

Verso un indice di deprivazione a livello aggregato da utilizzare su scala nazionale: metodi di validazione

L. Grisotto, D. Catelan, A. Biggeri

Università degli Studi di Firenze
Unità di Biostatistica – CSPO Firenze

Convegno AIE Roma 15 maggio 2008

SCOPO DEL LAVORO

Lo scopo del presente lavoro è validare l'indice di deprivazione prodotto a livello aggregato e su scala nazionale.

La validazione riguarda:

- la coerenza nella scelta delle variabili che compongono l'indice con la struttura di correlazione empirica,
- la robustezza dell'indice come risultante della scelta dei pesi o dell'importanza attribuita alle sue componenti,
- la predittività dell'indice rispetto a esiti sanitari che la letteratura associa alle dimensioni socio-economiche.

MATERIALI

DATI A DISPOSIZIONE

- Dati comunali al censimento 2001, fonte ISTAT;
- Archivio geografico informativo dei comuni italiani, fonte ISTAT;
- Automobili immatricolate per comune 2001, fonte ACI;

- Dati di mortalità comunali 1999-2001, fonte ISTAT.



DISEGNO dello STUDIO

Il progetto si sviluppa in 3 fasi:

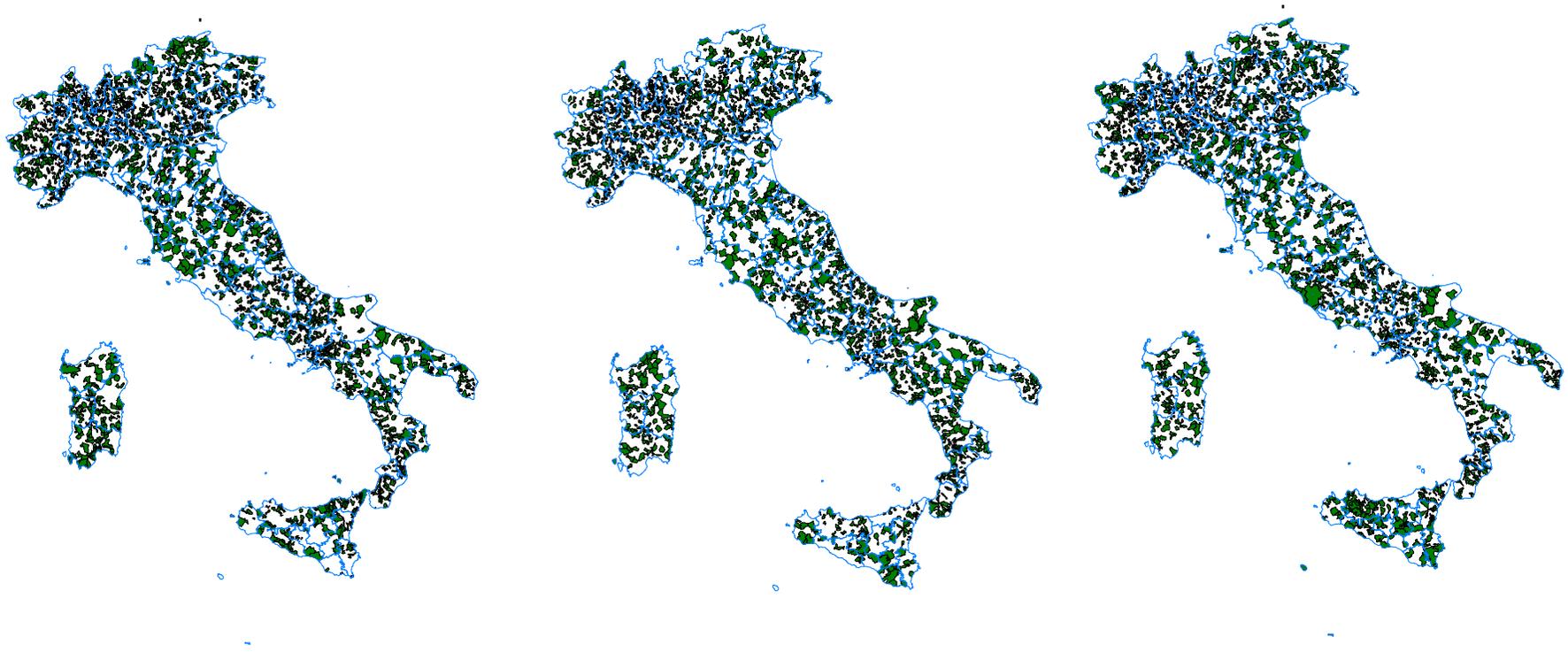
- Fase 1: Analisi Esplorativa delle relazioni
- Fase 2: Specificazione del modello e calcolo dell'indice
- Fase 3: Validazione predittiva dell'indice

Ogni fase viene eseguita su un DIFFERENTE campione casuale dei Comuni italiani

DISEGNO (segue)

Campioni

- Gli 8101 comuni italiani sono stati suddivisi in 3 campioni, **mutuamente esclusivi**. L'assegnazione di ogni comune al campione è avvenuta in modo casuale;



METODI – fase 1 e fase 2

- Analisi fattoriale esplorativa;
- Analisi fattoriale confermativa;

Entrambe le analisi si basano sui modelli a variabili latenti:

$$\mathbf{y}_j = \boldsymbol{\beta} + \Delta_y \boldsymbol{\eta}_j + \boldsymbol{\varepsilon}_j$$

Dove $j = 1 \dots N$ indica il comune

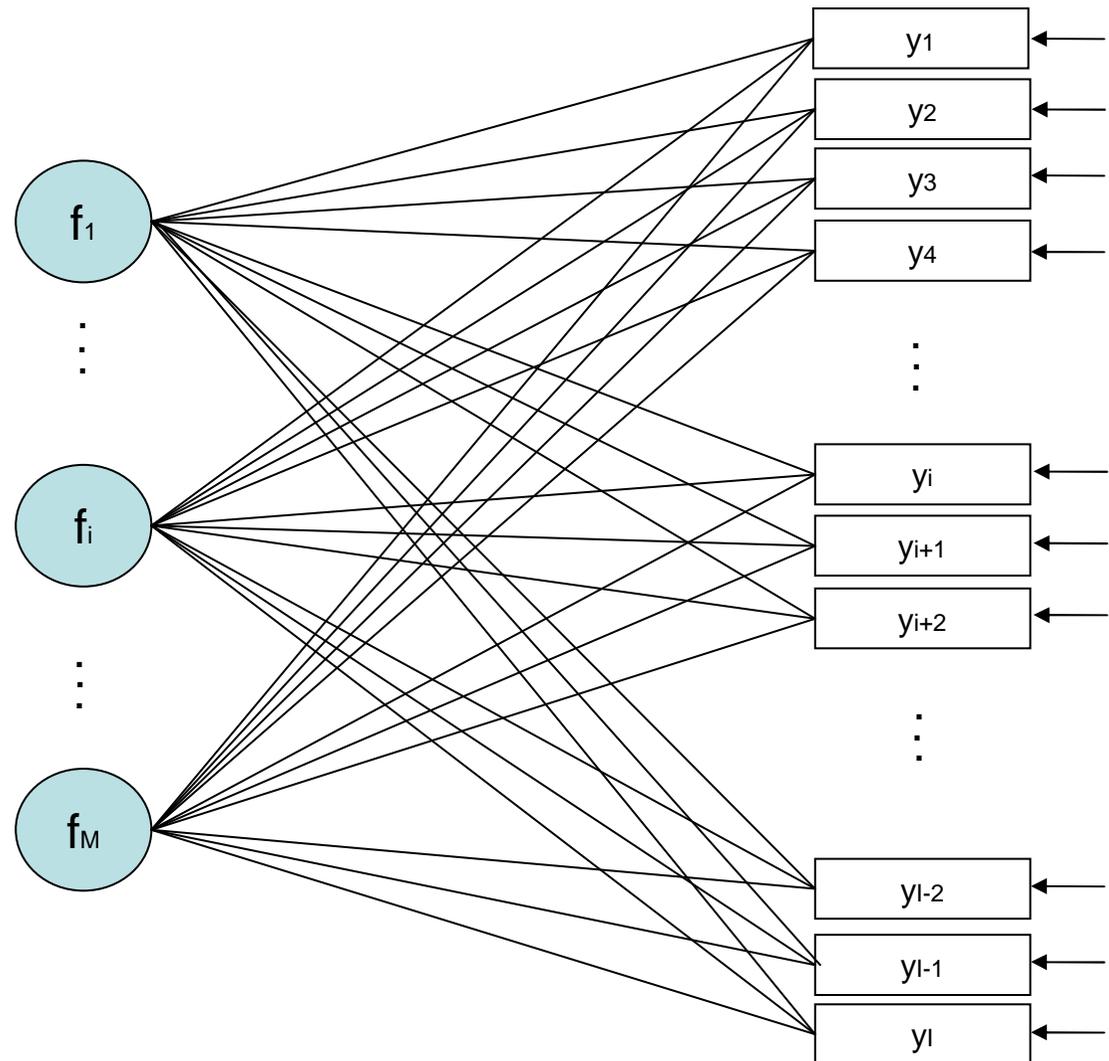
- \mathbf{y}_j è il vettore delle \mathbf{I} variabili osservate;
- $\boldsymbol{\beta}$ è il vettore delle medie, o messo se le \mathbf{I} variabili $\{\mathbf{y}\}$ sono centrate;
- Δ_y è la matrice dei pesi fattoriali;
- $\boldsymbol{\eta}_j$ è il vettore degli \mathbf{M} fattori comuni (variabili latenti);
- $\boldsymbol{\varepsilon}_j$ è il vettore dei fattori specifici, con matrice di covarianza diagonale

PATH DIAGRAM: ANALISI FATTORIALE ESPLORATIVA

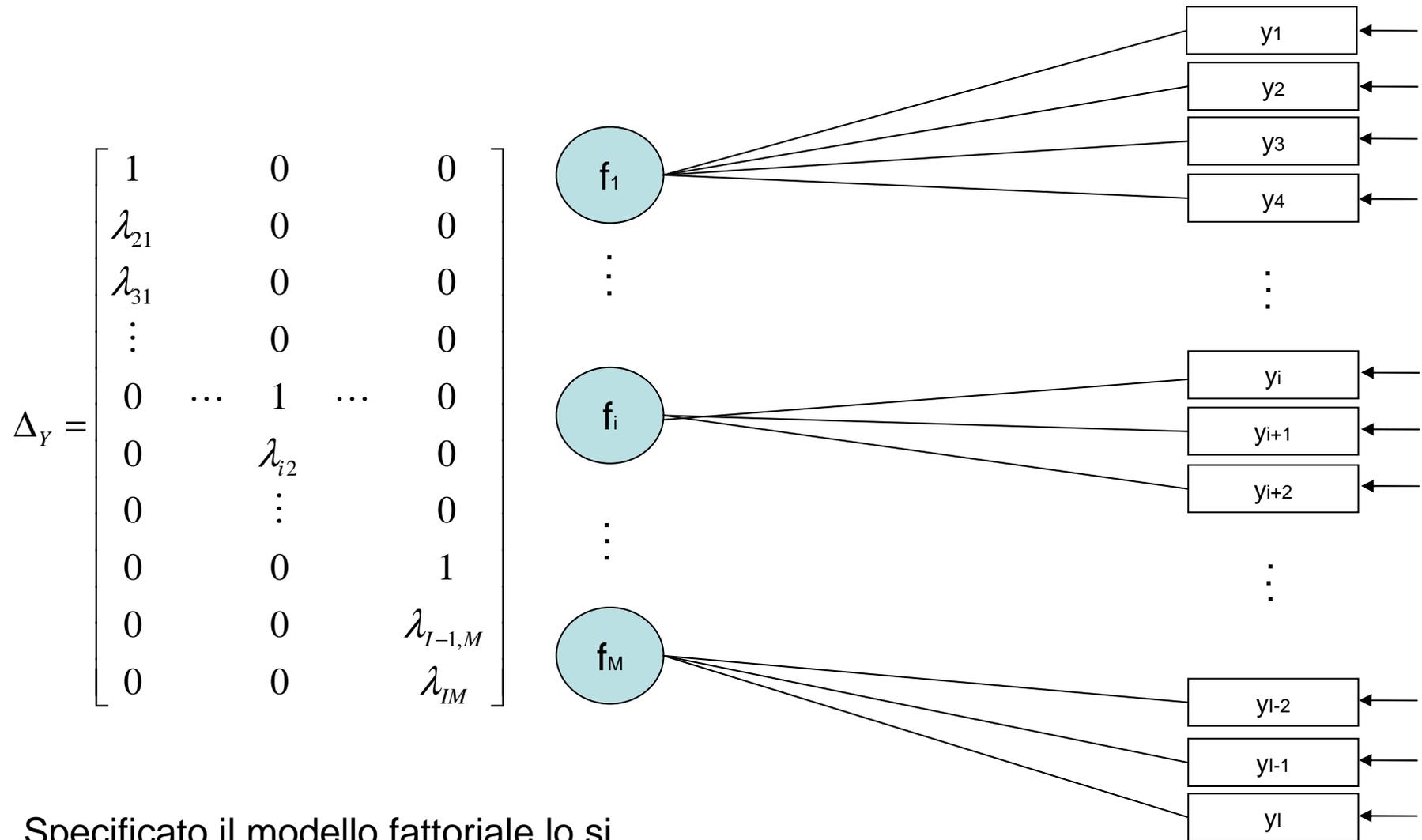
$$\Delta_Y = \begin{bmatrix} \lambda_{11} & \dots & \lambda_{1M} \\ \vdots & \ddots & \vdots \\ \lambda_{I1} & \dots & \lambda_{IM} \end{bmatrix}$$

alcuni pesi fattoriali saranno vicini a zero.

Le correlazioni tra le variabili sono spiegate da (pochi) fattori latenti



PATH DIAGRAM: ANALISI FATTORIALE CONFERMATIVA



Specificato il modello fattoriale lo si adatta su un differente campione

RISULTATI

Fase 1 : Analisi Esplorativa

L'obiettivo dell'analisi esplorativa è di verificare le relazioni tra le variabili.

Una volta individuati dalla revisione della letteratura gli indicatori semplici più utilizzati che potevano essere costruiti con i dati a nostra disposizione abbiamo effettuato:

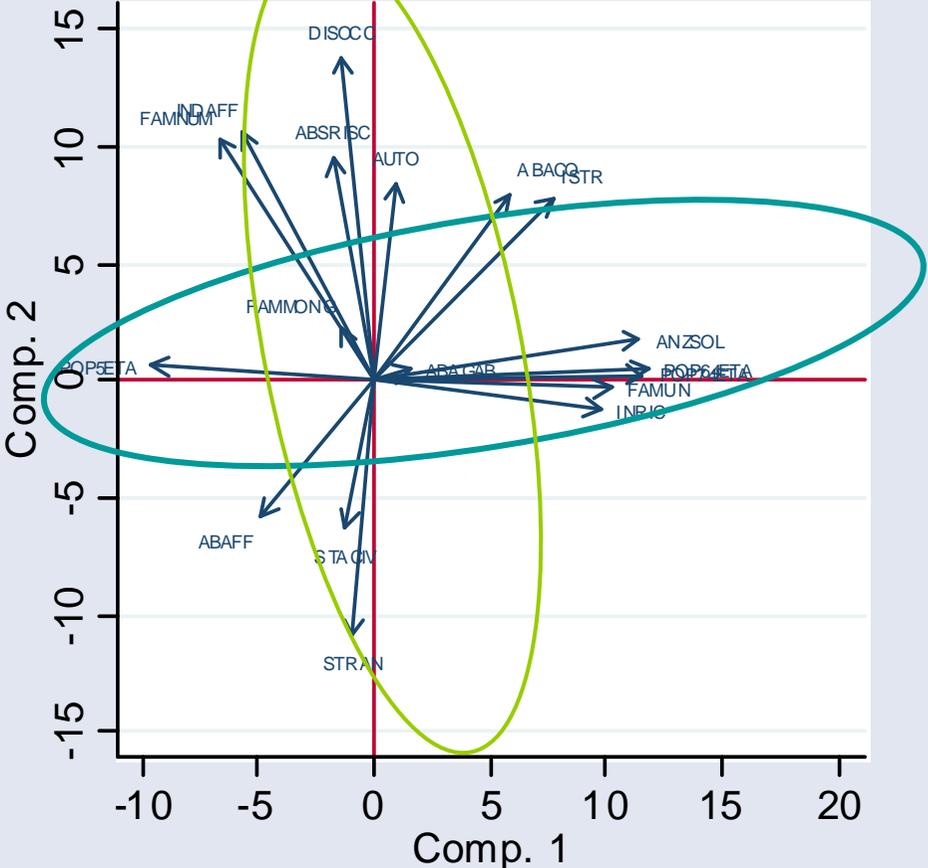
- Una rappresentazione grafica delle relazioni tra le variabili (biplot);
- Una analisi fattoriale esplorativa con il metodo della massima verosimiglianza:
 - prima non ponendo alcun vincolo sul numero di fattori;
 - in seguito vincolando il numero di fattori a 3;

INDICATORI SEMPLICI UTILIZZATI

- **DISOCC:** Popolazione 15 anni e più disoccupata o in cerca di occupazione su popolazione 15 anni e più appartenente alle forze di lavoro.
- **INDAFF:** superficie delle abitazioni occupate da persone residenti.
- **AUTO:** Popolazione residente totale su Numero di automobili immatricolate (2001 fonte ACI).
- **ISTR:** Popolazione residente 6 anni e più con licenza elementare, alfabeti e analfabeti su popolazione residente totale 6 anni e più
- **INDVEC:** Popolazione residente di 65 anni e più su popolazione residente con meno di 20 anni
- **POP<5:** Popolazione residente con meno di 5 anni su popolazione residente totale
- **POP64+:** Popolazione residenti di 65 anni e più su popolazione residente totale
- **POP74+:** Popolazione residenti di 75 anni e più su popolazione residente totale
- **FAMNUM:** Famiglie residenti con più di 5 componenti su famiglie residenti totale
- **FAMUNIP:** Famiglie unipersonali non in coabitazione su famiglie residenti totale
- **FAMMONG:** Famiglie mononucleari (con e senza membri isolati) madre o padre con figli su famiglie residenti totale
- **ANZSOLI:** Famiglie unipersonali non in coabitazione di 65 anni e più su famiglie residenti totale (anziani soli)
- **STACIV:** Popolazione residente coniugata o separata di fatto su popolazione residente totale
- **ABAFF:** Abitazioni occupate da persone residenti in affitto sulle abitazioni occupate da persone residenti
- **ABSRISC:** Abitazioni occupate da persone residenti non fornite di riscaldamento su abitazioni occupate da persone residenti
- **ABSACQ:** Abitazioni occupate da persone residenti senza disponibilità di acqua calda in bagno o cucina su abitazioni occupate da persone residenti
- **ABSGAB:** Abitazioni occupate da persone residenti senza acqua potabile e gabinetto su abitazioni occupate da persone residenti
- **STRAN:** Stranieri residenti in Italia su popolazione residente totale

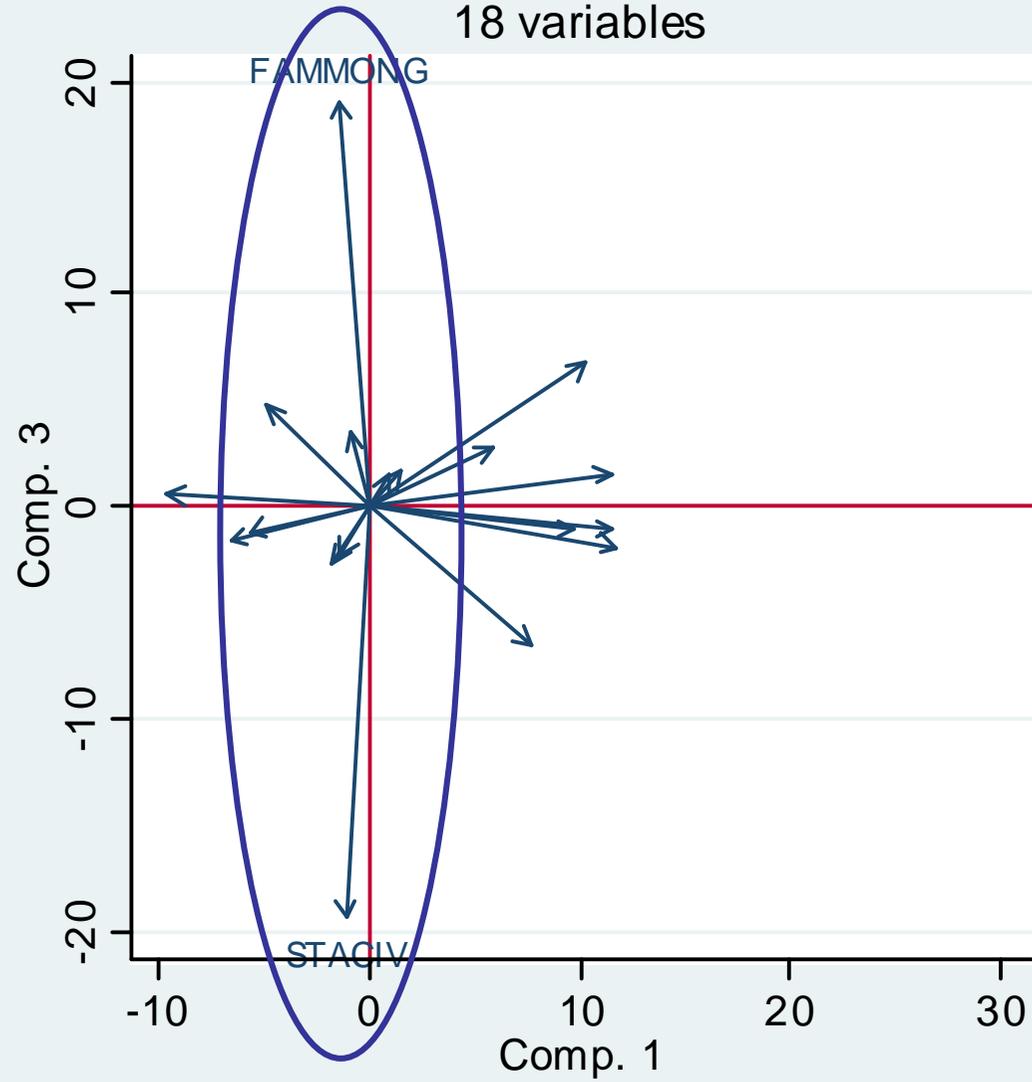
Principal-component Biplot

18 variables

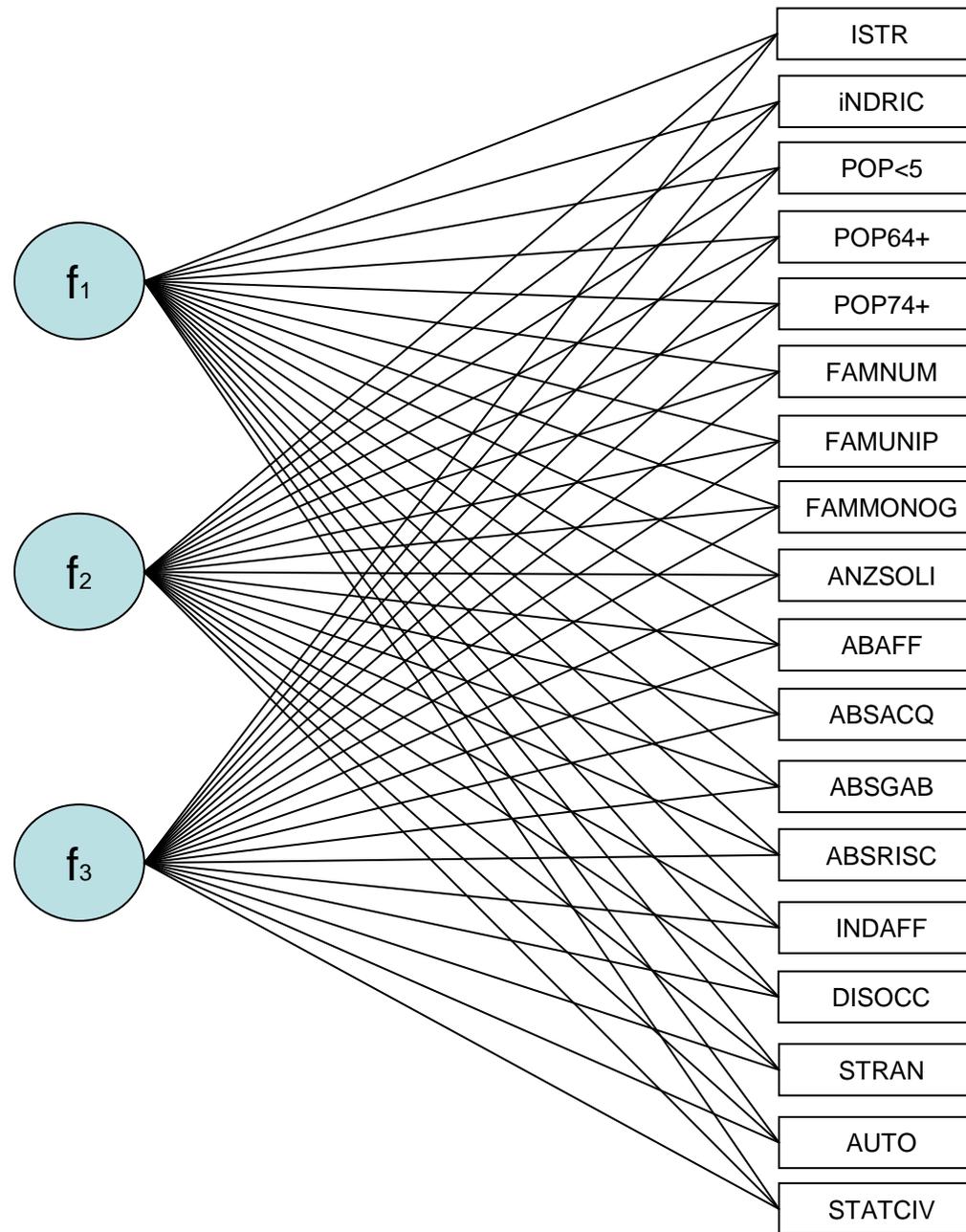


Principal-component Biplot

18 variables



ANALISI FATTORIALE ESPLORATIVA 3 FATTORI



RISULTATI ANALISI FATTORIALE ESPLORATIVA

METODO ML 3 FATTORI

Variable	Factor1	Factor2	Factor3	Uniqueness
DISOCC	-0.0317	0.7015	0.0692	0.5021
INDAFF	-0.3608	0.5768	0.0639	0.5330
AUTO	0.1057	0.3724	0.0777	0.8441
ISTR	0.6524	0.4018	-0.1038	0.4021
INDVEC	0.7430	-0.1355	0.0447	0.4276
POP<5	-0.7330	0.1180	0.0165	0.4485
POP64+	0.9946	-0.0426	-0.0247	0.0083
POP74+	0.9660	-0.0532	-0.0105	0.0640
FAMNUM	-0.4161	0.6504	-0.2171	0.3567
FAMUNIP	0.7334	-0.1725	0.6575	0.0000
FAMMONG	-0.1228	0.0695	0.1419	0.9600
ANZSOLI	0.8898	-0.0096	0.3228	0.1040
STACIV	-0.0362	-0.2443	-0.5469	0.6400
ABAFF	-0.3259	-0.2433	-0.0422	0.8328
ABSRISC	-0.0843	0.4381	0.0932	0.7923
ABSACQ	0.4081	0.3291	0.2509	0.6622
ABSGAB	0.0677	0.0063	0.0264	0.9947
STRAN	-0.0935	-0.4927	-0.0060	0.7484

RISULTATI

Fase 2: Specificazione del Modello e Calcolo dell'indice

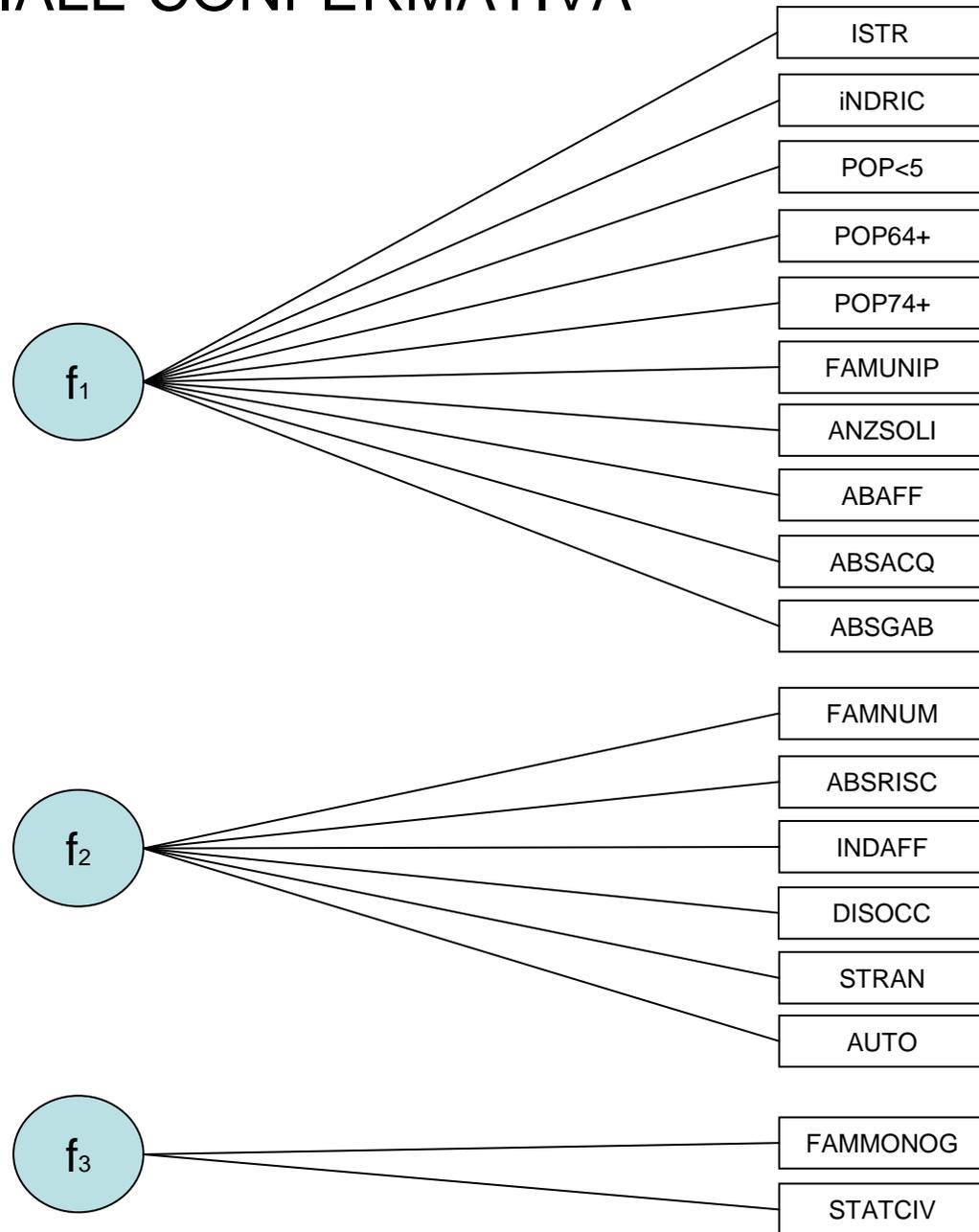
Utilizzando i risultati dell'analisi esplorativa si definisce l'indice, si valuta la bontà di adattamento e la coerenza geografica.

In questa fase viene deciso se calcolare o meno un indice per macroarea o solo a livello nazionale.

A questo scopo si verificano le associazioni tra le variabili originali e l'indice (*confirmatory analysis*) e tra l'indice e altri indici calcolati con metodi più semplici (somma punti z).

ANALISI FATTORIALE CONFERMATIVA

modello scelto



RISULTATI

ANALISI FATTORIALE CONFERMATIVA

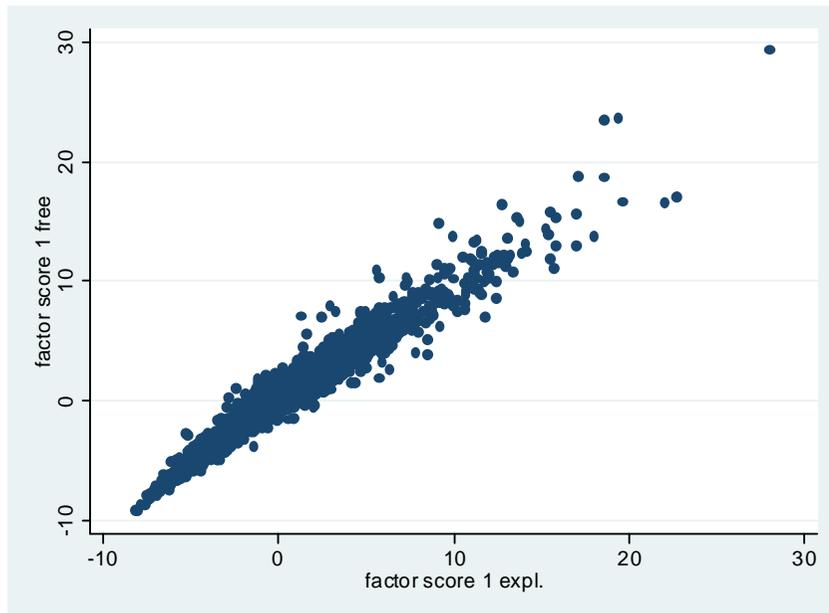
TESTS OF MODEL FIT	F1	F2
CFI	0.858	0.899
TLI	0.817	0.831
SRMR (Standardized RMS Residual)	0.067	0.046

Il terzo fattore è costituito sostanzialmente da due sole variabili !

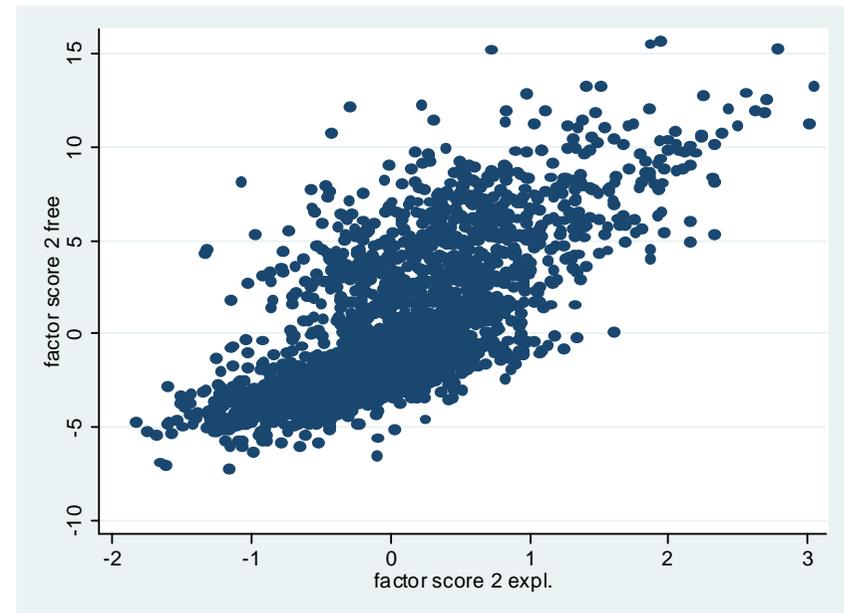
non è possibile effettuare un'analisi fattoriale confermativa (il modello non è identificato). Il terzo fattore è perciò dato dalla somma dei punteggi Z

Correlazione tra fattori con pesi vincolati e con pesi liberi

riproducibilità dell'analisi fattoriale

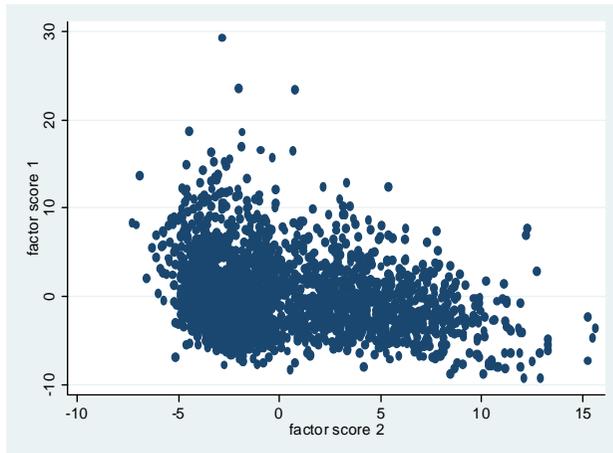


$$\text{Corr}(f1\text{free}, f1\text{expl}) = 0.9766$$

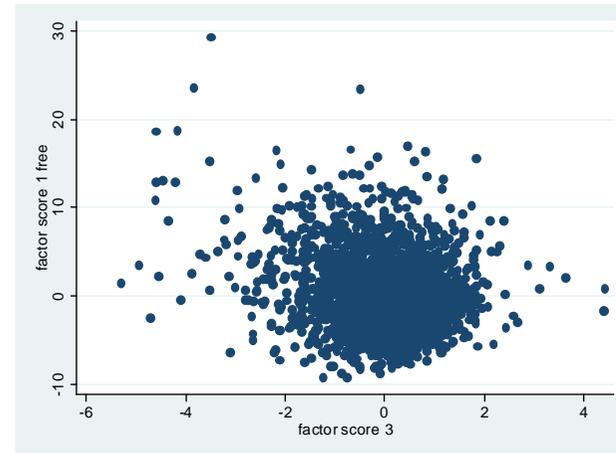


$$\text{Corr}(f2\text{free}, f2\text{expl}) = 0.7214$$

CORRELAZIONE TRA FATTORI

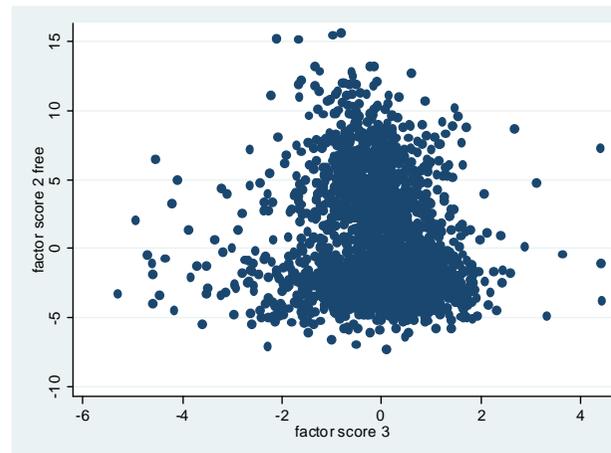


$$\text{Corr}(f1, f2) = -0.2263$$



$$\text{Corr}(f1, f3) = -0.1957$$

Se il modello è correttamente specificato ci aspettiamo ortogonalità dei fattori latenti



$$\text{Corr}(f2, f3) = -0.1451$$

ANALISI CONFERMATIVA sugli Indici sintetici

- come somma dei fattori latenti
 - factor score 1 free
 - factor score 2 free
 - factor 3 (famiglie monogenitore e stato civile)

- come somma dei fattori latenti vincolati
 - factor score 1 expl.
 - factor score 2 expl.
 - factor 3 (famiglie monogenitore e stato civile)

CONCORDANZA TRA INDICI (Classificazione in QUINTILI)

5 quantiles of indice free	5 quantiles of indice vincolato					Total
	1	2	3	4	5	
1	<u>351</u>	155	32	2	0	540
2	108	<u>200</u>	181	44	7	540
3	49	102	<u>184</u>	179	26	540
4	24	62	103	<u>207</u>	144	540
5	8	21	40	108	<u>362</u>	539
Total	540	540	540	540	539	2,699

Pearson's r = **0.749** Pr(r = 0) = 0.000 C_b = rho_c/r = 1.000

Agreement	Expected Agreement	Kappa	Std. Err.	Z	Prob>Z
48.31%	20.00%	0.3539	0.0096	36.77	0.0000

ANALISI CONFERMATIVA: confronto con l'indice nazionale

- come somma di punti Z
 - Disoccupazione
 - Indice di affollamento
 - Istruzione
 - Abitazioni in affitto
 - Famiglie monogenitore

CONCORDANZA TRA INDICI (Classificazione in QUINTILI)

5 quantiles of indice free	quintili indice nazionale					Total
	0	1	2	3	4	
1	<u>240</u>	149	87	46	18	540
2	144	<u>138</u>	125	114	19	540
3	76	129	<u>142</u>	113	80	540
4	53	83	111	<u>147</u>	146	540
5	27	40	75	120	<u>277</u>	539
Total	540	539	540	540	540	2,699

Pearson's r = **0.515** Pr(r = 0) = 0.000 C_b = rho_c/r = 0.800

Agreement	Expected Agreement	Kappa	Std. Err.	Z	Prob>Z
34.98%	20.00%	0.1872	0.0096	19.45	0.0000

Si ricorda che alcune variabili che compongono l'indice nazionale hanno una unicità relativamente alta

FASE 3 : VALIDAZIONE PREDITTIVA DELL'INDICE

Questa fase ha l'obiettivo di valutare la validità correlazionale con la mortalità per tutte le cause.

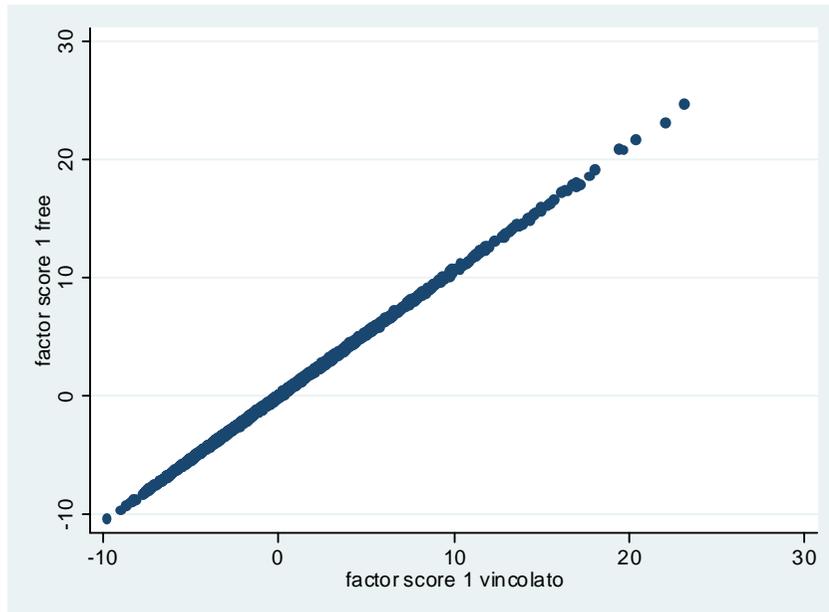
Indaga la predittività dell'indice rispetto a esiti sanitari che la letteratura associa alle dimensioni socio-economiche.

In pratica si misura la correlazione tra indice di deprivazione e il tasso grezzo di mortalità per tutte le cause (1999-2001).

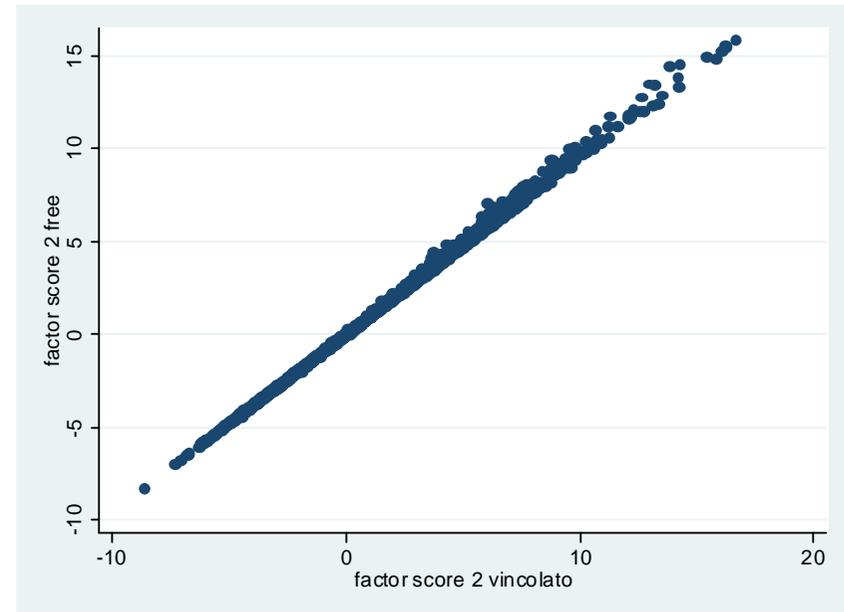
L'analisi è eseguita sul terzo campione casuale di Comuni italiani

Correlazione tra fattori con pesi vincolati e con pesi liberi
riproducibilità dell'analisi fattoriale

modello fattoriale confermativo del campione 2 su campione 3



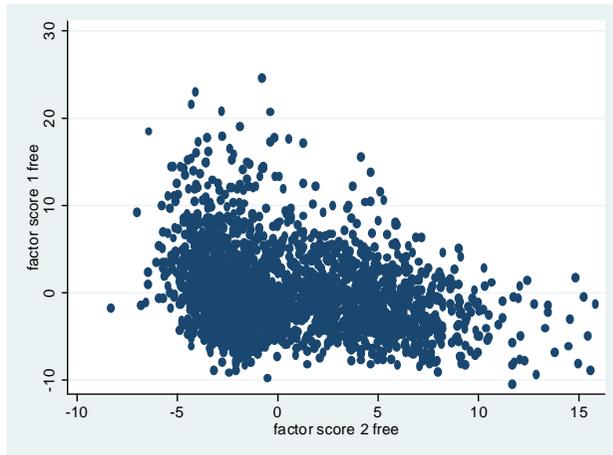
$$\text{Corr}(f1\text{free}, f1\text{vinc}) = 0.9999$$



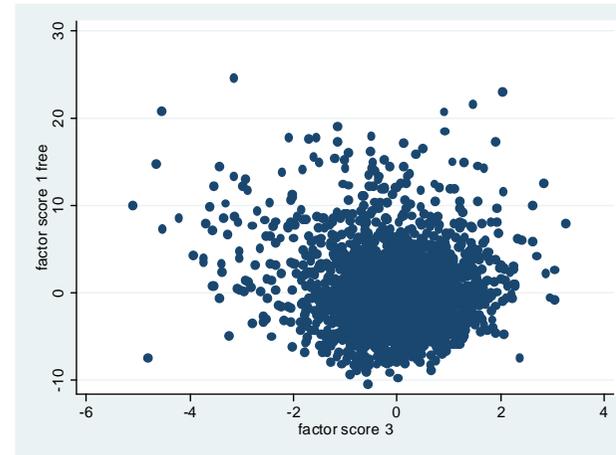
$$\text{Corr}(f2\text{free}, f2\text{vinc}) = 0.9996$$

CORRELAZIONE TRA FATTORI

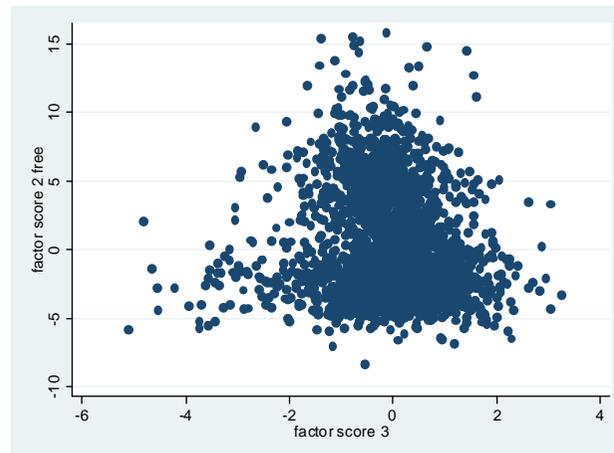
modello fattoriale confermativo del campione 2 su campione 3



$$\text{Corr}(f1, f2) = -0.2341$$



$$\text{Corr}(f1, f3) = -0.1124$$



$$\text{Corr}(f2, f3) = -0.1791$$

SMR mortalità per tutte le cause (maschi e femmine 1999-2001) per quintili degli indici di deprivazione

Indice di deprivazione			
Quintili	SMR	LI 90%	LS 90%
1	1.00	0.99	1.06
2	0.98	0.97	1.00
3	0.98	0.96	0.99
4	1.01	0.99	1.02
5	1.08	1.07	1.10

Indice nazionale			
Quintili	SMR	LI 90%	LS 90%
1	0.99	0.98	1.01
2	0.98	0.96	1.00
3	0.97	0.96	0.99
4	0.98	0.97	1.00
5	1.07	1.05	1.08

ALCUNI RIFERIMENTI BIBLIOGRAFICI

- Skrondal A, Rabe-Hesketh S. (2004) Generalized latent variable modeling. Multilevel, longitudinal, and structural equation models. *Chapman & Hall/CRC Interdisciplinary Statistics*.
- St Leger S (1995), Use of deprivation indices in small area studies of environment and health. *J Epidemiol Common Health*, **49** (Suppl 2).
- Morgan O, Baker A (2005), Measuring deprivation in England and Wales using 2001 Carstaris scores. *Health Statistics Quarterly*. 28-33.
- Challier B, Viel JF (2001), Relevance and validity of a new French composite index to measure poverty on a geographical level. *Rev Epidemiol Sante Publique*, **49**(1):41-50.
- Dominguez-Berjon MF, Borrell C, Benach J, Pasarín MI (2001), Measures of material deprivation in small area studies. *Gac Sani*, **15** (Suppl 4) 23-33.
- Dominguez-Berjon MF, Borrell C, Rodríguez-Sanz M, Pastor V (2005), The usefulness of area-based socioeconomic measures to monitor social inequalities in health in Southern Europe. *European Journal of Public Health*, **16**(1) 54-61.
- Eibner C, Sturm R (2006), US-based indices of area-level deprivation: Results from HealthCare for Communities. *Social Science & Medicine*, **62** 348-359.
- Fukuda Y, Nakamura K, Takano T (2007), Higher mortality in area of lower socioeconomic position measured by a single index of deprivation in Japan. *Public Health*, **121** 163-173.
- Biggeri A, Lagazio C, Catelan D, Pirastu R, Casson F, Terracini B (2006), Ambiente e salute nelle aree a rischio della Sardegna. *Epidemiologia & Prevenzione*, **30**(1) (Suppl. 1)

RINGRAZIAMENTI

Ricerca finalizzata Regione Valle d'Aosta – anno 2005

SLTo – piano 2008-2010

Università di Roma “La Sapienza” – anno 2007