

Un'indagine campionaria sugli studenti inattivi: il caso dell'Università di Pisa

Lucio Masserini

Osservatorio Statistico
Università di Pisa

l.masserini@adm.unipi.it

1. Introduzione

In questo lavoro si vuole soffermare l'attenzione sugli studenti inattivi, cioè gli studenti universitari che nella loro carriera hanno trascorso almeno un anno senza acquisire crediti formativi (o senza superare esami). L'interesse per questo argomento si spiega non solo per gli evidenti effetti negativi che la condizione di inattivo crea sugli studenti e sulle loro famiglie ma anche per gli oneri sul sistema universitario. Infatti, una quota sempre maggiore del Fondo di Finanziamento Ordinario (FFO) viene assegnata alle università tenendo conto dei risultati della carriera degli studenti. In particolare, il Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR), al fine di aumentare la qualità dell'offerta formativa e i risultati dei processi formativi, ha adottato due indicatori per la ripartizione del 34% (pari a 309 milioni di euro) della quota premiale di FFO nel 2012. Il primo indicatore (A1) è costruito a partire dagli studenti regolari nell'anno accademico 2010-'11 che hanno ottenuto almeno 5 crediti durante l'anno solare 2011; il secondo indicatore (A2) è basato invece sul rapporto tra i crediti acquisiti nell'anno solare 2011 e quelli previsti per gli studenti iscritti nell'anno accademico 2010-'11.

L'obiettivo di questo lavoro è quello di offrire un contributo alla comprensione dei fattori che favoriscono l'ingresso nella condizione di studente inattivo. I dati sono stati raccolti con un'indagine campionaria su un campione casuale stratificato di 1945 studenti iscritti all'Università di Pisa nell'anno accademico 2010-'11. La condizione di inattivo è stata considerata rispetto all'anno solare 2011. L'analisi è stata effettuata con un modello di regressione logistica a due livelli, rispettivamente per i corsi di laurea di primo e di secondo livello. L'indagine e l'elaborazione dei dati sono state realizzate dall'Osservatorio

Statistico dell'Università di Pisa. Dopo aver descritto il metodo di raccolta dei dati (Sezione 2), viene illustrato il modello statistico (Sezione 3) e, infine, sono presentati i risultati (Sezione 4) e le osservazioni conclusive (Sezione 5).

2. Dati

I dati sono stati raccolti mediante un'indagine campionaria realizzata con il metodo CATI (Computer Assisted Telephone Interviewing). Le unità campionarie sono state ottenute estraendo un campione casuale stratificato di 1945 studenti dalla popolazione dei 51.758 studenti iscritti all'Università di Pisa nell'anno accademico 2010-'11. La struttura del campione è stata definita a partire dalla condizione di inattività che identifica gli studenti con zero crediti formativi universitari (o esami superati) nell'anno solare 2011. I criteri di stratificazione sono stati: *status dello studente* (attivo, non attivo), *regolarità dell'iscrizione* (in corso, fuori corso da 1-2 anni, fuori corso da più di 2 anni), *status di matricola* (sì, no) ed *area disciplinare del corso di studio* (Sanitaria, Scienze e Matematica, Umanistica e Scienze economiche, giuridiche e sociali). Un numero quasi uguale di studenti attivi (968) ed inattivi (977) è stato ottenuto selezionando le unità campionarie con probabilità diseguali da ciascuno dei due diversi strati. La ripartizione degli studenti all'interno degli altri strati è invece proporzionale alla numerosità della popolazione. Le interviste sono state effettuate da un gruppo di studenti part-time, scelti tra coloro che avevano partecipato ad un periodo di formazione specifica. L'indagine è stata condotta dal 20 marzo al 5 maggio 2012. L'Osservatorio Statistico dell'Università di Pisa si è occupato di coordinare tutte le fasi dell'indagine.

3. Metodo

I dati mostrano una tipica struttura gerarchica (Snijders e Bosker, 2011), in cui le unità di livello inferiore (individui) sono comprese in unità di livello superiore (gruppi). Nel caso considerato, gli studenti sono raggruppati in corsi di studio che, a loro volta, sono compresi in facoltà e definiscono una struttura gerarchica a tre livelli. La variabile risposta è binaria e distingue gli studenti con crediti universitari in carriera dagli studenti senza crediti nell'anno solare 2011 tra gli iscritti all'Università di Pisa nell'anno accademico 2010-'11. Per considerare la struttura gerarchica dei dati e la natura della variabile risposta, l'analisi è stata eseguita stimando un modello di regressione logistica a due livelli con intercetta casuale (Goldstein, 2011). L'analisi a due livelli è stata preferita all'analisi a tre livelli poiché di solito il livello più basso di raggruppamento (corso di studio) ha un effetto maggiore sulle risposte e perché, in questo caso, è più importante ai fini di questa ricerca. Infatti, si ritiene che le caratteristiche dei corsi di studio riflettano sia le difficoltà dovute ai motivi di studio sia quelle relative all'organizzazione didattica e possono influenzare la probabilità di trovarsi nello stato inattivo più delle difficoltà dovute alle facoltà. Queste ultime sono di solito legate ad un più generale e meno definito effetto "ambientale", spesso limitato alla localizzazione e alla logistica delle aule e degli altri locali in cui si svolgono le attività didattiche.

Il modello è definito a partire dalla risposta binaria $y_{ij} \in (0,1)$, osservata sullo studente i , con $i = 1, 2, \dots, N$, appartenente al corso di studio j , con $j = 1, 2, \dots, G$. Questa assume il valore 0 se lo studente ha acquisito crediti nell'anno 2011 (studente attivo) ed il valore 1 in caso contrario (studente inattivo). Inoltre, si definisce la probabilità che y_{ij} assuma il valore 1, espressa come $P_{ij} = \Pr(y_{ij} = 1)$. Una trasformazione tipica per questa probabilità è la trasformazione logit, dove il logit è definito in termini di logaritmo naturale del rapporto di probabilità, indicato con $\ln(P_{ij}/(1-P_{ij}))$. Il modello di regressione logistica a due livelli con intercetta casuale può essere espresso come segue:

$$\ln\left(\frac{P_{ij}}{1-P_{ij}}\right) = \beta_0 + \beta_1 \mathbf{x}_{ij} + \gamma \mathbf{w}_j + u_j \quad (1)$$

dove \mathbf{x}_{ij} è un vettore di covariate riferite allo studente

i del corso di studio j , \mathbf{w}_j è un vettore di covariate che caratterizzano il corso di studio j , mentre $u_j \sim N(0, \sigma_u)$ è una quantità non osservabile condivisa dagli studenti di un particolare corso di studio e cattura i fattori non considerati dalle covariate del modello. Una questione a parte riguarda la stima in quanto la raccolta dei dati è stata concepita utilizzando un campione casuale stratificato, con proporzioni diverse tra gli strati di campionamento. In presenza di un disegno d'indagine complesso, due diversi approcci all'inferenza analitica sono il *design-based* e il *model-based* (Särndal, 1978). In un approccio *model-based*, i parametri del modello sono distorti solo se la distribuzione dei residui è influenzata dal disegno campionario. In questo caso il disegno si dice *informativo* (Pfeffermann, 1993) e il modello stimato sui dati campionari è diverso dal modello nella popolazione. Invece, se i residui del modello sono indipendenti dal disegno campionario e dai pesi di campionamento, allora il disegno è *non-informativo* e l'uso dei pesi campionari è irrilevante. In quest'ultimo caso, tenere conto del disegno di campionamento comporta una perdita di efficienza. Per comprendere se il disegno è informativo, sono state confrontate le stime ottenute con gli approcci *model-based* (non pesato) e *design-based* (pesato). Le stime *design-based* sono state ottenute con il metodo di pseudo-massima verosimiglianza proposto da Rabe-Hesketh e Skrondal (2006) per i modelli lineari generalizzati misti multilivello. I pesi sono definiti solo per le unità di primo livello (studenti) perché tutti i corsi di studio nella popolazione sono stati inclusi nel campione. Inoltre, i pesi sono inversamente proporzionali alla probabilità di inclusione e sono stati riscalati seguendo il "Metodo 2" in Pfeffermann *et al.* (1998):

$$w_i^* = w_i \frac{n_j}{\sum_j w_{i,j}} \quad (2)$$

Qui, $w_i = 1/\pi_i$, dove π_i definisce la probabilità che lo studente i sia incluso nel campione. Inoltre, n_j è il numero di studenti nel corso di studio j mentre $w_{i,j}$ è il peso riferito agli studenti del corso di studio j . Il confronto tra le stime *model-based* e *design-based* è stato effettuato con l'indice di informatività di Asparouhov (2006)¹.

¹ L'indice è dato dal rapporto tra la differenza delle stime *model-based* e *design-based* e la radice quadrata dell'errore standard della stima *model-based*.

4. Risultati

La percentuale di studenti inattivi nel 2011 rispetto agli iscritti all'Università di Pisa nell'anno accademico 2010-'11 è del 18,4 (9.506 su 51.758). Per gli studenti dei corsi di laurea del nuovo ordinamento (DM 509 e DM 270) tale percentuale è del 17,3 nei corsi di laurea triennale ma scende all'11,6 tra le matricole. Nei corsi di laurea specialistica/magistrale la percentuale è del 9,8.

L'analisi dei dati campionari è limitata agli studenti dei corsi di laurea del nuovo ordinamento ed è stata fatta stimando modelli separati per i corsi di primo e secondo livello. I risultati sono riportati in Tabella 1. Il confronto delle stime ottenute con gli approcci *model-based* e *design-based* porta a ritenere informativo il disegno di campionamento. Per questo, il commento dei risultati è riferito alle stime ricavate con il metodo di pseudo-massima verosimiglianza (Rabe-Hesketh e Skrondal, 2006), ottenuto utilizzando i pesi di campionamento.

4.1. Corsi di laurea di primo livello

L'analisi riguarda 1.382 studenti compresi in 139 corsi di studio. Il numero medio di studenti per corso è di 9,9 (min=5 e max=74). Il modello nullo mostra una deviazione standard a livello di corso significativa ($\sigma_u=0.565$, p-value<0.0001) e un valore piuttosto elevato del coefficiente di correlazione intra-classe ($\rho=0,088$) che indica una associazione significativa nelle risposte all'interno dei corsi di studio. Dopo aver introdotto le covariate a livello degli studenti, il modello presenta una deviazione standard di secondo livello inferiore ma ancora significativa ($\sigma_u=0,360$, p-value=0,0116) e, di conseguenza, un valore ridotto del coefficiente di correlazione intra-classe ($\rho=0,038$). Per questo modello, il valore della costante definisce un profilo di riferimento corrispondente ad uno studente ipotetico avente un valore nullo di tutte le variabili esplicative. Questo profilo rappresenta uno studente *iscritto allo stesso corso di immatricolazione, in corso o fuori corso da 1-2 anni, che non ha ripetuto anni durante la scuola superiore, che non ha un diploma liceale, che ha un voto di scuola superiore minore di 90, che frequenta le lezioni più o meno regolarmente, che ha rapporti con gli altri studenti anche al di fuori dell'università, che non lavora e non ha figli*. Per questo profilo anche l'effetto di corso è nullo. La probabilità di trovarsi nella condizione di inattivo

per uno studente avente il profilo di riferimento è piuttosto bassa (0,055). Tale probabilità aumenta in modo significativo soprattutto per i fuori corso da più di 2 anni (0,202) e per gli studenti che lavorano (0,155). Per quanto riguarda le variabili che caratterizzano l'esperienza universitaria, la probabilità di trovarsi nello stato inattivo è più elevata per i non frequentanti (0,104), per gli iscritti ad un corso diverso da quello di immatricolazione (0,097) e per chi non ha contatti con i compagni al di fuori dell'università (0,080). Infine, circa le variabili che caratterizzano l'esperienza prima di iscriversi all'università, la probabilità di essere inattivo è più alta per chi ripete anni durante la scuola superiore (0,078) e per chi ha un voto di maturità inferiore a 90 (0,088) ma è più bassa per gli studenti che hanno un diploma liceale (0,039).

4.2. Corsi di laurea di secondo livello

L'analisi riguarda 371 studenti compresi in 113 corsi di studio. Il numero medio di studenti per corso è 3,3 (min=2 e max=26). Il modello nullo mostra una deviazione standard a livello di corso non significativa ($\sigma_u=0.001$, p-value=1,000) e vuol dire che non esiste un effetto di corso. Per questo modello, il valore della costante definisce un profilo di riferimento corrispondente ad uno studente che *frequenta le lezioni più o meno regolarmente, che non lavora, non ha figli, è in corso o fuori corso da 1-2 anni e che non ha deciso di iscriversi per le opportunità di lavoro*. La probabilità di trovarsi nello stato inattivo associata a questo profilo è molto bassa (0,044). Rispetto a questo, la probabilità è più elevata per gli studenti fuori corso da più di 2 anni (0,235), per i non frequentanti (0,115), per chi lavora (0,091) e per coloro che hanno figli (0,077). Al contrario, questa probabilità è più bassa per gli studenti che si iscrivono spinti dalle future opportunità di lavoro (0,025).

Tabella 1: Stime dei parametri

Parametri	Corsi di primo livello				Corsi di secondo livello			
	Coeff.	SE	P	Prob	Coeff.	SE	P	Prob
Effetti fissi								
Costante (profilo di riferimento)	-2.839	0.211	0.000	0.055	-3.075	0.230	0.000	0.044
Iscritto ad corso diverso da quello di immatricolazione	0.609	0.163	0.000	0.097	-	-	-	-
Ha ripetuto anni durante le superiori	0.376	0.208	0.071	0.078	-	-	-	-
Frequenta le lezioni più o meno regolarmente	0.686	0.238	0.004	0.104	1.036	0.274	0.002	0.115
Rapporti con studenti anche fuori dell'università	0.399	0.200	0.047	0.080	-	-	-	-
Lavora	1.146	0.164	0.000	0.155	0.780	0.274	0.004	0.091
Ha figli	0.349	0.184	0.058	0.076	0.594	0.393	0.031	0.077
Voto di scuola superiore maggiore di 90	0.498	0.196	0.011	0.088	-	-	-	-
Diploma liceale	-0.366	0.168	0.029	0.039	-	-	-	-
Fuori corso da più di 2 anni	1.466	0.199	0.000	0.202	1.897	0.283	0.000	0.235
Iscrizione motivata dalle opportunità di lavoro	-	-	-	-	-0.549	0.283	0.052	0.025
Effetti casuali (deviazione standard a livello di corso)	0.360	0.100	0.012	-	-	-	-	-

5 Osservazioni conclusive

Il fenomeno degli studenti universitari inattivi è stato poco indagato finora e la letteratura su questo argomento è scarsa. Questo lavoro cerca di fornire un contributo alla comprensione delle motivazioni che possono favorire periodi di inattività durante la carriera di uno studente universitario. Questo problema riguarda ogni anno una percentuale piuttosto significativa di studenti ed è rilevante non solo per gli effetti negativi sugli studenti stessi e sulle loro famiglie ma anche perché una quota sempre maggiore di FFO viene attribuita alle università tenendo conto di indicatori relativi alla carriera degli studenti.

I risultati dell'indagine campionaria svolta dall'Osservatorio Statistico dell'Università di Pisa mostrano una notevole differenza nella probabilità di trovarsi nella condizione di inattivo tra i corsi di primo e di secondo livello. Un effetto di corso si osserva solo nei corsi di laurea triennale. Inoltre, il percorso scolastico che precede l'università svolge un ruolo fondamentale per gli studenti del primo livello ma è irrilevante per gli studenti dei corsi di laurea specialistica/magistrale. Per i primi è più importante avere contatti con i compagni di studio anche fuori dall'università mentre per i secondi sono più importanti i motivi dell'iscrizione che, se collegati alle opportunità di lavoro, sono associati ad una più bassa probabilità di essere inattivi. In comune tra primo e secondo livello ci sono l'esperienza universitaria (frequentare le lezioni ed essere in corso o fuori corso da 1-2 anni) con un effetto positivo e avere un lavoro, con un effetto negativo. In generale, questi risultati non sono in contrasto con la possibilità che una parte significativa di studenti inattivi siano studenti-lavoratori. Questi in genere non hanno la possibilità

di frequentare le lezioni, non hanno rapporti con gli altri studenti, sono iscritti da molto tempo e sono distanti dal mondo universitario. Per loro, il sistema universitario attuale, le cui attività (lezioni, counseling e tutoraggio) sono svolte soprattutto al mattino o nel primo pomeriggio, potrebbe essere difficilmente accessibile. Un modo per provare a controllare o anche ridurre la quota di studenti inattivi potrebbe essere quello di ampliare l'offerta dei corsi serali. Tuttavia vi è spazio anche per il miglioramento dei servizi di consulenza e di orientamento, ex ante e in itinere.

Riferimenti bibliografici

- Asparouhov, T. (2006), General Multilevel modeling with Sampling Weights, *Communications in statistics. Theory and Methods*, 35, 439-460.
- Goldstein H. (2011), *Multilevel Statistical Models*, Wiley, New York.
- Pfeffermann, D. (1993), The role of sampling weights when modeling survey data, *International Statistical Review*, 61, 317-337.
- Pfeffermann, D., Skinner, C. I., Holmes, D. I., Goldstein, H., and Rasbash, I. (1998), Weighting for unequal selection probabilities in multi-level models, *Journal of the Royal Statistical Society, Series B*, 60, 23-56.
- Rabe-Hesketh, S., and Skrondal, A.. Multilevel modelling of complex survey data, *Journal of the Royal Statistical Society, Series A*, 169, 805-827 (2006).
- Särndal, C.E. Design-based and model-based inference in survey sampling. *Scandinavian Journal of Statistics* 5, 27-52 (1978).
- Snijders, T.A. & Bosker, R.. *Multilevel analysis. An introduction to basic and advanced multilevel modelling. 2nd edition*, Thousand Oaks, CA: Sage (2011).