

Metodi di stima

Metodi iniziali

Viene fatta una prima stima iniziale dei parametri a partire dai dati. I risultati vengono usati per delle ulteriori stime iterative. Gli errori sono legati alle equazioni di partenza.

- IV (Instrumental Variable)
- TSLS (Two Stages Least Squares)

Metodi iterativi

1. Massima verosimiglianza (ML, Maximum likelihood):

$$\ln |\Sigma| - \ln |S| + \text{tr}(S\Sigma^{-1}) - J$$

con J =numero di osservate. Stimatore efficiente. Questo metodo cerca di “sparpagliare” l'errore totale un po' su tutti i parametri, in modo da renderli tutti sempre più piccoli. Nel caso in cui le variabili si distribuiscano normalmente. Se il modello da analizzare “è corretto”, è il metodo migliore. La matrice Cov/Cor dev'essere positivamente definita.

2. Minimi quadrati ponderati (WLS, Weighted Least Square):

$$(\sigma_{jk} - s_{ik})w^{-1}(\sigma_{jk} - s_{ik})'$$

dove σ_{jk} è riprodotto e s_{ik} è osservato. La loro differenza è il vettore dei residui. E' il metodo generico che include tutti gli altri come casi particolari.

Si può usare con dati ordinali o con dati quantitativi se ci sono molte variabili non normali.

3. ULS, Unweighted least square:

$$\frac{1}{2}\text{tr}[(S - \Sigma)^2]$$

somma dei residui al quadrato (minimi quadrati non ponderati) della differenza tra var/cov osservate e quelle riprodotte dal modello. E' uno stimatore consistente e veloce da

calcolare, ma non è efficiente, perché non minimizza gli errori. Però funziona anche se la matrice non è positivamente definita.

4. GLS (Generalized least squares):

$$\frac{1}{2}tr[(I - S^{-1}\Sigma)^2]$$

E' uno stimatore efficiente, ma produce errori standard grandi.

5. DWLS (Diagonally Weighted Least Square)

Quando una matrice è singolare?

Quando i dati contengono valori mancanti e la matrice di varianza/covarianza è calcolata con il metodo *pairwise*. Oppure quando ho una matrice di correlazione policorica.

Valori mancanti, missing: informazioni non raccolte per vari motivi (il soggetto non risponde, dà una risposta confusa, non c'era quando avete fatto quel particolare questionario, non capisce la domanda o non c'è la risposta che va bene a lui).

Si trattano in due modi:

1. **Listwise:** L'unità statistica che ha anche un solo valore mancante fra le variabili da analizzare viene eliminato completamente (una matrice di correlazione avrà lo stesso N per ogni cella). Se ci sono molti *missing* la numerosità totale diminuisce notevolmente; *però* la matrice delle cov. incrocia dati relativi agli stessi soggetti.
2. **Pairwise:** Per ogni coppia di variabili, vengono eliminati le sole unità statistiche che hanno un valore mancante in una delle due variabili da analizzare (la matrice di correlazione potrà avere celle con N diversi). Se ci sono molti *missing* la numerosità totale non diminuisce di molto, perché ogni cov. è calcolata con tutti i valori utilizzabili. *Però* le celle della matrice di COV, contiene informazioni che non combaciano perfettamente essendo calcolate su gruppi di soggetti diversi. Questo porta, spesso, a trovare che la matrice è singolare.

Correlazione policorica: quando le variabili da correlare sono espresse su scale diverse dalle quantitative (intervallo, rapporto). Non posso usare Pearson e uso invece un misto di metodi di correlazione: biseriale, punto-biseriale, tetracorica, ... Dalla matrice di correlazione policorica, si può stimare “all’indietro” una matrice di varianze/covarianze, con un metodo iterativo: questa matrice viene chiamata *asintotica* perché le stime dovrebbero avvicinarsi “all’infinito” alle vere varianze/covarianze fra le variabili.

Le correlazioni policoriche sono generalmente più alte del normale. Il χ^2 è generalmente pari a 2,5 volte i gl e tende ad essere significativo.

Usare prevalentemente ML. Quando non devo usare ML?

- Se non ho dati normali (ordinali o continui con lunghe code). Meglio usare WLS
- Se la matrice risulta singolare (non positivamente definita). Usare ULS.