

Che cos'è l'ESCS e come si valuta

Daniele Checchi
(Università degli Studi di Milano e Anvur^{*})

Roma 1 marzo 2016

^{*} I pareri espressi sono del tutto personali e non implicano minimamente le istituzioni di appartenenza.

Cosa vogliamo misurare ?

⇒ i fattori che contribuiscono alla formazione delle competenze misurate che non siano imputabili alla scuola.

In questo modo è teoricamente possibile depurare il contributo della scuola dall'effetto del background familiare.

Definiamo

- y_{ij} il livello di apprendimento conseguito dallo studente i nella scuola j ,
- x_i una qualsiasi misura del background familiare (per esempio ESCS) dello stesso studente
- z_j una qualsiasi misura delle risorse scolastiche della scuola (per esempio il livello di formazione degli insegnanti)
- ε_{ij} la componente che non si riesce a prevedere

Se ipotizziamo che valga una relazione del tipo

$$y_{ij} = \alpha + \beta x_i + \delta z_j + \varepsilon_{ij}$$

ovvero ① che gli apprendimenti dipendano linearmente ed indipendentemente da ambiente familiare e risorse scolastiche

② che la componente erratica non contenga elementi sistematici (cioè $Cov[x, \varepsilon] = 0$ e $Cov[z, \varepsilon] = 0 \rightarrow$ ipotesi sistematicamente violata quando non si riesca a controllare per fattori quali abilità, intelligenza)

allora la statistica ci assicura una stima corretta dei parametri $\hat{\alpha}$, $\hat{\beta}$ e $\hat{\gamma}$, da cui otteniamo il livello di apprendimento previsto per lo studente

$$\hat{y}_{ij} = \hat{\alpha} + \hat{\beta}x_i + \hat{\gamma}z_j$$

con errore di previsione $y_{ij} - \hat{y}_{ij} = \hat{\varepsilon}_{ij}$, $E[\hat{\varepsilon}] = 0$

Con questo modello possiamo costruire esercizi contro-fattuali quali per esempio

⇒ quale sarebbe il livello di apprendimento dello studente i se il suo background fosse pari a quello medio di scuola \bar{x}_j

$$\tilde{y}_{ij} = y_{ij} + \hat{\beta}(\bar{x}_j - x_i)$$

⇒ il punteggio medio di una scuola al netto degli effetti ambientali creati da contesti locali \bar{x}_j diversi da quelli medi nazionali \bar{x}

$$\tilde{y}_j = \bar{y}_j + \hat{\beta}(\bar{x} - \bar{x}_j)$$

Per questo è cruciale

- ✓ una buona misura quantitativa del background individuale x_i
- ✓ una stima corretta dell'effetto del background β (*gradiente*).

Sfortunatamente:

① non abbiamo misure delle potenzialità degli alunni all'ingresso
⇒ non possiamo quindi convalidare l'ipotesi $Cov[x, \varepsilon] = 0$

② non abbiamo informazioni sulla potenziale autoselezione degli alunni nelle diverse scuole (*self-sorting*)
⇒ non possiamo quindi convalidare l'ipotesi $Cov[z, \varepsilon] = 0$

③ abbiamo misure diverse (alternative ?) del background
⇒ non sappiamo come combinare occupazione (reddito ? ricchezza ? tempo disponibile ?) e istruzione di entrambi i genitori

④ non siamo sicuri che la linearità sia una buona approssimazione
⇒ se fosse per esempio $y_{ij} = \alpha + \beta x_i + \delta z_j + \eta(x_i \times z_j) + \varepsilon_{ij}$

Come affrontare questi ostacoli:

① non abbiamo misure delle potenzialità degli alunni all'ingresso
⇒ non possiamo quindi convalidare l'ipotesi $Cov[x, \varepsilon] = 0$

Se avessimo una misura delle potenzialità iniziali dell'alunno a_i , il modello statistico appropriato sarebbe

$$y_{ij} = \alpha + \beta x_i + \delta z_j + \sigma a_i + \omega_{ij}$$

Fattibile:

Per alunni scuola primaria → test posizionamento oppure giudizio insegnanti

Per alunni secondaria 1 grado → punteggio 5 primaria (perdita 30% dei codici SIDI nel passaggio!)

Per alunni secondaria 2 grado → terza prova dell'esame di stato

② non abbiamo informazioni sulla potenziale autoselezione degli alunni nelle diverse scuole (*self-sorting*)
⇒ non possiamo quindi convalidare l'ipotesi $Cov[z, \varepsilon] = 0$

Sappiamo poco del perché le famiglie scelgono alcune scuole piuttosto che altre.

Se ci fossero bacini di utenza chiusi, il problema si attenuerebbe.

Se ci fosse allocazione randomizzata di chi chiede l'ammissione, il problema si attenuerebbe.

Alternativamente bisognerebbe modellare la scelta. A questo scopo servirebbero

→ misure delle informazioni o aspettative (customer ?)

→ misure del grado di convinzione (distanza geografica ?)

→ misure dell'eventuale orientamento (indicazioni insegnanti ?)

③ abbiamo misure diverse (alternative ?) del background
⇒ non sappiamo come combinare occupazione (reddito ? ricchezza ? tempo disponibile ?) e istruzione di entrambi i genitori

Importante identificare misura unica: per esperienza diretta, il numero di libri in casa (variabile ordinale 1→5) ha il massimo contributo esplicativo.

Il peso relativo delle componenti: non lascerei alla procedura statistica (analisi fattoriale) variabile annualmente e fisserei i pesi delle variabili normalizzate (stessa scala di misura).

Valore più elevato o valore medio nella coppia ?

Risorsa rilevante: il tempo genitoriale ? l'istruzione ? il reddito ? libri ?

Importante controllo qualità informazione sui genitori (obbligo di indicare all'atto iscrizione)

Elementi da aggiornare/includere: cittadinanza – genitori separati – famiglie non standard.

④ non siamo sicuri che la linearità sia una buona approssimazione

⇒ se fosse per esempio $y_{ij} = \alpha + \beta x_i + \delta z_j + \eta(x_i \times z_j) + \varepsilon_{ij}$

Risorse familiari e risorse scolastiche possono non essere indipendenti (studenti figli di laureati traggono maggior beneficio dalla qualità dei propri insegnanti → $\hat{\eta} > 0$ – studenti figli di genitori analfabeti traggono maggior beneficio da buoni insegnanti → $\hat{\eta} < 0$).

Questo rende più complicata la misurazione di un benchmark per le scuole, in quanto viene a dipendere da una misurazione della qualità delle risorse impiegate nella scuola → il confronto tra scuole dipende dalla distribuzione delle risorse alle stesse!