

GLI APPROFONDIMENTI DEL SERVIZIO STATISTICO

La newsletter ufficiale del Servizio Statistico INVALSI



Gentilissimo/a,

in questo numero della newsletter *“Gli approfondimenti del Servizio Statistico”* parleremo dei paragrafi, inseriti nel Rapporto annuale INVALSI (<https://serviziostatistico.invalsi.it/wp-content/uploads/2025/07/Rapporto-prove-INVALSI-2025.pdf>), dal titolo suggestivo *“Cosa incide sui risultati?”*.

In essi vengono analizzate le differenze negli esiti delle prove di Italiano, Matematica ed Inglese in base alle seguenti caratteristiche degli studenti:

- genere;
- regolarità nel percorso di studi (studenti in pari o in ritardo di almeno un anno[1]);
- macro-indirizzo di studio (licei, istituti tecnici, istituti professionali);
- *background* socio-economico-culturale, misurato da un indicatore sintetico (ESCS) basato su condizione professionale e titolo di studio dei genitori e sulla disponibilità di risorse per lo studio (ad esempio, la possibilità di utilizzare un pc);
- *background* migratorio (Italiani, stranieri di prima e seconda generazione);
- area geografica di residenza (Nord Est, Nord Ovest, Centro, Sud e Sud-Isole).



Per dare l'idea di quanto ciascun aspetto incide sull'esito delle prove, si fa riferimento al risultato medio di uno studente tipo con le seguenti caratteristiche: maschio, in regola con gli studi, che frequenta un istituto tecnico del Centro Italia e di condizione socio-economica-culturale (ESCS) pari alla media nazionale. Questo profilo è convenzionale e serve unicamente per disporre di un termine di paragone.

Si stima quindi la differenza fra il risultato medio di questo studente tipo e quello di un altro che differisce soltanto per una delle caratteristiche considerate. Ad esempio, uno studente in tutto identico al profilo di riferimento ma iscritto a un liceo anziché a un istituto tecnico.

Il confronto viene ripetuto per ogni caratteristica – genere, regolarità ecc. – tenendo fisse tutte le altre. Le differenze per uno stesso grado scolastico e per una stessa materia sono mostrate sinotticamente in un unico grafico ove gli scostamenti positivi e negativi si dipartono da una linea orizzontale (in corrispondenza dello 0) che rappresenta lo studente tipo.

Ad esempio (grafico 1), nelle prove di italiano del secondo anno della scuola secondaria di secondo grado, a parità di altre caratteristiche, chi frequenta un liceo "tradizionale" (classico, scientifico o linguistico) ottiene in media circa **+21** punti rispetto al profilo tipo.

[1] Parleremo in questo caso di studenti "posticipatari".

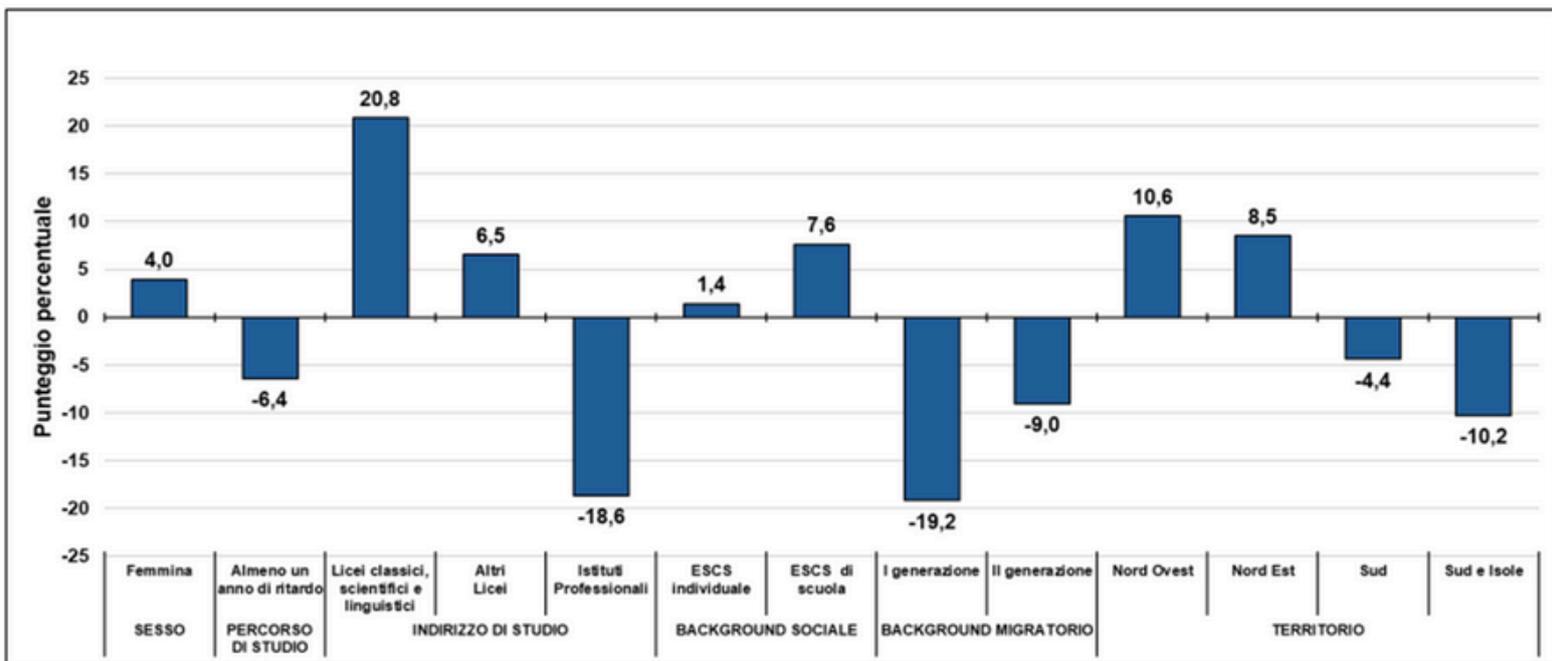


Grafico 1 – Esiti nella prova di Italiano in prima secondaria di secondo grado. Differenze rispetto al profilo tipo (fonte: Rapporto INVALSI 2025).

Guardando il grafico, si potrebbe pensare che tali differenze si sommino, creando divari abissali tra profili “favorevoli” e “sfavorevoli”. Come esempio di profilo favorevole, sulla base dei risultati medi, si pensi a una femmina italiana proveniente da una famiglia di livello socio-economico-culturale elevato, che frequenti un liceo classico del Nord Ovest; al polo opposto, si consideri un maschio straniero di estrazione socio-economico-culturale modesta, iscritto a un istituto professionale del Sud e Isole.

In realtà, la questione è un po’ più complessa. In molti casi, infatti, le caratteristiche interagiscono fra di loro, cosicché l’effetto dell’una può dipendere dal livello dell’altra. A volte ciò contribuisce ad amplificare i divari, altre volte ad attenuarli.

Questa newsletter inaugura una serie di approfondimenti dedicati proprio alle interazioni tra variabili. Analizzeremo cioè non solo gli effetti singoli, ma anche come essi cambino a seconda delle altre caratteristiche.

La prima variabile che prenderemo in esame è la **condizione di regolarità o ritardo** negli studi. Le domande che ci porremo sono del seguente tenore:

- la penalizzazione nei risultati delle prove INVALSI che, guardando al grafico, parrebbe interessare gli studenti posticipatari è la stessa in tutti gli indirizzi scolastici?
- nelle scuole del Nord come in quelle del Centro e del Sud?
- fra studenti italiani e studenti stranieri di prima o di seconda generazione?

Rispondere a questi quesiti è essenziale per comprendere le disuguaglianze della scuola italiana e per progettare politiche mirate.

GLI STUDENTI POSTICIPATARI

Nella terminologia qui adottata,[2] essere uno studente posticipatario significa frequentare un grado scolastico precedente a quello atteso in base all'età.

È il caso, ad esempio, di un quindicenne iscritto al primo anno di un liceo o di un diciottenne che frequenta il terzo anno di un istituto professionale. Non sempre il ritardo è dovuto alla mancata ammissione all'anno scolastico successivo a seguito di una valutazione negativa di profitto: può dipendere da problemi di salute che hanno condotto all'interruzione della frequenza o da percorsi migratori. La ripetizione è tuttavia la causa più frequente: essa interessa circa il 90% degli studenti posticipatari che hanno sostenuto le prove INVALSI dell'ultimo anno della scuola secondaria di secondo grado (13° grado) nel 2025.[3]

Gli studenti posticipatari sperimentano dunque una condizione di disallineamento anagrafico, spesso associata ad altre fragilità: la ripetenza di per sé, difficoltà linguistiche e culturali, problemi familiari o di salute.

Per questo motivo è importante misurare il divario di competenze di questi studenti rispetto ai loro colleghi "regolari". I dati INVALSI consentono di analizzare questo aspetto in relazione alle competenze di base (italiano, matematica, inglese), che pur non esaurendo la totalità degli apprendimenti sono fortemente correlate allo sviluppo e al successo personali.

[2] Si utilizza la terminologia del Rapporto INVALSI.

[3] Gli studenti posticipatari sono circa il 16% del totale degli studenti che hanno sostenuto le prove INVALSI di 13° grado nel 2025.

MODELLI DI ANALISI

Per valutare le interazioni fra condizione di posticipatario e altre caratteristiche socio-anagrafiche (genere, indirizzo di studio, area di residenza, background migratorio, background socio-economico-culturale) usiamo la regressione lineare.

Si tratta di un modello statistico-matematico in cui il punteggio atteso in una prova dipende da una combinazione di predittori (genere, indirizzo di studi, area geografica ecc.) e di coefficienti (parametri) che stimano il peso di ciascuno di essi.

In forma semplificata:

$$\text{punteggio}_i = b_0 + b_1 * \text{post}_i + b_2 * \text{genere}_i + b_3 * \text{tipo}_i + b_4 * \text{escs}_i + b_5 * \text{background_mig}_i + b_6 * \text{area}_i$$

dove:

- punteggio_i è il risultato nella prova (ad es. Italiano) per l'i-esimo studente;
- ripete_i è la variabile che misura la condizione di regolarità/ritardo per l'i-esimo studente;
- genere_i , escs_i , background_mig_i e area_i sono i valori degli altri predittori per l'i-esimo studente;
- b_0 , b_1 ecc. sono i parametri del modello.[4]

MISURE DI ASSOCIAZIONE

Dal modello possiamo calcolare varie misure di associazione fra punteggio nelle prove INVALSI e condizione di regolarità/posticipazione. Per i nostri scopi facciamo riferimento a una quantità chiamata effetto medio della variabile di interesse (in inglese *Average Treatment Effect*, abbreviato in ATE). Per ottenere l'ATE si confrontano due "cloni" della popolazione empirica costituita dagli studenti che hanno effettivamente sostenuto le prove: una popolazione-clone in cui la variabile che misura la regolarità degli studi è fissata a "posticipario" per tutti gli studenti (anche quelli che sono in realtà regolari) e una costituita invece da tutti "regolari". La differenza fra i punteggi medi delle due popolazioni fornisce la stima dell'effetto.

Questo procedimento analitico riproduce la logica dell'esperimento, in cui una variabile (detta "trattamento", in inglese *treatment*) viene manipolata dall'osservatore, mentre tutte le altre caratteristiche sono tenute fisse o sotto controllo. Ovviamente noi non stiamo conducendo un esperimento, pertanto l'ATE non può essere interpretato in senso causale. Piuttosto esso va inteso come una misura dell'associazione fra condizione di regolarità/ritardo degli studi e risultato nelle prove, nel senso di fornire una stima della differenza media fra chi è in pari e chi ha accumulato almeno un anno di ritardo: tanto maggiore è il divario e tanto più forte è l'associazione fra percorso di studi e punteggio, e viceversa.

Nel caso degli studenti che hanno sostenuto le prove INVALSI di 13° grado nel 2025, tali differenze o "contrast" sono pari, rispettivamente, a -10,4 punti per italiano e -9,4 punti per matematica.

In parole molto semplici, uno studente posticipatario ottiene in media circa 10 punti in meno di uno regolare.

[4] I parametri sono gli stessi per tutti gli studenti, dal momento che (come detto) stabiliscono il peso di ciascun carattere nella determinazione del risultato atteso.

STIME DI INTERAZIONI

Fin qui abbiamo seguito la logica applicata nel rapporto, che consiste nel guardare alle differenze di risultati considerando una variabile alla volta. Questo approccio è utile per una prima descrizione e valutazione del fenomeno. Esso, tuttavia, non è in grado di dare conto dell'eventualità che esistano interazioni fra variabili – ad esempio, rispetto al fatto che la penalizzazione empirica degli studenti posticipatari sia più alta in certe tipologie di scuole anziché che in altre; o in determinate aree del Paese anziché che in altre.

Per poter esplorare questi aspetti dobbiamo arricchire il modello di regressione, stimando nuovi parametri che esprimano l'interazione fra condizione di regolarità negli studi e altre caratteristiche socio-anagrafiche:

$$\text{punteggio}_i = b_0 + b_1 \cdot \text{post}_i + b_2 \cdot \text{genere}_i + b_3 \cdot \text{tipo}_i + b_4 \cdot \text{escs}_i + b_5 \cdot \text{background_mig}_i + b_6 \cdot \text{area}_i + b_7 \cdot \text{post}_i \cdot \text{tipo}_i + b_8 \cdot \text{post}_i \cdot \text{area}_i + b_9 \cdot \text{post}_i \cdot \text{background_mig}_i$$

dove b_7 , b_8 e b_9 sono nuovi parametri che colgono un'eventuale interazione fra condizione di posticiparietà e, rispettivamente, tipo di scuola, background migratorio e area di residenza.

Con questo nuovo modello possiamo calcolare una nuova quantità chiamata CATE (*Conditional Treatment Average Effect*), cioè l'effetto medio del ritardo condizionato al valore di un'altra variabile.

Di seguito, riportiamo le stime CATE dei divari fra studenti posticipatari e studenti regolari per differenti valori delle variabili socio-anagrafiche (tabella 1):

	Italiano			Matematica		
	Stima	Intervallo conf.za 95%		Stima	Intervallo conf.za 95%	
		Inferiore	Superiore		Inferiore	Superiore
Licei	-20,3	-21,0	-19,5	-18,6	-19,4	-17,8
Altri licei	-8,5	-9,3	-7,6	-13,1	-13,7	-12,4
Istituti Professionali	-3,4	-4,0	-2,7	-2,6	-3,1	-2,0
Istituti Tecnici	-11,3	-11,8	10,8	-10,6	-11,0	-10,2
	Stima	Intervallo conf.za 95%		Stima	Intervallo conf.za 95%	
		Inferiore	Superiore		Inferiore	Superiore
Nord Ovest	-11,8	-12,5	-11,1	-11,5	-12,0	-10,9
Nord Est	-13,3	-14,0	-12,5	-11,4	-12,1	-10,8
Centro	-13,7	-14,5	-12,9	-12,0	-12,7	-11,3
Sud	-17,1	-17,8	-16,5	-15,6	-16,2	-15,0
Sud e Isole	-12,2	-13,1	-11,3	-11,4	-12,2	-10,6
	Stima	Intervallo conf.za 95%		Stima	Intervallo conf.za 95%	
		Inferiore	Superiore		Inferiore	Superiore
Italiano	-14,4	-14,7	-13,8	-13,2	-13,6	-12,9
Straniero prima gen.	-9,1	-10,4	-7,8	-4,9	-6,1	-3,7
Straniero sec. gen.	-9,8	-10,7	-8,8	-6,5	-7,4	-5,7

Tabella 1 – Stime contrasti fra punteggi medi di studenti posticipatari e regolari che hanno sostenuto le prove INVALSI di 13° grado nel 2025 (fonte: nostra elaborazione su dati INVALSI).

I risultati mostrano che le differenze stimate fra studenti posticipatari e studenti regolari hanno sempre segno negativo – in nessuno dei casi ci si può attendere che i primi abbiano colmato il divario che li separa dai secondi; tuttavia, tali differenze variano tantissimo a seconda dell'indirizzo di studi. Ad esempio, nei licei classici e scientifici, il divario atteso è dell'ordine di ca. -20 punti per italiano e di ca. -19 punti per matematica; da notare che parliamo di stime (stiamo utilizzando un modello statistico) assolutamente robuste. Ad esempio, l'intervallo di confidenza al 95% del divario fra posticipatari e regolari in italiano va da -19 punti e mezzo a -21 punti.

CHE COSA SIGNIFICA TUTTO QUESTO?

In nessun caso gli studenti posticipatari colmano il loro svantaggio con i regolari. Tuttavia, l'entità del divario varia molto a seconda del contesto.

Una possibile interpretazione è che nei licei, dove il livello medio è più alto, gli studenti in ritardo faticano maggiormente a recuperare. Nei professionali, al contrario, le differenze appaiono più contenute perché il livello generale è meno elevato.

I dati, però, non offrono risposte definitive: servono per porre domande nuove e guidare ulteriori analisi. È in questo spazio che possiamo far crescere la conoscenza e sviluppare politiche più efficaci. Sarebbe ed è importante che da queste analisi partissero riflessioni metodologiche e pedagogiche che provino a fornire spiegazioni, da riscontrare empiricamente, delle importanti differenze messe in luce attraverso gli esempi forniti. In questo senso si realizza il mutuo vantaggio delle applicazioni di metodi quantitativi e qualitativi, da intendersi come complementari e mai alternativi.

A questo [link](#) è possibile recuperare i precedenti numeri della newsletter.