

Analisi Fattoriale - Applicazioni

Analisi Esplorativa

Aldo Solari





Outline



Dati Esami

- Voto agli esami
- $n = 202$ studenti maschi
- $p = 6$

Variabili:

- Gaelic (non-math)
- English (non-math)
- History (non-math)
- Arithmetic (math)
- Algebra (math)
- Geometry (math)



Dati Esami: Correlazione

$$\mathbf{R} = \begin{array}{c} \begin{array}{cccccc} & \text{Gaelic} & \text{English} & \text{History} & \text{Arithmetic} & \text{Algebra} & \text{Geometry} \end{array} \\ \left[\begin{array}{cccccc} 1.0 & .439 & .410 & .288 & .329 & .248 \\ & 1.0 & .351 & .354 & .320 & .329 \\ & & 1.0 & .164 & .190 & .181 \\ & & & 1.0 & .595 & .470 \\ & & & & 1.0 & .464 \\ & & & & & 1.0 \end{array} \right] \end{array}$$



Stima di massima verosimiglianza

Assunzione aggiuntiva

La variabile aleatoria $x_{p \times 1}$ segue una distribuzione Normale p -variata

Funzione di log-verosimiglianza

$$\ell(\Sigma) = -\frac{1}{2}n \log |2\pi\Sigma| - \frac{1}{2}n \text{tr}(\Sigma^{-1}S)$$

dopo aver sostituito $\mu_{p \times 1}$ con $\bar{x}_{p \times 1}$

Stima di MV: Sostituisci Σ con $\Lambda\Lambda' + \Psi$ e massimizza $\ell(\Sigma)$ per Λ e Ψ

Stima iterativa

- ❶ Per Ψ fissato, massimizza numericamente per Λ
 - ❷ Per Λ fissato, massimizza numericamente per Ψ
- Implementata nella funzione R `factanal()`
 - Possiamo ottenere casi di Heywood



Dati Esami: FA con $k = 2$ e stima di MV

Table 9.5			
Variable	Estimated factor loadings		Communalities \hat{h}_i^2
	F_1	F_2	
1. Gaelic	.553	.429	.490
2. English	.568	.288	.406
3. History	.392	.450	.356
4. Arithmetic	.740	−.273	.623
5. Algebra	.724	−.211	.569
6. Geometry	.595	−.132	.372

- Stima di MV: $\hat{h}_1^2 = \hat{\lambda}_{11}^2 + \hat{\lambda}_{12}^2 = (0.553)^2 + (0.429)^2 \approx 0.490$
- Primo fattore: *intelligenza generale*
- Secondo fattore: *abilità matematica vs abilità verbale*



Rotazione dei pesi fattoriali

- Per la rotazione dei pesi fattoriali $\Lambda_{p \times k}$, dobbiamo cercare una matrice ortogonale $A_{k \times k}$ ($A'A = AA' = I$) tale per cui i pesi fattoriali ruotati $\Lambda_{p \times k}^* = \Lambda_{p \times k} A_{k \times k}$ sono più facilmente interpretabili
- $A_{2 \times 2} = \begin{bmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{bmatrix}$ rotazione oraria per $k = 2$
- Questo non cambia la soluzione del modello, solo la sua descrizione
- Situazione desiderata per i fini interpretativi:
 - i pesi fattoriali sono tutti grandi e positivi o prossimi a 0 (con pochi valori intermedi)
 - ogni variabile osservabile è legata in modo pesante al più ad un solo fattore
- Per $k > 2$ il metodo *varimax* identifica la rotazione massimizzando un'opportuna funzione dei pesi fattoriali ruotati che misura la variabilità dei pesi



Dati Esami: rotazione dei pesi fattoriali

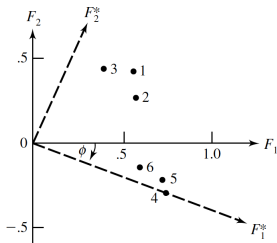


Figure 9.1 Factor rotation for test scores.

Table 9.6			
Variable	Estimated rotated factor loadings		Communalities $\hat{h}_i^{*2} = \hat{h}_i^2$
	F_1^*	F_2^*	
1. Gaelic	.369	.594	.490
2. English	.433	.467	.406
3. History	.211	.558	.356
4. Arithmetic	.789	.001	.623
5. Algebra	.752	.054	.568
6. Geometry	.604	.083	.372

- Primo fattore: *abilità matematica*
- Secondo fattore: *abilità verbale*



Outline



Table 8.4 Stock-Price Data (Weekly Rate Of Return)

Week	J P Morgan	Citibank	Wells Fargo	Royal Dutch Shell	Exxon Mobil
1	0.01303	-0.00784	-0.00319	-0.04477	0.00522
2	0.00849	0.01669	-0.00621	0.01196	0.01349
3	-0.01792	-0.00864	0.01004	0	-0.00614
4	0.02156	-0.00349	0.01744	-0.02859	-0.00695
5	0.01082	0.00372	-0.01013	0.02919	0.04098
6	0.01017	-0.01220	-0.00838	0.01371	0.00299
7	0.01113	0.02800	0.00807	0.03054	0.00323
8	0.04848	-0.00515	0.01825	0.00633	0.00768
9	-0.03449	-0.01380	-0.00805	-0.02990	-0.01081
10	-0.00466	0.02099	-0.00608	-0.02039	-0.01267
⋮	⋮	⋮	⋮	⋮	⋮
94	0.03732	0.03593	0.02528	0.05819	0.01697
95	0.02380	0.00311	-0.00688	0.01225	0.02817
96	0.02568	0.05253	0.04070	-0.03166	-0.01885
97	-0.00606	0.00863	0.00584	0.04456	0.03059
98	0.02174	0.02296	0.02920	0.00844	0.03193
99	0.00337	-0.01531	-0.02382	-0.00167	-0.01723
100	0.00336	0.00290	-0.00305	-0.00122	-0.00970
101	0.01701	0.00951	0.01820	-0.01618	-0.00756
102	0.01039	-0.00266	0.00443	-0.00248	-0.01645
103	-0.01279	-0.01437	-0.01874	-0.00498	-0.01637



Stock-Price Data

- Rendimento (settimanale) di cinque titoli
- Gen 04 - Dic 05
- $n = 103$
- $p = 5$

Variabili:

- JP Morgan (bank)
- Citibank (bank)
- Wells Fargo (bank)
- Royal Dutch Shell (oil)
- Exxon-Mobil (oil)



Stock-Price Data: correlazione

$$\bar{\mathbf{x}}' = [.0011, .0007, .0016, .0040, .0040]$$

$$\mathbf{R} = \begin{bmatrix} 1.000 & .632 & .511 & .115 & .155 \\ .632 & 1.000 & .574 & .322 & .213 \\ .511 & .574 & 1.000 & .183 & .146 \\ .115 & .322 & .183 & 1.000 & .683 \\ .155 & .213 & .146 & .683 & 1.000 \end{bmatrix}$$



Stock-Price Data: FA con $k = 2$ e stima di MV

Table 9.3						
Variable	Maximum likelihood			Principal components		
	Estimated factor loadings		Specific variances	Estimated factor loadings		Specific variances
	F_1	F_2	$\hat{\psi}_i = 1 - \hat{h}_i^2$	F_1	F_2	$\tilde{\psi}_i = 1 - \tilde{h}_i^2$
1. J P Morgan	.115	.755	.42	.732	-.437	.27
2. Citibank	.322	.788	.27	.831	-.280	.23
3. Wells Fargo	.182	.652	.54	.726	-.374	.33
4. Royal Dutch Shell	1.000	-.000	.00	.605	.694	.15
5. Texaco	.683	-.032	.53	.563	.719	.17
Cumulative proportion of total (standardized) sample variance explained	.323	.647		.487	.769	

- Stima di MV: $\hat{h}_1^2 = \hat{\lambda}_{11}^2 + \hat{\lambda}_{12}^2 = (0.115)^2 + (0.755)^2 \approx 0.58$
- Primo fattore: *mercato dei titoli*
- Secondo fattore: *bank vs oil*



Stock-Price Data: residui

Massima Verosimiglianza

$$\mathbf{R} - \hat{\mathbf{L}}\hat{\mathbf{L}}' - \hat{\Psi} = \begin{bmatrix} 0 & .001 & -.002 & .000 & .052 \\ .001 & 0 & .002 & .000 & -.033 \\ -.002 & .002 & 0 & .000 & .001 \\ .000 & .000 & .000 & 0 & .000 \\ .052 & -.033 & .001 & .000 & 0 \end{bmatrix}$$

Componenti principali

$$\mathbf{R} - \tilde{\mathbf{L}}\tilde{\mathbf{L}}' - \tilde{\Psi} = \begin{bmatrix} 0 & -.099 & -.185 & -.025 & .056 \\ -.099 & 0 & -.134 & .014 & -.054 \\ -.185 & -.134 & 0 & .003 & .006 \\ -.025 & .014 & .003 & 0 & -.156 \\ .056 & -.054 & .006 & -.156 & 0 \end{bmatrix}$$



Stima di MV: test sul numero di fattori

- Un vantaggio della stima di MV è che permette un test di ipotesi sul numero di fattori
- Ipotesi nulla: k fattori sono sufficienti
- Ipotesi alternativa: k fattori sono insufficienti
- Rifiuto l'ipotesi nulla con un p -value $< 5\%$
- Test sequenziali: parto da $k = 1$, se rifiuto proseguo con $k = 2, 3, \dots$ fino a quando fallisco di rifiutare l'ipotesi



Stock-Price Data: test di $k = 2$

Ipotesi nulla

$$H_0 : \Sigma = \Lambda\Lambda' + \Psi \quad (k = 2)$$

Statistica test

$$\frac{|\hat{\Lambda}\hat{\Lambda}' + \hat{\Psi}|}{S} = \frac{|\hat{\Lambda}_z\hat{\Lambda}'_z + \hat{\Psi}_z|}{R} = \frac{0.17898}{0.17519} = 1.0216$$

Richiede la stima di massima verosimiglianza

p -value

$$\mathbb{P}(\chi_1^2 > n \ln(1.0216)) \approx 0.138 > 5\%$$

(p -value = 0.15 utilizzando la correzione di Bartlett)



Stock-Price Data: rotazione dei pesi fattoriali

Table 9.8					
Variable	Maximum likelihood estimates of factor loadings		Rotated estimated factor loadings		Specific variances $\hat{\psi}_i^2 = 1 - \hat{h}_i^2$
	F_1	F_2	F_1^*	F_2^*	
J P Morgan	.115	.755	.763	.024	.42
Citibank	.322	.788	.821	.227	.27
Wells Fargo	.182	.652	.669	.104	.54
Royal Dutch Shell	1.000	-.000	.118	.993	.00
ExxonMobil	.683	.032	.113	.675	.53
Cumulative proportion of total sample variance explained	.323	.647	.346	.647	

- Primo fattore: *bank*
- Secondo fattore: *oil*



Stima dei punteggi fattoriali

Metodo di Bartlett (1937)

- $\hat{f}_i = (\hat{\Lambda}'\hat{\Psi}^{-1}\hat{\Lambda})^{-1}\hat{\Lambda}'\hat{\Psi}^{-1}x_i$

Metodo di Thompson (1951)

- $\hat{f}_i = \hat{\Lambda}'\hat{\Sigma}^{-1}x_i$



Stock-Price Data: punteggi fattoriali

Decomposizione di R

$$\hat{\mathbf{L}}_z^* = \begin{bmatrix} .763 & .024 \\ .821 & .227 \\ .669 & .104 \\ .118 & .993 \\ .113 & .675 \end{bmatrix} \quad \text{and} \quad \hat{\Psi}_z = \begin{bmatrix} .42 & 0 & 0 & 0 & 0 \\ 0 & .27 & 0 & 0 & 0 \\ 0 & 0 & .54 & 0 & 0 \\ 0 & 0 & 0 & .00 & 0 \\ 0 & 0 & 0 & 0 & .53 \end{bmatrix}$$

Una osservazione

$$\mathbf{z}' = [.50, -1.40, -.20, -.70, 1.40]$$

Metodo di Bartlett

$$\hat{\mathbf{f}} = (\hat{\mathbf{L}}_z^* \hat{\Psi}_z^{-1} \hat{\mathbf{L}}_z^*)^{-1} \hat{\mathbf{L}}_z^* \hat{\Psi}_z^{-1} \mathbf{z} = \begin{bmatrix} -.61 \\ -.61 \end{bmatrix}$$

Metodo di Thompson

$$\hat{\mathbf{f}} = \hat{\mathbf{L}}_z^* \mathbf{R}^{-1} \mathbf{z} = \begin{bmatrix} .331 & .526 & .221 & -.137 & .011 \\ -.040 & -.063 & -.026 & 1.023 & -.001 \end{bmatrix} \begin{bmatrix} .50 \\ -1.40 \\ -.20 \\ -.70 \\ 1.40 \end{bmatrix} = \begin{bmatrix} -.50 \\ -.64 \end{bmatrix}$$



Stock-Price Data: punteggi fattoriali

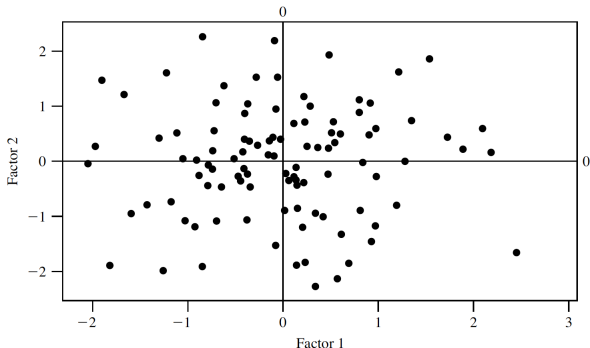


Figure 9.4 Factor scores using (9-58) for factors 1 and 2 of the stock-price data (maximum likelihood estimates of the factor loadings).

