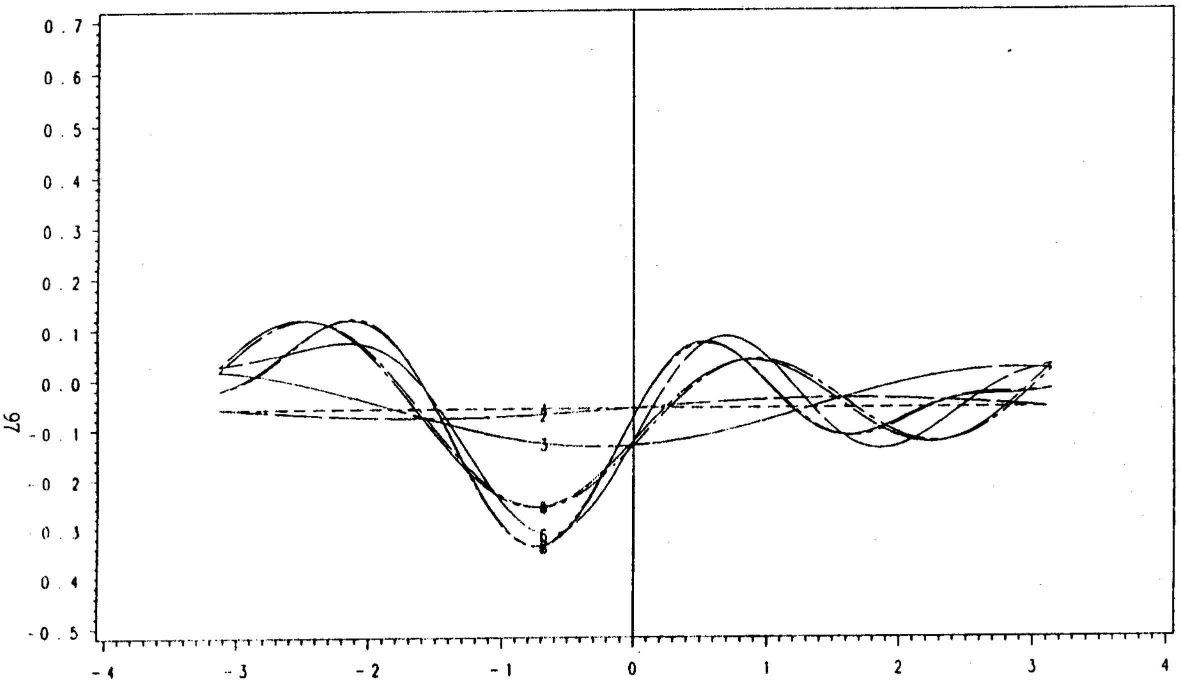
Slika 11: Andrewsov grafikon prve, druge in tretje skupine profilov.



Slika 12: Andrewsov grafikon četrte in peto skupine profilov.



Slika 13: Andrewsov grafikon tistih profilov, ki ne oblikujejo skupin



Slika 14: Andrewsov grafikon profila oseb, ki bivajo brezplačno (STN4) (na podlagi 1-8 PGK).