

# Iscrizioni in scuole private

Gianluca Covini

Maggio 2023

## 1 Introduzione

Nel seguente report ci occupiamo di indagare i fattori che determinano il numero di iscritti in scuole private. In particolare faremo riferimento al dataset "Private school enrolment and determinants" dal libro di Ramanathan, *Introduction to Econometrics* (dataset numero 4.16), presente tra i dati d'esempio di GRETL. L'obiettivo dell'analisi è studiare l'impatto di alcuni fattori rilevanti sulla percentuale di studenti iscritti in scuole private: in particolare, indagheremo l'influenza del reddito pro capite dello stato sulla percentuale di iscritti in scuole private per stato e della percentuale di spesa pubblica in istruzione, in modo da mettere in evidenza sia la situazione familiare dello studente che il ruolo dell'intervento pubblico.

## 2 Descrizione dei dati

Il dataset utilizzato riporta i dati riguardanti iscrizioni in scuole private e altri fattori rilevanti, raccolti per i 50 stati americani e per il distretto di Columbia; la maggior parte dei dati sull'istruzione provengono dal Digest of Education Statistics del 1996. In particolare il dataset distingue 8 categorie differenti di informazioni: la categoria `cathol` rappresenta la percentuale di popolazione cattolica per stato, `pupil`, il rapporto tra studenti e insegnanti per scuole pubbliche per stato, `white`, la percentuale di popolazione bianca per stato, `admexp`, la percentuale di spese da parte dell'amministrazione di ciascuno stato in istruzione, `revenue`, i fondi in istruzione per studente in migliaia di dollari, `memnea`, la percentuale di insegnanti delle scuole pubbliche che sono membri della National Education Association, `income`, il reddito pro capite in migliaia di dollari, `college`, la percentuale della popolazione dello stato che ha completato almeno 4 anni di college.

Come prima cosa, riportiamo nella tabella 1 alcune statistiche sui dati, ad esempio la media, la deviazione standard, il massimo e il minimo delle grandezze su cui ci concentreremo, cioè, in particolare, la percentuale di studenti iscritti in scuole private, il reddito pro capite per stato e la percentuale di spesa pubblica in istruzione da parte di ciascuno stato.

Variabile	Media	Deviazione standard	Max	Min
<code>enroll</code>	9,118%	3,972%	17,39%	1,860%
<code>income</code>	18,25	3,134	26,38	12,71
<code>admexp</code>	7,533%	1,232%	11,93%	5,400%

Tabella 1: Statistiche descrittive delle grandezze principali

Possiamo studiare, ora, il legame tra le variabili: abbiamo riportato nella figura 1 due grafici a dispersione, con già rappresentata, grazie a una funzionalità di GRETL, la retta dei minimi quadrati, in modo da visualizzare maggiormente la correlazione con `enroll` di `income` e `admexp`.

Inoltre, nella tabella 2, abbiamo riportato tutti gli indici di correlazione calcolati delle variabili con `enroll`.

<code>cathol</code>	<code>pupil</code>	<code>white</code>	<code>admexp</code>	<code>revenue</code>	<code>memnea</code>	<code>income</code>	<code>college</code>
0,5061	-0,2698	-0,3735	-0,0903	0,3254	-0,1730	0,5760	0,3188

Tabella 2: Indici di correlazione tra le grandezze ed `enroll`

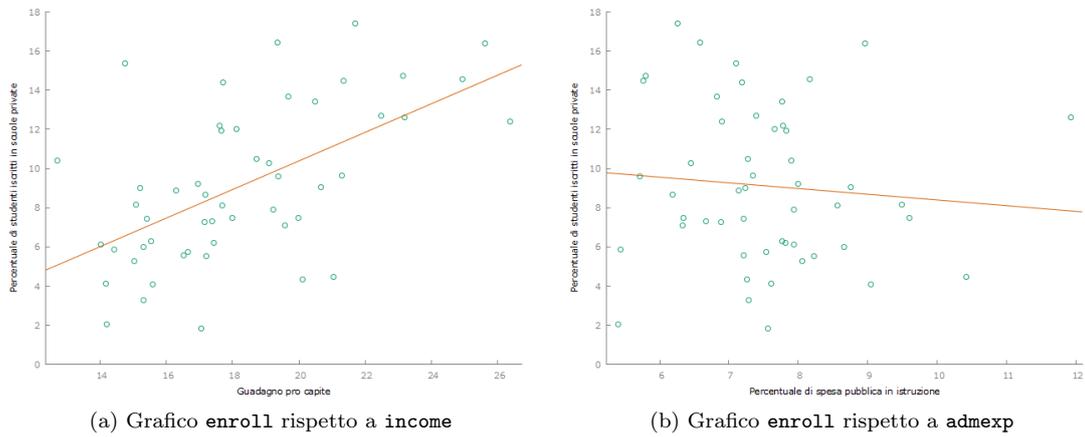


Figura 1: Grafici a dispersione con retta dei minimi quadrati

Notiamo che **income** possiede l'indice di correlazione con **enroll** più alto tra le variabili prese in considerazione; risulta, quindi, particolarmente interessante studiare il suo impatto su tramite un'analisi di regressione. Notiamo che, invece, l'altra variabile di nostro interesse, **admexp**, ha un indice di correlazione piuttosto basso, il che può suggerirci un minore impatto, che va, però, verificato tramite un'analisi di regressione.

### 3 Regressione lineare

Eseguiamo, ora, 4 regressioni lineari per studiare l'impatto di **income** e **admexp** su **enroll**. I risultati sono riportati nella tabella 3. In particolare, abbiamo inserito nella colonna (1) i risultati della regressione di **enroll** su **income**, nella colonna (2) la regressione su **admexp**, nella colonna (3) su entrambi i regressori, mentre nella colonna (4) introduciamo come variabili di controllo **cathol** e **white**. Per ogni riga riportiamo i valori del coefficiente, il suo errore standard e, per le variabili di interesse, anche l'intervallo di confidenza al 95%. Nelle ultime tre righe, infine, sono riportati l'errore standard della regressione, l' $R^2$  corretto e il numero di osservazioni.

Regressione	(1)	(2)	(3)	(4)
<b>income</b>	0,729 (0,139) [0,449; 1,010]		0,745 (0,145) [0,454; 1,036]	0,350 (0,198) [-0,050; 0,749]
<b>admexp</b>		-0,291 (0,513) [-1,322; 0,740]	-0,454 (0,348) [-1,154; 0,244]	-0,310 (0,308) [-0,931; 0,311]
<b>cathol</b>				0,127 (0,041)
<b>white</b>				-0,098 (0,030)
Intercetta	-4,206 (2,633)	11,31 (3,972)	-1,063 (3,829)	10,73 (5,286)
<b>Statistiche:</b>				
SER	3,280	3,996	3,265	2,898
$\bar{R}^2$	0,318	-0,012	0,324	0,468
<i>n</i>	51	51	51	51

Tabella 3: Risultati delle analisi di regressione

Possiamo riportare anche dei risultati per alcuni test eseguiti sulla regressione. I test sui coefficienti pari a zero con livello al 5% si possono eseguire naturalmente osservando l'appartenenza o meno di 0 all'intervallo di confidenza riportato nella tabella. Possiamo notare che, nelle regressioni riportate, l'ipotesi nulla viene rifiutata sempre nel caso del coefficiente di `income`, mentre non si può mai rifiutare al livello 5% per il coefficiente di `admexp`. Anche effettuando un test con un livello più alto, ad esempio del 10%, l'ipotesi nulla non può essere respinta per nessuna delle regressioni effettuate.

## 4 Commento

I risultati della regressione ci permettono di trarre importanti conclusioni per la nostra ricerca. Le prime osservazioni che possiamo fare riguardano i valori effettivi dei coefficienti: in particolare, nelle specificazioni (1) e (3) una variazione unitaria del reddito pro capite (corrispondente a una variazione di 1000 dollari) porta a un aumento di circa lo 0,7% del numero di studenti iscritti in scuole private. Possiamo interpretare questo dato alla luce del fatto che le scuole private in genere richiedono il pagamento di una retta più alta rispetto alla scuola pubblica, che può essere completamente gratuita. Tuttavia è importante notare che questo legame non implica necessariamente un rapporto di causalità tra le due cose: è naturale, però, ritenere, ad esempio, che le persone con redditi molto bassi si rivolgano quasi esclusivamente a scuole pubbliche, dove possono beneficiare di agevolazioni statali.

Un'osservazione interessante sui risultati della regressione è che aggiungendo le variabili di controllo nella colonna (4) il coefficiente di `income` si riduce sensibilmente, divenendo meno della metà. Questo fenomeno ci suggerisce che si aveva una distorsione da variabili omesse nelle regressioni (1) e (3) per via dell'assenza delle grandezze `cathol` e `white`. Possiamo verificare questo fatto studiando le correlazioni delle nuove variabili con `income` (rispettivamente 0,4871 e -0,2972), e notare che sono effettivamente significative, e osservando che queste due variabili hanno un effettivo impatto sulla variabile dipendente, tramite due test d'ipotesi sui coefficienti di `cathol` e `white`. Questi permettono, in entrambi i casi, di respingere l'ipotesi nulla all'1%.

Possiamo, poi, fare alcune considerazioni sui risultati relativi alla variabile `admexp`. Nelle regressioni effettuate, infatti, si nota che una variazione dell'1% della percentuale di spesa pubblica destinata all'istruzione porta a una diminuzione della percentuale di iscritti in scuole private che va dallo 0,3% allo 0,45% circa. I test d'ipotesi effettuati, tuttavia, non permettono di considerare rilevante con una probabilità sufficientemente alta l'impatto di tale variabile. Possiamo, quindi, affermare che un aumento della percentuale di spesa pubblica destinata all'istruzione non porta necessariamente a un aumento di iscritti nelle scuole pubbliche a scapito di quelle private. Ciò può essere dovuto a vari motivi riconducibili alla teoria economica: ad esempio, una spiegazione può risiedere nel fatto che l'aumento della spesa pubblica non è sufficiente se a questo non corrisponde un'allocatione efficiente delle risorse che permetta un miglioramento della qualità dell'educazione pubblica. Altre spiegazioni sono possibili: come, per esempio, il fatto che i genitori possono scegliere di mandare i figli in scuole private non basandosi sulla qualità dell'educazione ma per altri motivi. Il peso, nella regressione (4), della variabile `cathol`, per la quale il test d'ipotesi suggerisce un rifiuto dell'ipotesi nulla all'1%, suggerisce, ad esempio, che alcuni genitori possano scegliere di mandare i figli in scuole private di ispirazione cattolica per motivi primariamente religiosi.

Possiamo, infine, riportare alcune considerazioni sulle misure di bontà della regressione. È interessante notare che l' $R^2$  corretto aumenta con l'aumentare del numero di regressori, anche se l'aggiunta della sola variabile `admexp` non porta a un aumento significativo. Il valore dell' $R^2$  corretto, inoltre, è alto nella regressione (4), dimostrando che una parte significativa delle variazioni delle percentuali di iscrizioni in scuole private è spiegata tramite i regressori scelti, mentre è addirittura negativo nella regressione (2) che include solamente la variabile `admexp`, segnalando che l'aggiunta del regressore non spiega meglio le variazioni della variabile dipendente rispetto alla sola costante.

## 5 Conclusioni

Le conclusioni della nostra ricerca ci portano ad affermare che un ruolo rilevante nella scelta dell'iscrizione in scuole private è dato dalla situazione familiare dell'alunno: in particolare, abbiamo visto come il reddito e la situazione religiosa possono essere due variabili che influenzano tale scelta. D'altra parte, abbiamo anche osservato che l'intervento pubblico tramite un aumento della percentuale di spesa in istruzione, non ha un effetto sufficientemente rilevante dal punto di vista statistico.