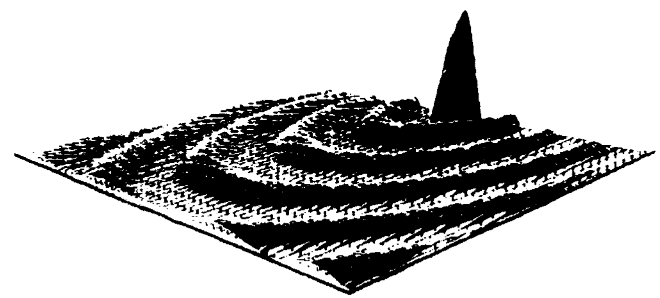
Estratto da:

ATTI DELLE GIORNATE DI STUDIO SU:

ANALISI STATISTICA DI DATI TERRITORIALI: METODI, TECNOLOGIE, APPLICAZIONI



con il patrocinio della
Società Italiana di Statistica

Cacucci Editore
Bari, 1989

Atti delle Giornate di studio su Analisi Statistica di Dati Territoriali (Bari, 1989): pp. 485-494.

SULL'USO DELL'ANALISI FATTORIALE PER IL TRATTAMENTO DI DATI TERRITORIALI

Giorgio Vittadini

1. Introduzione

Nell'analisi di dati territoriali sono stati sovente utilizzati, a fini classificatori, i punteggi fattoriali soluzione dell'analisi fattoriale. Impiegata in studi basati sia su variabili quantitative osservate su unità territoriali elementari (u.t.e.), sia su flussi di uomini e cose fra u.t.e., l'analisi fattoriale è di solito criticata per la non unicita' dei metodi di individuazione e rotazione dei fattori e quindi in definitiva per la non oggettività dei criteri di identificazione dei suoi parametri.

Si può però portare una critica ancor più radicale. E' stato dimostrato che le soluzioni del modello fattoriale non sono uniche quand'anche siano stati decisi univocamente metodi di individuazione e rotazione dei fattori e il modello sia perfettamente identificato. Poichè tali soluzioni possono essere addirittura correlate tra loro negativamente, i risultati ottenuti con l'ausilio dei punteggi fattoriali perdono il loro valore "oggettivo".

In questa nota partendo da queste considerazioni si dimostra che l'uso del modello fattoriale conduce a risultati, per certi aspetti, arbitrari.

2. Non unicità dei "factor scores" dell'analisi fattoriale

Sia $\underline{x}=(x_1, x_2, \ldots, x_m)'$ un vettore casuale m-dimensionale di variabili osservate centrate. Si assume che esistano $q < m$ variabili casuali standardizzate non osservabili $\underline{f}=(f_1, f_2, \ldots, f_q)'$ chiamate "fattori comuni" (factor scores) e m variabili casuali standardizzate non osservabili $\underline{u}=(u_1, u_2, \ldots, u_m)'$ chiamate fattori specifici.

Il modello fattoriale può quindi essere scritto nel seguente modo:

$$\underline{x} = \underline{B}\,\underline{f} + \underline{K}\,\underline{u} \tag{1}$$

con B matrice dei coefficienti di regressione tra fattori comuni e variabili osservate (m,q) (denominata anche matrice dei "factor loadings") e con $\underline{K}$ matrice diagonale dei parametri dei fattori specifici.

Per le assunzioni del modello:

$$E(\underline{x})=0 \; ; \; E(\underline{f})=0 \; ; \; E(\underline{u})=0 \; ; \; cov(\underline{u}, \underline{u})=\underline{I}_m \tag{2}$$

Si assume inoltre che:

$$cov(\underline{f},\underline{u})=0 \; ; \; \underline{K}=diag(k_1,k_2,\ldots,k_m) \; ; \; (k_i>0) \tag{3}$$

$$rango(\underline{B})=q \; ; \tag{4}$$

Per le assunzioni fatte si ha:

$$cov(\underline{x},\underline{x})=\underline{B}\ \underline{ff}'\underline{B}'+K^2 \tag{5}$$

A riguardo delle soluzioni del modello (1) si vuole, in questa sede, ricordare solo che una volta prescelto il metodo di estrazione e rotazione dei fattori e identificate univocamente le matrici di parametri $\underline{B}$ e $\underline{K}$, i valori dei "factor scores" non sono unici (Guttman 1955). Percio', in seguito, si considereranno solo equazioni inerenti realizzazioni di variabili casuali. Si indichi quindi con $\underline{X}$ la matrice (n,m) dei valori osservati di $\underline{x}$, con $\underline{F}$ la corrispondente matrice dei punteggi fattoriali di $\underline{f}$ (n,q), e con $\underline{U}$ la matrice dei punteggi fattoriali di $\underline{u}$ (n,m). Il problema dell'unicità dei punteggi fattoriali coincide con il problema della ricerca di soluzioni $\underline{F}$ del modello:

$$\underline{X}=\underline{F}\ \underline{B}'+\underline{U}\ \underline{K} \tag{6}$$

dati $\underline{X}$,$\underline{B}$,$\underline{K}$; ovvero, in modo più compatto, denominando:

$$\underline{R}=[\underline{F},\underline{U}] \; ; \; \underline{C}=[\underline{B}',\underline{X}] \tag{7}$$

con il problema della ricerca di soluzione $\underline{R}$ del modello:

$$\underline{X}=\underline{R}\ \underline{C} \tag{8}$$

dati $\underline{X}$ e $\underline{C}$, con: $cov(\underline{R},\underline{R}) = \underline{S}_R$.

Poichè il numero di vettori linearmente indipendenti in $\underline{R}$ eccede generalmente il numero dei vettori linearmente indipendenti in $\underline{X}$:

$$rango(\underline{R}) = m + q \; ; \; rango(\underline{X}) = m \tag{9}$$

il modello (1) fornisce soluzioni non uniche per $\underline{R}$. Infatti si dimostra anche (Guttman 1955) che una volta scelto il metodo di estrazione e rotazione dei fattori e quindi identificato il modello con date matrici $\underline{C}$ e $\underline{S}_R$, il modello, quando la matrice $\underline{S}_x$ è invertibile, fornisce le seguenti soluzioni per $\underline{R}$ e di $\underline{F}$ in termini di $\underline{X}$:

$$\underline{R}= \underline{X}\ \underline{S}_x^{-1}\ \underline{C}'\ \underline{S}_R + \underline{N}'\ \underline{P}\ [\underline{I} - \underline{C}\ \underline{S}_x^{-1}\ \underline{C}'\ \underline{S}_R]; \tag{10a}$$

$$\underline{F}= \underline{X}\ \underline{S}_x^{-1}\ \underline{B}\ \underline{S}_F + \underline{N}'\ \underline{P}_1\ [\underline{I} - \underline{B}'\ \underline{S}_x^{-1}\ \underline{B}\ \underline{S}_F] \tag{10b}$$

con: $\underline{P}$ matrice di dimensioni (q+m,q+m) tale che: $cov(\underline{P}, \underline{P}) = \underline{S}_R$; $\underline{N}$ matrice qualunque di dimensioni (q+m,n) tale che:

$$cov(\underline{N},\underline{N})=\underline{I}_1, \; cov(\underline{N},\underline{X}) = 0; \tag{11}$$

$\underline{P}_1$ sottomatrice delle prime q colonne di $\underline{P}$. I primi addendi delle due soluzioni (10a) e (10b) sono le stime ottime dei minimi quadrati di $\underline{R}$ ed $\underline{F}$ attraverso $\underline{X}$ costruiti per tutte le soluzioni $\underline{R}$ ed $\underline{F}$ caratterizzate da medesime matrici di varianze-covarianze $\underline{S}_R$ ed $\underline{S}_F$; i secondi addendi sono gli errori di stima che variano per ogni soluzione $\underline{R}$ ed $\underline{F}$ in quanto $\underline{N}$ può assumere qualunque valore subordinatamente alle condizioni (11). Ne discende la non unicità dei punteggi fattoriali $\underline{R}$ ed $\underline{F}$ anche per modelli identificati con la stessa matrice $\underline{C} = [\underline{B}', \underline{K}]$.

Quando $\underline{S}_x$ non e' invertibile le soluzioni (10a) e (10b) divengono:

$$\underline{R} = \underline{X}\ \underline{V} + \underline{N}'\underline{P}(\underline{I} - \underline{C}\ \underline{V}); \tag{12a}$$

$$\underline{F} = \underline{X}\ \underline{V}_1 + \underline{N}'\underline{P}_1(\underline{I}-\underline{B}'\underline{V}_1); \tag{12b}$$

con $\underline{V}$ e $\underline{V}_1$ matrici (m, m+q) e (m, q) tali che:

$$\underline{S}_x\ \underline{V} = \underline{C}'\ \underline{S}_R \; ; \; \underline{S}_x\ \underline{V}_1 = \underline{B}\ \underline{S}_F \tag{13}$$

ricavabili come stime ottime dei minimi quadrati di $\underline{R}$ e $\underline{F}$ attraverso $\underline{X}$. Le matrici $\underline{V}$ e $\underline{V}_1$ esistono sempre ma non sono uniche. Non essendo costruito in questo secondo caso neanche il primo addendo della soluzione si conferma la non unicità dei punteggi fattoriali $\underline{R}$ ed $\underline{F}$.

3. L'analisi fattoriale con dati di flusso

Il primo utilizzo dell'analisi fattoriale per effettuare partizioni del territorio in aree basati su dati di flusso risale alla prima meta' degli anni '60 (Berry 1964) con lo scopo di ottenere "mappe delle aree residenziali non limitate alle rappresentazioni di un fattore sociale per volta.... ma di una continuazione di fattori che accadono simultaneamente in ogni area " (Sforzi ed al. 1982,p.7). Successivamente numerose sono state le applicazioni come ad esempio: movimenti di taxi nella grande Londra (Goddard 1970); chiamate telefoniche in Galles (Clark 1973), chiamate telefoniche in Canada (Nader 1980, 1981). I lavori di Davies, inerenti regioni cittadine in un distretto amministrativo (1972), flussi di pendolarismo nel Galles (Davies-Musson 1979, Davies 1980), collegamenti urbani nel Montana (1979) segnano una definitiva sistematizzazione anche sul piano teorico dell'uso dell'analisi dei fattori per la costruzione di aree basati su dati di flusso.

Sia data una matrice quadrata $\underline{H}^*$ (m,n), il cui generico elemento h_{ij} indica il generico flusso di uomini (o cose) dall' u.t.e. u_i all'u.t.e. u_j. Le u.t.e. destinazione dei flussi, associate alle colonne della matrice $\underline{H}^*$ sono considerate alla stregua di variabili, mentre le u.t.e. origine dei flussi associate alle righe della matrice $\underline{H}^*$, sono considerate alla stregua di osservazioni. Dapprima le colonne della matrice $\underline{H}^*$ vengono normalizzate per eliminare l'effetto della diversa numerosità complessiva dei flussi diretti ad ogni u.t.e. sui risultati. Ottenuta così una matrice di flussi relativi $\underline{H}$ si calcola la matrice di covarianza $\underline{S}_H = cov(\underline{H},\underline{H})$.

Si estraggono poi dalla matrice $\underline{S}_H$ i fattori iniziali, ruotandoli poi generalmente in modo obliquo per ricavare fattori comuni ruotati $\underline{F}$. In questo modo nel solco della tradizionale teoria sul modello fattoriale, il modello:

$$\underline{X} = \underline{F}^{*} \underline{B}^{*} + \underline{U}\,\underline{K} \qquad (14)$$

con $\underline{F}^{*}$ e $\underline{B}^{*}$ matrici dei punteggi fattoriali e dei "factor loadings" dopo la rotazione, fornisce una soluzione fattoriale dalla struttura semplice, in cui ogni fattore e' fortemente correlato con un numero limitato di u.t.e. destinazione.

La matrice $\underline{B}^{*}$ viene utilizzata per definire insiemi di u.t.e. destinazione con identica struttura di relazioni con u.t.e. origine: ogni u.t.e. destinazione viene allocata al fattore comune con il quale ha più alta correlazione. I punteggi più elevati dei fattori comuni $\underline{F}$ individuano le principali origini dei flussi diretti verso tali insiemi di u.t.e. In questo modo si ricavano regioni del primo ordine: successivamente si possono costruire regioni di ordine superiore effettuando ulteriori analisi dei fattori sulla matrice dei punteggi fattoriali $\underline{F}$ trattata alla stregua di una matrice di variabili osservate.

Numerose sono le opzioni e le varianti possibili a questa metodologia: l'utilizzo di matrici rese simmetriche sommando flussi in entrata e uscita dalle u.t.e. o di matrici in cui si considerano come variabili le u.t.e. origine in luogo che le u.t.e.destinazione; l'adozione di diversi metodi di standardizzazione dei dati e di estrazione, fissazione del numero e rotazione dei fattori; la scelta di differenti criteri di allocazione di u.t.e. ai fattori in base ai valori della matrici $\underline{F}^{*}$ e $\underline{B}$.

Senza addentrarci nel dibattito su tali opzioni, ciò che interessa rilevare in questa sede, è il fatto che mentre l'individuazione di insiemi di u.t.e. destinazione sulla base dei valori della matrice $\underline{B}$ è del tutto lecito, una volta giustificati i metodi attraverso cui si ricavano i fattori, completamente arbitraria invece è l'allocazione di u.t.e. origine a tali insiemi di u.t.e.destinazione attraverso i valori dei punteggi fattoriali della matrice $\underline{F}$, qualunque siano le scelte effettuate in precedenza. Infatti, per quanto ricordato nel paragrafo 2, la matrice $\underline{F}$ non è unica anche per una data matrice di "factor loading" $\underline{B}$ e una data matrice di correlazione S_F.

Perciò, attraverso l'analisi fattoriale si possono ottenere determinati insiemi di u.t.e destinazione ma non le corrispondenti principali u.t.e. origine, in quanto esse variano a seconda delle soluzioni $\underline{F}^{*}$ utilizzate per individuarle. Nel successivo paragrafo 4 si esemplifica tale conclusione con una applicazione.

4. Un'applicazione empirica

Al fine di verificare anche empiricamente la validità di quanto affermato nel precedente paragrafo, si costituiscono regioni sulla base dei flussi telefonici esistenti fra i 20 compartimenti telefonici italiani (fonte SIP 1988). Si effettua quindi, nel modo indicato nel paragrafo 3, l'analisi dei fattori considerando come variabili i flussi normalizzati con identica destinazione. Si ricavano i fattori con il metodo della massima verosimiglianza mentre il numero dei fattori viene fissato con il criterio degli autovalori di Kaiser (Takeuchi ed al., 1982, p.319), e la rotazione obliqua viene effettuata attraverso il metodo "promax" (Takeuchi ed al., 1982, p.319). Ai quattro fattori conservati e' associata una quota rispettivamente pari al 27.1%, 20.8%, 12.9%, 9.2% della varianza complessiva. Nella tabella 1 appaiono i valori dei "factor loadings" e dei punteggi dei

fattori comuni di due soluzioni fattoriali valide per un modello ugualmente identificato.

Tab. 1: Matrice dei coefficienti di regressione tra u.t.e destinazione e fattori comuni; matrici dei punteggi fattoriali delle u.t.e.origine

Coeff. regressione				Punt.fatt.1ˆsoluz.				Punt.fatt.2ˆsoluz.			
1	2	3	4	1	2	3	4	1	2	3	4
.032	.021	.056	.107	-.407	.170	-.106	-1.001	1.071	-.094	.402	1.385
.047	.030	.029	-.117	-1.506	-.332	-.111	-1.743	1.630	.451	.521	2.342
.043	-.005	.097	.052	-.471	.051	-.511	-.526	-.273	-.045	.137	.030
-.002	-.012	.197	-.037	1.635	2.023	-.070	.090	-.670	-2.007	-.090	-.173
.000	.014	.109	-.011	.323	-.381	1.137	-.806	.153	.421	-1.195	.707
.091	.019	.040	.070	1.101	.616	.102	1.374	-.630	-.697	-.429	-1.823
.042	.102	.031	-.007	.917	2.267	.716	-.548	-.540	-2.229	-.531	.237
-.056	.191	-.015	-.055	1.549	-1.200	-.113	.639	.531	1.149	.095	-.859
-.003	.173	.014	-.033	.146	.754	-.109	.076	-1.156	-.764	-.061	-.433
.234	-.039	.009	-.050	-1.476	-1.322	-.662	-.406	1.506	1.326	.744	.260
.060	.030	.000	-.003	-.425	-.032	.066	-.197	-.267	.836	-1.107	.153
.024	.052	.033	.020	-.134	-.027	.252	1.095	-1.520	.005	-.171	-.516
.100	-.055	-.006	.006	-.430	-.900	-.505	-.691	.545	.924	.514	.413
.000	-.025	-.011	.003	.747	.273	1.320	1.797	-.311	-.321	-.902	-1.049
-.004	.130	-.024	.024	.215	-.104	1.710	.409	-.005	.004	-1.672	-.464
.044	.040	-.012	.010	1.265	1.271	1.431	1.316	-1.711	-1.316	-1.606	-1.273
.072	.057	-.013	.002	-1.002	-.009	-2.304	-.731	1.953	.906	1.900	.418
.127	.060	-.001	-.009	-.615	-1.067	-.713	-.935	-.255	.986	.355	.223
.115	.020	.000	.012	.137	-.437	-.090	-1.068	-.329	.503	1.460	1.330
.097	.027	-.013	-.001	-.600	.070	-1.202	1.050	.206	-.116	[illegible]	[illegible]

Nella tabella 2 si indicano gli insiemi di comprensori destinazione caratterizzati da flussi con comune origine mentre nella tabella 3 i relativi comprensori origine più importanti deducibili della matrice dei fattori comuni. Gli insiemi di comprensori destinazione con relativi principali comprensori origine dei flussi sono poi visualizzati nella figura 1.

Tab. 2: Aree di comprensori destinazione con comune origine dei flussi;

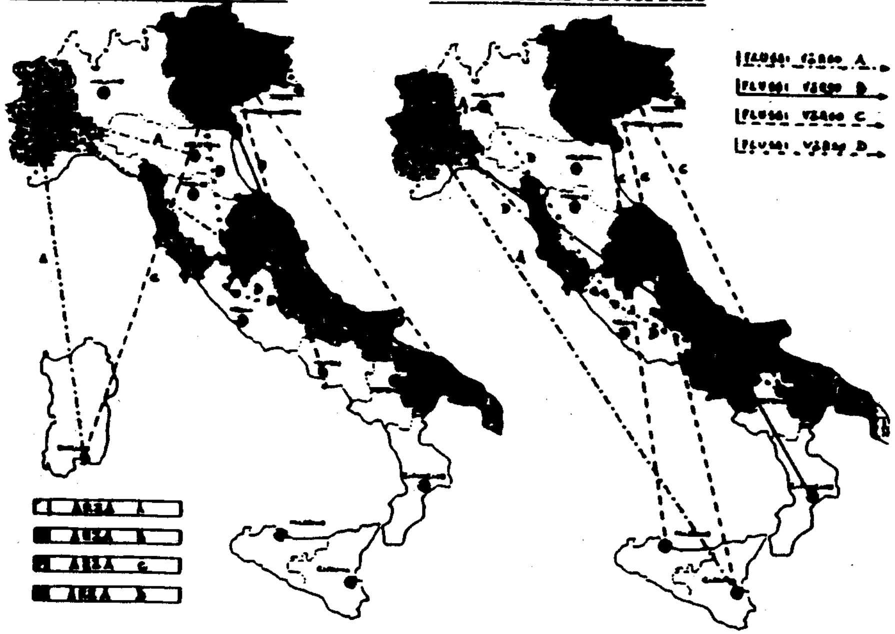
Area A:	Ancona, Perugia, Pescara, Pisa, Napoli, Bari
Area B:	Bolzano, Venezia, Bologna
Area C:	Milano, Bologna, Genova, Firenze, Roma, Cagliari, Potenza, Catanzaro, Catania, Palermo
Area D:	Torino

Tab. 3: Principali comprensori origine dei flussi per ogni insieme di comprensori destinazione

	1 soluzione	2 soluzione
Area A:	Ancona, Bolzano, Bari	Genova, Perugia, Catanzaro
Area B:	Napoli, Bari, Cagliari	Potenza, Palermo, Catania
Area C:	Bolzano, Perugia, Bari	Potenza, Milano, Genova
Area D:	Cagliari, Bologna, Bari	Milano, Catania, Torino

Fig.1: Aree di comprensori destinazione e principali comprensori origine dei flussi secondo la 1ˆ e la 2ˆ soluzione fattoriale

1ˆ soluzione fattoriale   2ˆ soluzione fattoriale



Dalle tabelle 1 e 3 e dalla figura 1 si osserva che, mentre sono individuabili univocamente gli insiemi di comprensori aventi flussi comune origine dei flussi, e' impossibile individuare in modo univoco quali siano i comprensori origine dei flussi.

5. L'analisi fattoriale utilizzata con caratteristiche misurabili

Per la costruzione di aree omogenee in base a caratteristiche misurabili delle u.t.e. si effettua usualmente una cluster analysis, utilizzando sovente come input, quando si voglia ridurre la dimensione dei dati, le prime q (q<m) componenti principali (c.p.) cui sia associata una quota elevata della varianza delle variabili di partenza.

Data una matrice $\underline{X}$ di dimensione (n,m) delle osservazioni di m variabili su n u.t.e. si definisce come analisi delle c.p. di $\underline{X}$ la ricerca di una serie di trasformate lineari della matrice $\underline{X}$

$$\underline{w}_r = \underline{X}\ \underline{d}_r \quad (r= 1,2,\ldots, m) \qquad (15)$$

dette "c.p." che spieghino quanta più parte possibile delle variabili originali, subordinatamente alla condizione di ortonormalità dei vettori ignoti $\underline{d}_r$. Si dimostra, attraverso il metodo dei moltiplicatori di Lagrange che i vettori $\underline{d}_r$ (r= 1,2,...,m) sono gli autovettori successivi associati alla matrice di correlazione (o covarianza) $cov(\underline{X},\underline{X}) = \underline{S}_x$, mentre la varianza

associata ad ogni c.p. è pari al corrispondente autovalore di $\underline{S}_x$. La matrice di dimensioni (n, m) delle c.p. è quindi:

$$\underline{W} = \underline{X}\ \underline{D} \qquad (16)$$

in cui $\underline{W}$ è la matrice le cui colonne sono le c.p. e $\underline{D}$ è la matrice di dimensioni (m,m), le cui colonne sono i rispettivi autovettori $\underline{d}_r$ ricavati da $\underline{S}_x$. L'analisi dei gruppi viene quindi effettuata sulla matrice:

$$\underline{W}_q = \underline{X}\ \underline{D}_q \qquad (17)$$

in cui $\underline{W}_q$ e $\underline{D}_q$ (n,q) e (m, q) sono le matrici delle prime q componenti e dei primi q autovettori cui sia associata una quota elevata della varianza iniziale.

In molti lavori in luogo dell'analisi delle c.p. si utilizza o si consiglia l'utilizzazione dell'analisi di fattori, per lo più estratti con il metodo delle componenti principali (Spence 1968, Zani 1974, Good 1977, Brunetta 1983). Le due analisi non sono però equivalenti. Mentre infatti il rango della matrice dei fattori come si è visto in (9), eccede generalmente il rango della matrice delle variabili osservate, il rango della matrice $\underline{W}$ è sempre minore o uguale del rango della matrice delle variabili osservate. Infatti supponendo per semplicità che $\underline{X}$ sia di rango pieno si ha:

$$\text{rango } (\underline{W} = \underline{X}\ \underline{D}) = m \ ; \ \text{rango } (\underline{X}) = m \qquad (18)$$

Se ne deduce che la matrice delle c.p. $\underline{W}$ è unica a differenza della matrice dei punteggi fattoriali $\underline{F}^*$. Perciò non appare metodologicamente corretto utilizzare come input di una cluster analysis in luogo delle c.p. punteggi fattoriali $\underline{F}^*$, anche se estratti con il metodo delle c.p.. Infatti dalla medesima matrice di variabili osservate $\underline{X}$, adottando questo identico metodo di estrazione dei fattori e un identico criterio di rotazione si possono ricavare differenti matrici di factor scores da utilizzare come input nella cluster analysis. Come si mostra in un secondo esempio nel successivo paragrafo, utilizzando l'analisi dei fattori come criterio di riduzione delle variabili si ottengono quindi aggregazioni delle u.t.e. in aree omogenee del tutto arbitrarie.

6. Una seconda applicazione empirica

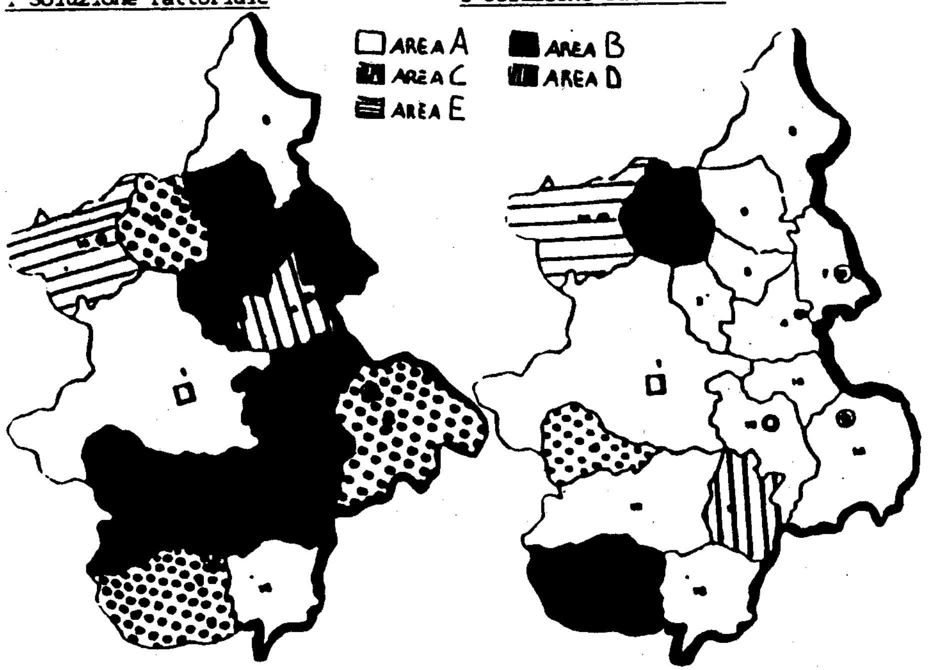
Si effettua una cluster analysis tra i 17 bacini del mercato del lavoro in Piemonte e Val d'Aosta. Le variabili su cui si effettua l'analisi sono gli indici di dotazione (rapporti tra addetti nel settore e popolazione residente) per agricoltura, industria e a servizi suddivisi in quattro settori (fonte: Martini,1989,pp.129-130). Per verificare quanto visto nel paragrafo precedente si effettua sulle 6 variabili un'analisi fattoriale ricavando i fattori con il metodo della massima verosimiglianza, fissando il numero dei fattori con il metodo degli autovalori di Kaiser e il criterio di rotazione obliqua "promax". Ai tre fattori conservati e' associata una quota rispettivamente pari al 33.7%, al 26.2%, al 25.4% della varianza complessiva. La successiva cluster analysis è di tipo gerarchico, agglomerativo, e utilizza la distanza euclidea e il legame singolo. Nella tabella 4 sono indicati i valori dei punteggi dei fattori comuni di due soluzioni fattoriali valide per un modello identicamente identificato. Nella figura 2 sono indicati,a titolo di esempio, le suddivisioni in cinque aree omogenee dei bacini sulla base del primo fattore delle due soluzioni fattoriali. Nella tabella 5 e' indicato il valore dell'indice normalizzato di concordanza della configurazione dei gruppi (Fowlkes e Mallows 1983) per diversi passi delle coppie di procedure agglomerative basate sui primi 3,2,1 fattori delle due soluzioni fattoriali:

Tab.4: Valori dei punteggi dei fattori comuni

	1 Soluzione fattoriale			2 Soluzione fattoriale		
	1	2	3	1	2	3
1	.369	1.502	.908	.287	1.400	.800
2	-.698	-.792	.618	-.476	-.752	.674
3	-.695	-1.471	-.713	-1.057	-1.535	-.801
4	-.804	.257	-.308	-.090	.241	-.410
5	-.749	.522	1.994	-.275	.606	2.100
6	-.223	-1.023	2.291	-.753	-1.117	2.163
7	-.431	.729	.549	.039	.011	.665
8	.318	-1.042	.269	.376	-1.030	.283
9	.855	1.409	-.477	1.303	1.409	-.369
10	-1.090	-.075	-.867	-.776	-.021	-.709
11	-.337	.671	-.326	-1.403	.469	-.606
12	-.468	-1.434	-.658	-.292	-1.402	-.614
13	-.965	.328	-1.419	-.375	.432	-1.277
14	.705	1.236	-.654	-.097	1.094	-.846
15	-.614	.152	-.562	-.340	.200	-.494
16	2.969	.330	-.035	2.677	.270	-.105
17	1.065	-1.301	-.524	1.329	-1.253	-.460

Fig 2: Suddivisione in cinque aree omogenee nella cluster analysis basata sul primo fattore

1 Soluzione fattoriale 2 Soluzione fattoriale



Tab.5: Indice di concordanza della configurazione dei gruppi tra le cluster an. basate sulle due soluzioni fattoriali

	Analisi dei gruppi		
passi	con 3 fatt.	con 2 fatt.	con 1 fatt.
7	.647	.517	.380
10	.733	.595	.355
12	.634	.595	.582

Dal basso valore degli indici di concordanza e dalla grande differenza nella configurazione dei gruppi viene confermata l'arbitrarietà delle classificazioni ottenute con l'ausilio dell'analisi fattoriale.

7. Conclusioni

L'utilizzo di punteggi fattoriali per l'analisi dei dati territoriali fornisce risultati arbitrari. L'indicazione operativa è quella di ricorrere ad altri metodi di costruzione di aree basati su dati di flusso (es. misure di associazione) e di riduzione delle variabili nel caso di costruzione di aree omogenee.

BIBLIOGRAFIA

BERRY J.B.L.(1964),"Cities as systems within systems of cities", Papers and Proceedings of Regional Science Association, vol.13,pp.157-176.

BIANCHI G., OPENSHAW S., SFORZI F. e WYMER C. (1982), "Recenti sviluppi nell'analisi delle aree sociali", III° Conferenza Statistica di Scienze Regionali, Venezia, 10-12 novembre.

BRUNETTA R. (1983), Economia del lavoro, Marsilio Venezia.

CLARK D. (1973), "The formal and functional structures of Wales", Annals of the Association of American Geographer, vol. 63 pp. 71-85.

DAVIES W.K.D. (1972), "Conurbation and city region in an administrave borderland", Regional Studies, vol. 6, pp. 217-236.

DAVIES W.K.D. (1979), "Urban connectivity in Montana", Annals of Regional Science, vol. 13, pp. 29-46.

DAVIES W.K.D. (1980), "High order factor analysis and functional regionalisation: a cose study in south Wales, 1971", Environment and planning, vol. 12, pp. 685-701.

DAVIES W.K.D. e MUSSON T. C. (1978), "Spatial patterns of commuting in south Wales 1951-1971: factor analysis definition", Regional Studies, vol. 12, pp. 353-366.

FOWLKES B. e MALLOWS C.L. (1983), "A method for comparing two hierarchical clustering", Journal of the American Statistical Association, vol. 73, pp. 553-564.

GODDARD J.B. (1970), "Functional Regions within the city center. A study by factor analysis of taxi flows in central London", Transaction of the Institute of British Geographics, vol. 49, pp. 161-182.

GOOD I.S. (1977), "The botryology of botryology" in Van Ryzin ed., Classification and clustering, New York, pp. 73-94.

GUTTMAN L. (1955), "The determinacy of factor score matrices with implication for five other basic problems of commun factor theory", The British Journal of Statistical Psichology, vol. 9, pp. 65-81.

MARTINI M. (1989), "La nuova geografia dello sviluppo economico italiano" in A.A.V.V. Il sistema terziario in Italia, ISCOM, Roma, pp. 83-156.

NADER G.A. (1980), "An economic regionalisation of Canada: the validity of provinces as regions for the conduct of regional economic policy", Canadian Journal of Regional Science, vol. 2, pp. 117-138.

NADER G.A. (1981), " The delineation of hierarchy of nodal regions by means of higher-order factor analysis", Regional Studies, vol. 15, pp. 475-492.

SPENCE N.A.(1968),"A multifactor uniform regionalization of british counties on the basis of employment data for 1961",Regional Studies , vol.2, pp.87-104

TACHEUCHI K., YANAY L. e MURHERJEE (1982),"Multivariate analysis",New York, Wiley.

ZANI S.(1974),"Classificazione dei comuni parmensi secondo le caratteristiche delle aziende agrarie , in "Studi e Ricerche",vol.X-XI, Facolta' di Economia e Commercio, Universita' di Parma.