

Piero Demetrio Falorsi,
Patrizia Falzetti, Roberto Ricci

**LE METODOLOGIE
DI CAMPIONAMENTO
E SCOMPOSIZIONE
DELLA DEVIANZA NELLE
RILEVAZIONI NAZIONALI
DELL'INVALSI**

LE RILEVAZIONI DEGLI APPRENDIMENTI
A.S. 2018-2019



FrancoAngeli
OPEN  ACCESS





Il presente volume è pubblicato in open access, ossia il file dell'intero lavoro è liberamente scaricabile dalla piattaforma **FrancoAngeli Open Access** (<http://bit.ly/francoangeli-oa>).

FrancoAngeli Open Access è la piattaforma per pubblicare articoli e monografie, rispettando gli standard etici e qualitativi e la messa a disposizione dei contenuti ad accesso aperto. Oltre a garantire il deposito nei maggiori archivi e repository internazionali OA, la sua integrazione con tutto il ricco catalogo di riviste e collane FrancoAngeli massimizza la visibilità, favorisce facilità di ricerca per l'utente e possibilità di impatto per l'autore.

Per saperne di più:

http://www.francoangeli.it/come_publicare/publicare_19.asp

I lettori che desiderano informarsi sui libri e le riviste da noi pubblicati possono consultare il nostro sito Internet: www.francoangeli.it e iscriversi nella home page al servizio "Informatemi" per ricevere via e-mail le segnalazioni delle novità.

Piero Demetrio Falorsi,
Patrizia Falzetti, Roberto Ricci

LE METODOLOGIE
DI CAMPIONAMENTO
E SCOMPOSIZIONE
DELLA DEVIANZA NELLE
RILEVAZIONI NAZIONALI
DELL'INVALSI

LE RILEVAZIONI DEGLI APPRENDIMENTI
A.S. 2018-2019



FrancoAngeli
OPEN  ACCESS



Progetto grafico di copertina: Alessandro Petrini

Copyright © 2019 by FrancoAngeli s.r.l., Milano, Italy.

L'opera, comprese tutte le sue parti, è tutelata dalla legge sul diritto d'autore ed è pubblicata in versione digitale con licenza Creative Commons Attribuzione-Non Commerciale-Non opere derivate 4.0 Internazionale (CC-BY-NC-ND 4.0)

L'Utente nel momento in cui effettua il download dell'opera accetta tutte le condizioni della licenza d'uso dell'opera previste e comunica sul sito
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.it>

Copyright © 2019 by FrancoAngeli s.r.l., Milano, Italy. ISBN 9788891798206

Indice

1. Introduzione	pag. 7
2. Le rilevazioni INVALSI	» 10
3. Tecnica di misurazione	» 11
4. Obiettivi dell'indagine	» 13
5. Campionamento	» 18
6. Stime campionarie	» 31
7. Variabilità campionaria	» 53
8. Il <i>cheating</i> e l'affidabilità dei dati	» 76
Bibliografia	» 81
Appendice – Approfondimenti metodologici	» 83
Indice delle tabelle	» 103
Indice delle figure	» 107

Si ringraziano tutti i colleghi dell'Area 2 – Servizio Statistico dell'INVALSI, in particolare Andrea Bendinelli e Chiara Sacco.

1. Introduzione

L'INVALSI (Istituto nazionale per la valutazione del sistema educativo di istruzione e formazione, cfr. d.lgs. n. 286/2004), tra i suoi compiti, ha quello di “attuare verifiche periodiche e sistematiche sulle conoscenze e abilità degli studenti”.

Tale compito è realizzato effettuando ogni anno una *Rilevazione sugli apprendimenti* che, secondo quanto stabilito dal DPR n. 80/2013 e dal d.lgs. n. 62/2017, è rivolta ai seguenti *gradi scolari*¹:

- le classi II e V della scuola primaria, corrispondenti ai gradi scolari 2° e 5° nelle quali le prove sono somministrate attraverso fascicoli cartacei. Gli allievi della II primaria svolgono una prova di Italiano, una di Matematica e, limitatamente agli allievi delle classi campione, una prova a tempo di decodifica strumentale della lettura. Gli allievi della V primaria svolgono una prova di Italiano, una di Matematica e una di Inglese, articolata in due parti: lettura (*Reading*) e ascolto (*Listening*);
- le classi III della scuola secondaria di primo grado corrispondente al grado scolare 8° nelle quali le prove vengono somministrate tramite modalità *Computer based testing* (CBT), introdotte dal decreto legislativo n. 62 del 13 aprile 2017, con conseguente rilascio di certificazione delle competenze per ciascuna disciplina, contenenti il posizionamento delle abilità in 5 livelli descrittivi per Italiano e Matematica, in 3 livelli corrispondenti al Quadro comune europeo di riferimento per la conoscenza delle lingue (QCER) per Inglese (*Listening* e *Reading*). Lo svolgimento della prova è requisito di accesso all’esame di Stato a conclusione del I ciclo d’istruzione;
- le classi II e V della scuola secondaria di secondo grado, corrispondenti ai gradi scolari 10° e 13°, nelle quali le prove vengono somministrate con modalità CBT. Per il grado 13, le prove, introdotte dal decreto legislativo n. 62 del 13 aprile 2017, prevedono il rilascio di certificazioni delle competenze per ciascuna disciplina, contenenti il posizionamento delle abilità in 5 livelli descrittivi per Italiano

¹ Il presente lavoro fa riferimento nello specifico alle Rilevazioni nazionali svolte durante l’anno scolastico 2018/19.

e Matematica, in 3 livelli corrispondenti al Quadro comune europeo di riferimento per la conoscenza delle lingue (QCER) per Inglese (*Listening e Reading*). Lo svolgimento della prova sarà requisito di accesso all'esame di Stato a conclusione del II ciclo d'istruzione a partire dall'a.s. 2019/20².

La Rilevazione sui gradi scolari sopra citati è condotta su tutte le scuole, statali e paritarie, del Paese e su tutti gli studenti delle suddette classi. Le informazioni di interesse sono raccolte mediante la somministrazione agli alunni di prove su ambiti disciplinari specifici (Italiano³, Matematica e Inglese)⁴.

La Rilevazione è costituita dall'unione di due indagini distinte che differiscono per la modalità di somministrazione delle prove che rimangono comunque le medesime in entrambe le rilevazioni:

- nell'*Indagine Standard* (di seguito denominata IS) le prove sono effettuate avvalendosi di somministratori appartenenti al corpo docente della scuola. L'indagine è effettuata sulla quasi totalità delle classi per gli anni scolari oggetto di interesse;
- nell'*Indagine Campionaria di Controllo* (di seguito denominata ICC) le prove sono svolte sotto il controllo di osservatori esterni al corpo docente della classe e della scuola. L'ICC coinvolge la parte rimanente delle classi non interessate dall'IS. Tale sottoinsieme viene scelto con criteri probabilistici.

Il compito dell'osservatore esterno nell'ICC è quello di monitorare la correttezza della somministrazione, a garanzia del rispetto delle procedure, e di riportare le risposte fornite dagli allievi su apposite schede elettroniche predisposte da INVALSI per i gradi 2 e 5 mentre per i gradi 8, 10 e 13 l'acquisizione delle risposte, grazie alla modalità di somministrazione CBT, è automatizzata. Pertanto l'intervento dell'osservatore è in particolar modo a salvaguardia del protocollo di somministrazione. La scelta di prevedere un'indagine campionaria con osservatori esterni è connessa al fatto che questi ultimi sono meno propensi ad aiutare gli allievi nella risoluzione delle prove (fenomeno noto con il termine *cheating*)⁵ e, di conseguenza, le valutazioni che si ottengono con l'ICC, pur se soggette all'errore di campionamento, sono meno esposte all'errore di misura derivante dal *cheating* che tende a sovrastimare i livelli effettivi delle competenze. Il fenomeno del *cheating* viene trattato in modo più approfondito nel capitolo 8 del presente lavoro.

² Ai sensi dell'art. 6, comma 3-*septies* della legge 108/2018 gli effetti dell'art. 13, comma 2, lettera b) del d.lgs. 62/2017 sono rinviati all'1.9.2019, pertanto per l'a.s. 2018-19 le prove INVALSI non costituiscono requisito di ammissione all'esame di maturità.

³ La prova di Italiano comprende una parte relativa alla comprensione della lettura e una parte relativa alla riflessione sul funzionamento della lingua italiana (Grammatica).

⁴ Ci si riferisce alle prove di Italiano, Matematica, Inglese *Reading* e Inglese *Listening* (le prove di Inglese sono somministrate al grado 5, al grado 8 e al grado 13). I livelli di competenza sono calcolati per Italiano, Matematica (grado 8, grado 10 e grado 13); Inglese *Reading e Listening* (grado 5, grado 8 e grado 13).

⁵ Con il termine *cheating* in ambito educativo si indicano le pratiche messe in atto dagli studenti (*student cheating*) o dagli insegnanti (*teacher cheating*) per alterare, sovrastimandoli, i risultati delle prove standardizzate di valutazione.

Per l'anno scolastico 2018/19, la Rilevazione nel suo complesso ha coinvolto 138.051 classi di circa 13.000 istituti, statali e paritari, e tutti gli studenti delle suddette classi, ossia 2.666.931 alunni. L'ICC ha coinvolto 166.105 allievi in 8.218 classi.

Tab. 1 – Totale delle classi e degli studenti per grado scolastico coinvolti nella Rilevazione per l'anno scolastico 2018/19

<i>Grado scolastico</i>	<i>Totale classi</i>	<i>Totale classi campione</i>	<i>Totale studenti</i>	<i>Totale studenti campione indagine ICC</i>
2°: II primaria	28.546	1.362	522.266	26.143
5°: V primaria	29.481	1.362	556.875	26.959
8°: III secondaria primo grado	27.333	1486	563.726	31.393
10°: II secondaria secondo grado	26.406	1.922	536.782	40.893
13°: V secondaria secondo grado	26.285	2.086	487.282	40.717
Totale	138.051	8.218	2.666.931	166.105

Il presente rapporto è finalizzato a illustrare le principali caratteristiche dell'*Indagine campionaria di controllo* ed è articolato nel modo seguente: i capitoli 2, 3 e 4 descrivono rispettivamente il sistema delle indagini, le tecniche di misurazione utilizzate e gli obiettivi statistici che vengono perseguiti. I capitoli 5, 6 e 7 approfondiscono le parti relative al campionamento, alla produzione delle stime e al calcolo degli errori campionari. Infine, il capitolo 8 esamina la problematica del *cheating*.

2. Le rilevazioni INVALSI

Negli ultimi due anni il disegno delle prove INVALSI è stato profondamente modificato, in particolare per gli effetti prodotti dal d.lgs. n. 62/2017. A partire dal 2018 le prove INVALSI della scuola secondaria (gradi 8, 10 e 13) sono svolte interamente al computer e online. Tale soluzione colloca l'Italia in una posizione di avanguardia internazionale, tanto da essere divenuto un caso di studio osservato con interesse da molti Paesi (Francia, Regno Unito, Brasile, Germania, Irlanda ecc.).

Tuttavia, la novità principale delle nuove prove INVALSI non risiede solo nella soluzione tecnica di somministrazione, ma anche nella modalità di restituzione dei risultati. Essi sono forniti in modo da garantire la comparabilità diacronico-longitudinale degli esiti. In questo modo è possibile valutare, in modo scientificamente corretto, l'impatto delle politiche scolastiche, a livello *macro*, *meso* e *micro*. Inoltre, la somministrazione computerizzata consente di valutare gli esiti degli studenti attraverso un numero molto più elevato di quesiti. Ciò permette di passare da una valutazione puramente numerica, tramite un punteggio medio di sintesi, a una valutazione di competenza, legata ai traguardi delle Indicazioni nazionali/Linee guida (Italiano e Matematica) e del QCER per l'Inglese. A ben vedere, quest'ultimo aspetto attenua il confine tra una valutazione puramente sommativa e una formativa, rendendo più semplici ed empiricamente verificabili le azioni di miglioramento, volte a innalzare i livelli di apprendimento degli studenti delle scuole italiane.

3. Tecnica di misurazione

La somministrazione delle prove INVALSI viene effettuata in genere nell'arco temporale compreso tra marzo e maggio. Per l'anno scolastico 2018/19 le prove sono iniziate il 4 marzo e sono terminate il 18 maggio con la somministrazione alla classe II della scuola secondaria di secondo grado. Il tempo previsto per ciascuna prova è differenziato in base al grado scolare. Di seguito si riportano i tempi di somministrazione di ciascuna prova e del questionario rivolto allo studente, ove previsto.

Tab. 2 – Tempi di somministrazione

Prova	Cartaceo			CBT	
	II primaria	V primaria	III sec. I gr.	II sec. II gr.	V sec. II gr.
Prova preliminare di lettura	2 min.	–	–	–	–
Prova di Italiano	45 min.	75 min.	90 min.	90 min.	120 min.
Prova di Matematica	45 min.	75 min.	90 min.	90 min.	120 min.
Prova di Inglese <i>Reading</i>	–	30 min.	45 min.	–	90 min.
Prova di Inglese <i>Listening</i>	–	30 min.	30 min.	–	60 min.

Come tutti gli anni, nella scuola primaria, vista l'età degli alunni coinvolti, le prove si sono svolte in due giornate distinte al fine di evitare l'effetto di affaticamento che si sarebbe potuto verificare se le prove fossero state somministrate nello stesso giorno, una di seguito all'altra.

Alle prove hanno partecipato tutti gli studenti delle classi interessate. Per quanto riguarda, in particolare, gli alunni con bisogni educativi speciali, le modalità di partecipazione alle prove variano a seconda che si svolgano prove di tipo cartaceo o di tipo CBT. Il d.lgs. n. 62/2017 definisce i criteri in base ai quali sono previste misure compensative (tempo aggiuntivo, formati speciali, sintetizzatore vocale per la lettura dei testi scritti) o dispensative, queste ultime riservate solo agli allievi disabili certificati secondo la legge n. 104/1992 e in base al PEI (Piano educativo individualizzato) elaborato dalle scuole.

Al termine di ciascuna giornata di somministrazione, l'INVALSI invia, tramite e-mail, a tutte le scuole le griglie di correzione delle prove cartacee. Per le rilevazioni

svolte tramite somministrazione informatizzata la correzione è centralizzata al fine di garantire sia per il campione, ma anche per l'intera popolazione, omogeneità nel criterio di correzione. Tutto ciò che può essere codificato in modo univoco, quindi domande che prevedono risposte chiuse, *cloze*, un ordinamento, un *gap match* ecc. viene corretto automaticamente, senza interventi particolarmente complicati, poiché si tratta di definire la risposta corretta e il sistema codifica di conseguenza tutte le risposte degli studenti. Per le domande a risposta aperta (per es. scrivere in una casella di testo un procedimento, un'argomentazione, un calcolo ecc.) il procedimento di codifica richiede, invece, l'implementazione di procedure più complesse per il trattamento di stringhe più o meno articolate.

In una fase che precede le somministrazioni il team di correzione costituito da statistici si confronta con il gruppo degli autori degli item per definire i criteri di correzione, ossia una serie di regole che determinano la classificazione di ciascuna risposta a un determinato item come corretta o errata.

Individuati i criteri di correzione questi vengono tradotti in *patterns* logici informatici che codificano una serie di modi possibili di dare una risposta corretta. Il dialogo con gli autori consente, inoltre, al gruppo di correzione di individuare una serie di operazioni sui dati testuali che permettono di rimuovere elementi non utili alla classificazione in corretta o errata. Per esempio, in alcuni item delle prove di Italiano l'algoritmo svolge preliminarmente le operazioni che seguono: individuazione e rimozione della punteggiatura, dei caratteri speciali, degli articoli, delle congiunzioni, lemmatizzazione delle parole ecc.

Un altro elemento affrontato con gli autori nelle fasi preliminari è la valutazione, per ciascun item, dell'opportunità di sottoporre le stringhe di risposta a correzione automatizzata degli errori di ortografia e digitazione: l'algoritmo implementato è in grado di individuare e sostituire le parole "fuori vocabolario" (*OOV words*).

Implementato l'algoritmo sulla base di quanto emerso dalle procedure fin qui illustrate, l'efficienza e la precisione dello stesso viene sottoposta a test sulla base delle risposte registrate e già corrette "a video" nei pretest: lo scopo di questa fase è di raggiungere una perfetta concordanza tra gli esiti della correzione automatizzata e quelli della correzione manuale.

Terminate le fasi di collaudo, si calendarizzano con gli autori incontri periodici. Questi incontri hanno lo scopo di sottoporre agli autori dei report in cui sono individuate delle risposte dubbie che richiedono una valutazione *ad hoc* che può determinare delle modifiche alle regole di correzione.

L'approccio garantisce un progressivo affinamento della precisione di correzione e della sua coerenza con le indicazioni degli autori. Un approccio metodologico analogo è stato utilizzato tra gli altri dai lavori dell'*Automated Student Assessment Prize* (ASAP) sponsorizzato da William and Flora Hewlett Foundation (Hewlett Foundation) (Tandalla, 2012) ed è definibile come correzione automatizzata supervisionata, che rappresenta un valido compromesso tra una codifica manuale e una totalmente automatizzata (per es. algoritmi di *machine learning*).

4. Obiettivi dell'indagine

L'ICC ha la finalità di fornire stime di:

- valori medi dei punteggi riportati alle prove;
- differenti componenti della devianza tra i punteggi, distinguendo la variabilità individuale all'interno della classe, da quella tra classi dello stesso istituto e quella tra istituti differenti;
- livelli descrittivi di apprendimento secondo i QdR (per le prove di Inglese i livelli provengono dal Quadro comune europeo di riferimento per la conoscenza delle lingue – QCER)¹.

Le stime di cui sopra devono essere fornite separatamente per ciascuna tipologia di prova (Italiano, Matematica, Inglese *Reading* e Inglese *Listening*) e per domini di stima².

Nella rilevazione in oggetto, i domini di stima sono individuati dall'incrocio tra le diverse aree territoriali e le sottopopolazioni di alunni definite dalla *tipologia di scuola* e dal grado scolastico di interesse.

I principali domini di studio sono i 22 domini territoriali:

- l'Italia;
- le province autonome di Trento e Bolzano (lingua italiana)³ e le restanti 19 regioni.

Si noti che le stime possono essere anche prodotte per ripartizione geografica. Tuttavia il disegno di campionamento di seguito illustrato è realizzato considerando come prioritari i 22 domini sopra citati.

Si distinguono 7 *tipologie di scuola*:

- 1) scuola primaria;
- 2) scuola secondaria di primo grado;

¹ Si rimanda al documento *I livelli per la descrizione degli esiti delle prove INVALSI* disponibile al seguente link: https://invalsi-areaprove.cineca.it/docs/2018/Livelli_INVALSI_g8.pdf.

² I domini di stima sono sotto-insiemi della popolazione oggetto d'indagine (gli alunni della scuola primaria e secondaria in Italia) per i quali si vuole fissare a priori la dimensione del campione, ossia quanti alunni appartenenti alle sottopopolazioni saranno rilevati nell'indagine campionaria.

³ Nel seguito per semplicità espositiva, il termine regione sarà usato in modo generico, indicando anche la provincia autonoma.

- 3) licei scientifici;
- 4) licei classici;
- 5) altri licei;
- 6) istituti tecnici;
- 7) istituti professionali.

I gradi scolari oggetto di rilevazione sono:

- il secondo e il quinto anno di corso della scuola primaria;
- l’ottavo anno scolastico relativo al terzo anno di corso della scuola secondaria di primo grado;
- il decimo e il tredicesimo anno scolastico, relativi rispettivamente al secondo e al quinto anno di corso della scuola secondaria di secondo grado.

L’incrocio tra gradi scolari e tipologia di scuola porta a definire 13 distinte sottopopolazioni.

Tab. 3 – Sottopopolazioni individuate dal grado scolastico e dalla tipologia di scuola

<i>Tipologia di scuola</i>	<i>Grado scolastico</i>	<i>Sottopopolazione</i>
Scuola primaria	2°	1
	5°	2
Scuola secondaria di primo grado	8°	3
	10°	4
Licei scientifici	13°	5
	10°	6
Licei classici	13°	7
	10°	8
Altri licei	13°	9
	10°	10
Istituti tecnici	13°	11
	10°	12
Istituti professionali	13°	13

In totale quindi, l’ICC deve produrre stime per 286 domini di stima, dove tale numero è ottenuto dal prodotto di 22 (domini territoriali) per 13 (sottopopolazioni).

4.1. Notazione di base

Al fine di descrivere formalmente i parametri di interesse è utile considerare la notazione sotto riportata.

Si indichi con:

t la tipologia di scuola: $t=1, \dots, 7$;

r il singolo anno scolastico: $r=1, \dots, 13$; con $r=1, \dots, 5$ per $t=1$; $r=6, 7, 8$ per $t=2$; $r=9, \dots, 13$ per $t>2$;

- h la regione o la provincia autonoma;
 - i il plesso scolastico (in seguito indicato in modo generico anche come scuola);
 - j la sezione nell'ambito dell'istituto;
 - k il singolo alunno;
 - A l'insieme degli alunni e il numero di alunni;
 - S l'insieme delle strutture scolastiche. A livello Italia o a livello regionale, il simbolo S denota l'insieme delle scuole; quando si considera una specifica scuola, indica i plessi o le sezioni nell'ambito della stessa e anche il numero delle strutture stesse.
- A* e *S* sono ulteriormente denotati mediante i pedici *t* e *r* (nell'ordine) alla loro sinistra e con gli indici *h*, *i* e *j* (sempre nell'ordine) alla loro destra. La mancanza di un indice denota che la quantità rappresenta il totale rispetto all'indice mancante. Al fine di rendere più chiaro l'utilizzo della notazione introdotta, ne riportiamo alcuni esempi nel box seguente.

Box 4.1 – Esempi di utilizzo della notazione

${}_{tr}A$ indica il numero (e la popolazione corrispondente) degli scolari che in Italia che frequentano l'anno *r* nella tipologia di scuola *t*;

${}_{tr}S$ denota il numero (e la popolazione corrispondente) dei plessi scolastici per l'anno *r* nella tipologia di scuola *t*;

${}_{tr}A(h)$ rappresenta il numero (e la popolazione corrispondente) degli scolari che in Italia frequentano l'anno *r* nella tipologia di scuola *t* nella regione o provincia autonoma (*h*);

${}_{tA(h)ij}$ indica il numero degli alunni che frequentano la sezione *j* nella scuola *i* di tipologia *t* nella regione *h*;

${}_{tS(h)i}$ indica il numero delle sezioni nella scuola *i* di tipologia *t* nella regione *h*.

Dato ${}_{tr}y(h)_{ijk}$ il punteggio (o voto alla prova) *y* ottenuto dall'alunno *k* che frequenta l'anno *r* nella sezione *j* della scuola *i* di tipologia *t* nella regione *h*, si denota con

$${}_{tr}\bar{y}(h)_{ij} = \frac{{}_{tr}Y(h)_{ij}}{{}_{tr}A(h)_{ij}}, \quad {}_{tr}\bar{y}(h)_i = \frac{{}_{tr}Y(h)_i}{{}_{tr}A(h)_i},$$

rispettivamente il valore medio del punteggio della prova *y* per l'anno scolastico *r* per gli scolari della sezione *j* della scuola *i* e il valore medio per la scuola, dove

$${}_{tr}Y(h)_{ij} = \sum_{k=1}^{{}_{tr}A(h)_{ij}} {}_{tr}y(h)_{ijk}, \quad {}_{tr}Y(h)_i = \sum_{k=1}^{{}_{tr}S(h)_i} {}_{tr}y(h)_{ijk}$$

sono i rispettivi totali del punteggio nella sezione *j* e nella scuola *i*.

4.2. Punteggi medi

A partire dalla notazione precedentemente introdotta, i parametri d'interesse dell'ICC relativi ai valori medi dei punteggi regionali e nazionali alle prove sono espressi rispettivamente come:

$${}_{tr}\bar{y}^{(h)} = \frac{1}{{}_{tr}A^{(h)}} \sum_{i=1}^{{}_{tr}S^{(h)}} \sum_{j=1}^{{}_{tr}S^{(h)}i} \sum_{k=1}^{{}_{tr}A^{(h)}ij} {}_{tr}y^{(h)ijk}, \quad t=1,\dots,5; r \in {}_tR; h=1,\dots,21 \quad [4.1]$$

$${}_{tr}\bar{y} = \frac{1}{{}_{tr}A} \sum_{h=1}^{21} \sum_{i=1}^{{}_{tr}S^{(h)}} \sum_{j=1}^{{}_{tr}S^{(h)}i} \sum_{k=1}^{{}_{tr}A^{(h)}ij} {}_{tr}y^{(h)ijk} \quad t=1,\dots,5; r \in {}_tR; \quad [4.2]$$

con ${}_tR: \{r=2, 5 \text{ per } t=1; r=8 \text{ per } t=2; r=10, 13 \text{ per } t>2\}$.

In appendice A.1, sono riportate alcune utili riformulazioni dei parametri precedenti.

4.3. Componenti della devianza

La devianza complessiva, ${}_{tr}D$, tra i punteggi ${}_{tr}y^{(h)ijk}$ degli allievi che frequentano l'anno scolastico r nella tipologia di scuola t è data da:

$${}_{tr}D = \sum_{h=1}^{21} \sum_{i=1}^{{}_{tr}S^{(h)}} \sum_{j=1}^{{}_{tr}S^{(h)}i} \sum_{k=1}^{{}_{tr}A^{(h)}ij} ({}_{tr}y^{(h)ijk} - {}_{tr}\bar{y})^2 \quad t=1,\dots,7; r \in {}_tR \quad [4.3]$$

Aggiungendo e sottraendo all'interno della differenza quadratica i valori medi ${}_{tr}\bar{y}^{(h)ij}$ e ${}_{tr}\bar{y}^{(h)i}$, la devianza ${}_{tr}D$ può essere espressa come somma di tre componenti (vedi appendice A.2):

$${}_{tr}D = {}_{tr}D1 + {}_{tr}D2 + {}_{tr}D3$$

che denotano rispettivamente la devianza tra scuole (${}_{tr}D1$), quella tra classi della medesima scuola (${}_{tr}D2$) e quella tra gli alunni di una stessa classe (${}_{tr}D3$), dove

$${}_{tr}D1 = \sum_{h=1}^{21} \sum_{i=1}^{{}_{tr}S^{(h)}} \sum_{r} A_{(h)i} ({}_r\bar{y}^{(h)i} - {}_r\bar{y})^2$$

$${}_{tr}D2 = \sum_{h=1}^{21} \sum_{i=1}^{{}_{tr}S^{(h)}} \sum_{j=1}^{{}_{tr}S^{(h)}i} \sum_{r} A_{(h)j} ({}_r\bar{y}^{(h)j} - {}_r\bar{y}^{(h)i})^2$$

$${}_{tr}D3 = \sum_{h=1}^{21} \sum_{i=1}^{{}_{tr}S^{(h)}} \sum_{j=1}^{{}_{tr}S^{(h)}i} \sum_{k=1}^{{}_{tr}A^{(h)}ij} ({}_r y^{(h)ijk} - {}_r\bar{y}^{(h)j})^2$$

Le espressioni della devianza e delle relative componenti a livello regionale sono determinate mediante le espressioni precedenti, limitando le sommatorie alla specifica regione.

L'analisi delle componenti della devianza fornisce una chiave di interpretazione degli esiti delle prove standardizzate di considerevole rilevanza. In un sistema scolastico "ideale", la devianza totale ${}_tD$ dovrebbe essere spiegata dalle differenze osservate tra gli studenti (${}_tD3$) e non dalla loro suddivisione nelle classi (${}_tD2$) o nelle scuole (${}_tD1$). Infatti, tanto più un sistema è caratterizzato da bassa devianza tra le classi e tra le scuole, tanto maggiore è l'equità di quel sistema, ipotizzando che i risultati medi generali siano elevati. Infatti, in questo caso l'incidenza dell'inserimento in una data classe o in una data scuola ha poco peso e buoni risultati sono garantiti a tutti e a ciascuno, sostanzialmente in ragione delle caratteristiche individuali dell'allievo, indipendentemente dalla particolare scuola o classe frequentata.

Si può vedere dalle tabelle successive (dalla 19 alla 28) e dai grafici (figure dalla 7 alla 10) come si distribuisce la variabilità tra classi, nelle classi e tra scuole per materia e grado scolastico all'interno del Paese.

5. Campionamento

Al fine di stimare i parametri ${}_{tr}\bar{y}(h)$ e ${}_{tr}\bar{y}$, si seleziona un campione probabilistico a due stadi di selezione, stratificato al primo stadio, in cui le unità del primo stadio sono i plessi scolastici (le scuole) e quelle del secondo stadio le sezioni.

5.1. Primo stadio di campionamento

Le scuole sono stratificate per tipologia, t , e regione, h .

Questo tipo di stratificazione garantisce che i 286 domini di stima corrispondano a uno strato o possano essere ottenuti mediante aggregazioni degli strati dell'indagine. In tal modo è possibile pianificare la numerosità campionaria nei domini di stima e controllare in fase di disegno l'errore di campionamento.

Nell'ambito dello specifico strato, ${}_{tS}(h)$, si seleziona senza re-immissione e con probabilità variabili un campione probabilistico, ${}_{tS}(h)$, di istituti scolastici (tale simbolo denota anche la dimensione del campione stesso).

La i -esima scuola dello strato viene selezionata con probabilità di inclusione, ${}_{t\pi}(h)_i$, proporzionale al numero di alunni della scuola:

$${}_{t\pi}(h)_i = {}_{tS}(h) {}_{t\alpha}(h)_i \quad [5.1a]$$

dove

$${}_{t\alpha}(h)_i = \frac{{}_{t\ddot{A}}(h)_i}{\sum_{i=1}^{{}_{tS}(h)} {}_{t\ddot{A}}(h)_i} \quad [5.1b]$$

in cui

$$\begin{aligned} - {}_{t\ddot{A}}(h)_i &= \left[\frac{1}{2} ({}_{t2}A(h)_i + {}_{t5}A(h)_i) \right] \frac{2}{{}_{tS}(h)_i} && \text{per } t=1, \\ - {}_{t\ddot{A}}(h)_i &= {}_{t6}A(h)_i \frac{2}{{}_{tS}(h)_i} && \text{per } t=2, \end{aligned}$$

$$- \quad {}_t\ddot{A}(h)_i = {}_t10A(h)_i \frac{2}{{}_tS(h)_i} \quad \text{per } t > 2$$

e

$${}_tS(h)_i = \min(2, {}_tS(h)_i) \quad [5.2]$$

Il rapporto $2/{}_tS(h)_i$, pari a 1 per gli istituti scolastici con almeno due sezioni e uguale a 2 per le piccole scuole con una sola sezione, raddoppia la probabilità di inclusione dei piccoli istituti scolastici assicurando un'uguale probabilità di inclusione finale a tutti gli alunni e a tutte le sezioni di un medesimo strato (vedi paragrafo 5.3) a prescindere dalla dimensione, in termini di alunni, della scuola.

Il campione è selezionato con il metodo di Pareto (Särndal e Lundström, 2005), mediante una procedura articolata nei seguenti passi:

- per ogni scuola i , si genera la variabile casuale ${}_t\omega(h)_i$ come realizzazione di una distribuzione uniforme $Unif(0,1)$ e si calcola

$${}_tq(h)_i = \frac{{}_t\omega(h)_i (1 - {}_t\pi(h)_i)}{{}_t\pi(h)_i (1 - {}_t\omega(h)_i)}$$

- il campione è costituito dalle ${}_tS(h)$ scuole a cui corrispondono i valori più bassi delle quantità ${}_tq(h)_i$.

Il numero atteso di classi campione in ciascuno strato è pari a

$${}_t\bar{S}(h) = \sum_{i=1}^{{}_tS(h)} {}_t\pi(h)_i \cdot {}_tS(h)_i \cdot$$

5.2. Secondo stadio

Nella scuola campione $(h)_i$ della tipologia t viene selezionato, a probabilità uguali e senza re-immissione, un campione, di ${}_tS(h)_i$ sezioni. In tal modo, in ciascuna scuola si seleziona un campione di 2 sezioni, a eccezione delle scuole con un'unica sezione che viene selezionata con certezza nel campione di secondo stadio.

Pertanto, la singola sezione viene selezionata con probabilità di secondo stadio pari a:

$${}_t\pi(h)_j|i = \frac{{}_tS(h)_i}{{}_tS(h)_i} \quad [5.3]$$

Nelle sezioni selezionate nel campione vengono sottoposti alla prova gli alunni delle classi i cui gradi scolari sono d'interesse per la rilevazione. Di conseguenza, tutti gli ${}_{tr}A(h)_{ij}$ alunni appartenenti all'anno di interesse r ($r \in {}_tR$) per la tipologia t della sezione estratta vengono inclusi nel campione.

Il totale degli alunni estratti nella scuola campione per il grado scolare r è

$${}^{tr}a_{(h)i} = \sum_{j \in {}_i S_{(h)}} {}^{tr}A_{(h)ij}.$$

Il valore atteso di alunni campione a livello regionale e nazionale è pari a:

$${}^{tr}\bar{a}_{(h)} = \sum_{i=1} {}^t S_{(h)} \sum_{j=1} {}^t S_{(h)i} {}^t \pi_{(h)i} {}^t \pi_{(h)j|i} {}^{tr}A_{(h)ij}.$$

5.3. Probabilità di inclusione finale

Il singolo alunno k che frequenta l'anno r nella scuola i della tipologia t dello strato h viene selezionato con probabilità di inclusione finale pari a:

$${}^{tr}\pi_{(h)ijk} = {}^t \pi_{(h)i} {}^t \pi_{(h)j|i} = {}^t S_{(h)} \frac{{}^t \ddot{A}_{(h)i}}{\sum_{i=1} {}^t S_{(h)} {}^t \ddot{A}_{(h)i}} \frac{{}^t S_{(h)i}}{{}^t S_{(h)i}} \quad [5.4]$$

Nell'ipotesi che la dimensione delle classi in termini di alunni sia pressappoco costante nello strato e che ogni classe sia approssimativamente costituita da ${}^t \bar{A}_{(h)}$ alunni, si dimostra che vale la seguente condizione:

$${}^{tr}\pi_{(h)ijk} \cong \frac{2 {}^t S_{(h)}}{{}^t \bar{S}_{(h)}}, \quad [5.5]$$

in cui ${}^t \bar{S}_{(h)} = {}_{(>1)} {}^t \bar{S}_{(h)} + 2 {}_{(=1)} {}^t \bar{S}_{(h)}$, dove ${}_{(>1)} {}^t \bar{S}_{(h)}$ denota il numero delle sezioni appartenenti alle scuole con almeno due sezioni e ${}_{(=1)} {}^t \bar{S}_{(h)}$ indica il numero delle sezioni nello strato appartenenti alle scuole con un'unica sezione.

La [5.5] è una relazione importante in quanto dimostra che con il criterio di selezione prescelto ciascun alunno e ciascuna classe dello strato hanno la medesima probabilità di essere selezionati nel campione, indipendentemente dalla dimensione della scuola selezionata nel primo stadio.

La condizione di *autoponderazione* (approssimativa) espressa dalla [5.5] rappresenta un fattore di *ottimalità del disegno* in quanto tende ad annullare l'incremento di variabilità derivante dalla variabilità dei pesi finali.

5.4. Allocazione del campione

Il numero di scuole a livello nazionale della tipologia t , s , è definito cercando di individuare la dimensione che preveda costi inferiori al *budget* disponibile e che al

contempo sia sostenibile dalla complessa organizzazione logistica che è necessario mettere in atto per tutte le fasi della rilevazione.

Tale analisi ha portato a definire una dimensione campionaria, t^S , differenziata per tipologia di scuola.

Questo numero è stato suddiviso tra due sottoinsiemi di regioni:

- l'insieme delle regioni Auto Rappresentative (AR), costituito dalle piccolissime regioni con meno di 20 istituti scolastici; in particolare l'insieme AR comprende la Valle d'Aosta, la Provincia autonoma di Bolzano (lingua italiana);
- l'insieme delle regioni Non Auto Rappresentative (NAR), costituito dalle rimanenti 19 regioni/province.

All'insieme AR viene assegnato un campione di t^S_{AR} scuole, essendo

$$t^S_{AR} = t^S_{AR} = \sum_{h \in AR} t^S(h)$$

All'insieme NAR, con numero di scuole pari a $t^S_{NAR} = t^S - t^S_{AR}$, viene assegnata la parte rimanente del campione nazionale $t^S_{NAR} = t^S - t^S_{AR}$.

Il campione t^S_{NAR} viene quindi allocato tra le regioni dell'insieme NAR mediante una procedura di compromesso che tenta di individuare una soluzione ragionevole tra l'esigenza di:

- trovare un criterio ottimale che porti a minimizzare la varianza della stima nazionale, $V(t_r \hat{y})$;
- garantire una varianza di campionamento approssimativamente uguale alle stime regionali $V(t_r \hat{y}(h))$.

Fissata la dimensione del campione di scuole a livello nazionale t^S_{NAR}

- il criterio al punto (a) porta a definire un campione regionale proporzionale al numero di scuole della regione:

$$t^S(h) = t^S_{NAR} \frac{t^S(h)}{t^S_{NAR}};$$

- il criterio al punto (b) porta a definire un campione approssimativamente uguale a livello di regione:

$$t^S(h) = \frac{t^S_{NAR}}{19}.$$

Il criterio di compromesso si fonda sulla definizione di un valore $t\beta$ compreso tra 0 e 1 che definisce la dimensione del campione mediante la seguente combinazione lineare convessa:

$$t^S(h) = \beta_t \left(t^S_{NAR} \frac{t^S(h)}{t^S_{NAR}} \right) + (1 - \beta_t) t^S_{NAR} \frac{t^S(h)}{t^S_{NAR}} \quad [5.6]$$

Il valore di ${}_t\beta$ viene definito in modo tale che la dimensione campionaria assicuri che l'errore di campionamento a livello nazionale e regionale rimanga sotto determinate soglie ritenute accettabili per la pubblicazione dei dati:

$${}_t\beta: \begin{cases} CV_{(tr, \hat{y}(h))} \leq {}_t\overline{CVH} \\ CV_{(tr, \hat{y})} \leq {}_t\overline{CV} \end{cases} \quad h \in NAR$$

in cui $CV_{(tr, \hat{y}(h))}$ e $CV_{(tr, \hat{y})}$ denotano i coefficienti di variazione della stima a livello regionale e nazionale e ${}_t\overline{CVH}$ e ${}_t\overline{CV}$ denotano i valori soglia di tali coefficienti ritenuti accettabili per l'allocazione, essendo

$$CV_{(tr, \hat{y}(h))} = \sqrt{V_{(tr, \hat{y}(h))}} / {}_{tr}\hat{y}(h) \quad [5.7]$$

e

$$CV_{(tr, \hat{y})} = \sqrt{V_{(tr, \hat{y})}} / {}_{tr}\hat{y} \quad [5.8]$$

Dal punto di vista computazionale, i valori per il calcolo dei coefficienti di variazione sono desunti dalla rilevazione dell'anno precedente¹. Le formule di calcolo adottate sono illustrate nel seguente paragrafo 6. La variabile y presa a riferimento per il calcolo delle numerosità campionarie è la media dei punteggi complessivi. In genere, il valore di ${}_t\beta$ prescelto si attesta intorno a 0,3.

Le allocazioni campionarie per l'anno 2019 per le varie tipologie di scuole evidenziano una similarità di fondo; pertanto, ci limiteremo alla discussione di quella relativa alla scuola secondaria di primo grado riportata immediatamente di seguito (tab. 4) e alla novità del 2019 relativa all'ultimo anno della scuola secondaria di secondo grado (tabelle dalla 5 alla 9).

¹ Per il grado 13, trattandosi del primo anno di rilevazione, si sono presi come valori di riferimento quelli del grado 10 dell'anno precedente.

Tab. 4 – Allocations del campione nella scuola secondaria di primo grado classi III: anno 2018/19

Regione e/o Provincia autonoma	Totale scuole	Totale alunni	Campione				CV%	
			Allocazione proporzionale tra le regioni		Allocazione uguale tra le regioni			Allocazione di compromesso $\beta=0,3$
			Scuole	CV%	Scuole	CV%		
Valle d'Aosta	17	1.165	17	0,4	17	0,4	17	0,4
Piemonte	394	38.494	48	0,8	38	1,0	41	0,9
Liguria	144	12.755	16	1,5	38	0,9	32	1,0
Lombardia	963	94.446	110	0,5	38	0,9	61	0,7
Prov. Aut. di Bolzano (l. it.)	17	1.429	17	0,4	17	0,4	17	0,4
Prov. Aut. di Trento	58	5.734	5	1,8	38	0,5	28	0,7
Veneto	465	47.322	45	0,7	38	0,7	40	0,7
Friuli-Venezia Giulia	123	10.388	10	1,4	38	0,7	29	0,8
Emilia-Romagna	371	39.804	44	0,8	38	0,8	39	0,8
Toscana	320	33.150	40	0,9	38	0,9	38	0,9
Umbria	79	7.831	8	1,7	38	0,6	28	0,8
Marche	156	13.778	13	1,2	38	0,7	62	0,8
Lazio	560	53.485	69	0,7	38	1,0	49	0,9
Abruzzo	134	11.577	12	1,4	38	0,7	32	0,8
Molise	34	2.589	3	3,2	32	0,2	19	0,6
Campania	594	65.415	107	0,8	38	1,4	58	1,1
Puglia	354	40.844	53	0,9	38	1,0	44	1,0
Basilicata	77	5.317	6	2,3	38	0,7	28	0,9
Calabria	248	18.776	25	1,4	38	1,1	34	1,2
Sicilia	525	51.955	85	0,9	38	1,4	53	1,2
Sardegna	171	13.755	18	1,5	38	0,9	32	1,1
Italia	5.804	570.009	750	0,2	750	0,3	750	0,3

Si notano i seguenti aspetti:

- l’allocazione proporzionale, porta al valore minimo del coefficiente di variazione della stima nazionale (pari allo 0,22%). Tale metodo porta a privilegiare le grandi regioni; mentre le piccole e medie regioni risultano svantaggiate. Ciò nonostante, vi è solo una regione con valore del CV% superiore al 3% (cella in grigio) e nessuna regione ha un CV superiore al 4%;
- l’allocazione uguale, conduce a un peggioramento del coefficiente di variazione della stima nazionale (che passa dallo 0,22% allo 0,30%). D’altra parte, vi è un maggiore equilibrio negli errori delle stime regionali. Nessuna regione presenta un CV superiore all’1,5%;
- l’allocazione di compromesso è un’ottima soluzione, in quanto si controlla il coefficiente di variazione della stima nazionale mantenendolo al di sotto dell’allocazione equiripartita ma si avvicina a quest’ultima sotto il profilo dell’andamento dei CV regionali.

Di seguito si può vedere l’allocazione relativa alle 5 tipologie di istituto scelte per la rappresentatività del campione: licei scientifici, licei classici, altri licei, istituti tecnici e istituti professionali. Nel paragrafo 6.6 si esplicita meglio il perché della scelta di queste 5 tipologie e non di altre.

Tab. 5 – Allocations del campione nella scuola secondaria di secondo grado classi V – Licei scientifici: anno 2018/19

Regione e/o Provincia autonoma e tipologia di scuola	Totale scuole		Totale alunni		Campione			
	Scuole	CV%	Scuole	CV%	Allocazione proporzionale tra le regioni		Allocazione uguale tra le regioni	
					Scuole	CV%	Scuole	CV%
Valle d'Aosta	2	167	2	–	2	–	2	–
Piemonte	96	6.823	13	1,9	11	2	13	1,9
Liguria	35	2.536	5	3,1	11	1,8	9	2,2
Lombardia	233	15.503	31	1,2	11	2,1	17	1,7
Prov. Aut. di Bolzano (l. it.)	7	238	7	–	7	–	7	–
Prov. Aut. di Trento	14	803	2	5,8	11	1,6	6	2,7
Veneto	92	7.070	11	1,6	11	1,6	12	1,6
Friuli-Venezia Giulia	31	1.852	4	3,8	11	2	9	2,4
Emilia-Romagna	81	6.413	12	1,9	11	1,9	12	1,8
Toscana	82	5.697	12	2,2	11	2,4	12	2,3
Umbria	19	1.757	3	3,4	11	1,5	8	1,8
Marche	30	2.561	5	2,8	11	1,5	10	1,6
Lazio	164	12.128	35	1,5	11	3,2	18	2,4
Abruzzo	34	2.857	5	2,8	11	1,6	10	1,7
Molise	12	671	2	2,3	11	0,5	7	1,0
Campania	176	15.597	25	1,3	11	2	15	1,7
Puglia	96	8.370	15	1,7	11	2	12	1,9
Basilicata	20	1.273	3	4,6	11	0,7	7	2,0
Calabria	59	4.527	8	2,4	11	1,9	11	1,9
Sicilia	124	9.371	18	1,7	11	2,3	13	2,1
Sardegna	45	2.629	5	3,4	11	2,1	9	2,5
Italia	1.453	108.843	223	0,5	222	0,7	218	0,6

Tab. 6 – Allocations del campione nella scuola secondaria di secondo grado classi V – Licei classici: anno 2018/19

Regione e/o Provincia autonoma e tipologia di scuola	Totale scuole		Totale alunni		Campione			
	Scuole	CV%	Scuole	CV%	Allocazione proporzionale tra le regioni		Allocazione uguale tra le regioni	
					Scuole	CV%	Scuole	CV%
Valle d'Aosta	1	43	1	—	1	—	1	—
Piemonte	37	1.467	3	4,6	5	3,5	6	3,2
Liguria	16	763	2	4,7	5	2,6	4	3,1
Lombardia	72	2.771	6	3,2	5	3,5	5	3,6
Prov. Aut. di Bolzano (l. it.)	3	51	3	—	3	—	3	—
Prov. Aut. di Trento	5	148	2	6,3	5	2,8	2	6,3
Veneto	44	1.642	5	5	5	5	6	4,6
Friuli-Venezia Giulia	12	448	2	4,1	5	0,6	4	1,9
Emilia-Romagna	24	1.230	5	6,1	5	5,7	6	5,3
Toscana	36	1.270	9	7,6	14	5,8	13	6,1
Umbria	11	527	2	6,2	5	3,7	4	4,3
Marche	18	739	5	7	8	5	7	5,6
Lazio	87	4.360	18	3,7	10	5,2	12	4,7
Abruzzo	13	692	5	7,4	9	5,4	7	6,2
Molise	5	177	5	7,5	5	7,2	3	10,3
Campania	90	4.305	11	3,3	5	5	10	3,5
Puglia	54	2.544	9	4,5	7	5,1	8	4,7
Basilicata	11	476	2	5,3	5	2,8	4	3,4
Calabria	34	1.562	4	5,4	5	4,5	5	4,6
Sicilia	81	4.022	15	3,9	9	5,2	11	4,7
Sardegna	21	1.011	4	7,9	8	5,1	7	5,5
Italia	675	30.248	118	1,2	132	1,4	128	1,3

Tab. 7 – *Allocazione del campione nella scuola secondaria di secondo grado classi V – Altri licei: anno 2018/19*

Regione e/o Provincia autonoma e tipologia di scuola	Totale scuole		Totale alunni		Campione				
	Scuole	CV%	Scuole	CV%	Allocazione proporzionale tra le regioni		Allocazione uguale tra le regioni		Allocazione di compromesso $\beta\beta=0,4$
					Scuole	CV%	Scuole	CV%	
Altri licei	5	289	5	–	5	–	5	–	–
Valle d'Aosta	82	6.399	13	1,4	11	1,5	11	1,5	1,5
Piemonte	24	2.539	5	2	11	0,8	9	0,8	1,1
Liguria	229	15.404	40	1	14	1,8	22	1,4	1,4
Lombardia	6	361	6	–	6	–	6	–	–
Prov. Aut. di Bolzano (l. it.)	16	1.169	2	2,7	7	0,5	5	1,1	1,1
Prov. Aut. di Trento	104	7.709	22	1,3	14	1,8	17	1,6	1,6
Veneto	24	1.766	3	2,4	11	0,9	9	1,1	1,1
Friuli-Venezia Giulia	70	6.844	15	1,3	11	1,5	12	1,4	1,4
Emilia-Romagna	93	6.998	19	1,5	14	1,9	15	1,8	1,8
Toscana	22	1.487	2	2,6	11	0,7	6	1,3	1,3
Umbria	40	3.104	5	1,7	11	0,8	9	1,0	1,0
Marche	154	11.007	14	1	11	1,1	12	1,1	1,1
Lazio	28	2.138	3	2,3	11	1	8	1,2	1,2
Abruzzo	8	514	2	2,4	6	0,5	5	0,9	0,9
Molise	170	11.777	21	1,1	11	1,6	14	1,4	1,4
Campania	88	6.841	10	1,2	11	1,1	11	1,1	1,1
Puglia	17	916	4	3,2	11	0,3	9	1,2	1,2
Basilicata	44	2.514	5	2,6	11	1,6	9	1,8	1,8
Calabria	152	9.443	19	1,3	11	1,8	13	1,6	1,6
Sicilia	40	2.619	4	2,3	11	1,3	9	1,5	1,5
Sardegna	1.416	101.838	218	0,4	219	0,5	216	0,4	0,4
Italia									

Tab. 8 – Allocations del campione nella scuola secondaria di secondo grado classi V – Istituti tecnici: anno 2018/19

Istituti tecnici	Totale scuole		Totale alunni		Campione			
	Provincia autonoma e tipologia di scuola		Allocazione proporzionale tra le regioni		Allocazione uguale tra le regioni		Allocazione di compromesso $\beta=0,4$	
	Scuole	CV%	Scuole	CV%	Scuole	CV%	Scuole	CV%
Valle d'Aosta	4	317	4	—	4	—	4	—
Piemonte	136	10.509	17	0,5	17	0,5	18	0,5
Liguria	41	3.084	5	1,2	17	0,3	14	0,5
Lombardia	304	23.969	44	0,4	17	0,7	26	0,6
Prov. Aut. di Bolzano (l. it.)	7	343	7	—	7	—	7	—
Prov. Aut. di Trento	26	1.726	5	1,4	8	0,4	7	0,8
Veneto	147	14.133	27	0,5	17	0,7	20	0,6
Friuli-Venezia Giulia	43	3.219	6	1	12	0,2	11	0,4
Emilia-Romagna	119	11.061	21	0,6	17	0,7	18	0,6
Toscana	110	8.109	25	0,8	19	1	22	0,9
Umbria	26	2.111	4	1,6	17	0,4	11	0,7
Marche	59	4.436	9	1	17	0,5	15	0,6
Lazio	183	13.064	27	0,7	19	0,9	22	0,8
Abruzzo	49	3.267	6	1,3	17	0,6	14	0,7
Molise	11	849	2	2	7	0,2	6	0,6
Campania	320	22.952	48	0,5	19	0,9	28	0,7
Puglia	162	12.676	32	0,6	19	1	23	0,8
Basilicata	37	1.780	2	0,7	13	0	11	0,1
Calabria	94	6.319	11	1	17	0,7	15	0,8
Sicilia	245	15.099	22	0,6	17	0,7	18	0,6
Sardegna	73	4.011	10	1,4	19	0,8	16	0,9
Italia	2.198	163.034	336	0,2	319	0,2	326	0,2

Tab. 9 – Allocations del campione nella scuola secondaria di secondo grado classi V – Istituti professionali: anno 2018/19

Istituti professionali Regione e/o Provincia autonoma e tipologia di scuola	Totale scuole		Totale alunni		Campione				
					Allocazione proporzionale tra le regioni		Allocazione uguale tra le regioni		Allocazione di compromesso $\beta=0,4$
	Scuole	CV%	Scuole	CV%	Scuole	CV%	Scuole	CV%	
Valle d'Aosta	4	213	4	—	4	—	4	—	
Piemonte	88	6.576	18	2,1	14	2,4	15	2,3	
Liguria	26	2.195	6	3,3	10	1,9	9	2,2	
Lombardia	162	13.890	28	1,2	10	2,2	16	1,7	
Prov. Aut. di Bolzano (l. it.)	3	243	3	—	3	—	3	—	
Prov. Aut. di Trento	2	175	2	—	2	—	2	—	
Veneto	91	8.417	19	1,6	12	2,2	14	2,0	
Friuli-Venezia Giulia	24	1.669	3	3,4	10	1,4	8	1,7	
Emilia-Romagna	100	7.175	20	2,1	14	2,6	16	2,4	
Toscana	84	5.651	13	2,1	11	2,3	11	2,3	
Umbria	24	1.372	2	4	10	1,5	8	1,9	
Marche	38	2.752	6	3,2	10	2,3	9	2,5	
Lazio	94	7.773	15	1,7	10	2,1	12	1,9	
Abruzzo	24	1.654	3	3,7	10	1,6	9	1,8	
Molise	10	459	2	7	10	2,3	5	4,1	
Campania	180	12.837	21	1,4	10	2	15	1,7	
Puglia	112	7.876	13	1,7	10	1,9	12	1,8	
Basilicata	23	1.162	2	4,8	10	1,7	8	2,1	
Calabria	66	3.470	6	2,6	10	1,8	9	2,0	
Sicilia	137	8.486	15	1,8	10	2,3	12	2,1	
Sardegna	39	2.269	5	3,9	11	2,6	9	2,9	
Italia	1.333	96.314	205	0,5	207	0,6	206	0,6	

Si notano i seguenti aspetti:

- l’allocazione proporzionale, porta al valore minimo del coefficiente di variazione della stima nazionale in tutti gli strati osservati. Come già detto per il grado 8 questo metodo porta a privilegiare le grandi regioni a scapito di quelle piccole. I valori dei CV sono in genere bassi, soprattutto negli strati molto grandi, tuttavia essi aumentano per i licei classici (dove 5 regioni sono superiori addirittura al 7%) e per gli istituti professionali (dove 3 regioni sono superiori al 4% e una, il Molise, superiore al 7%);
- l’allocazione uguale porta a un peggioramento generalizzato del CV a livello nazionale in tutte le tipologie tranne che in quella più grande degli istituti tecnici dove rimane uguale; migliorano molto gli errori nelle stime regionali;
- l’allocazione di compromesso tiene sotto controllo le stime nazionali e regionali pertanto per ogni strato relativo alla tipologia di istituto riusciamo ad avere valori ottimizzati.

6. Stime campionarie

6.1. Descrizione generale

A ciascun alunno viene attribuito un peso campionario che indica quanti scolari, non coinvolti nell'ICC sono rappresentati dall'alunno incluso nel campione. Per esempio, un peso pari a 50, indica che l'alunno rappresenta sé stesso più altri 49 scolari non coinvolti nell'indagine.

Il peso campionario (o semplicemente peso) di un alunno viene ottenuto moltiplicando:

- il peso della scuola (peso di primo stadio);
- il peso della sezione all'interno della scuola (peso di secondo stadio).

La somma del peso di secondo stadio su tutti gli alunni campione della scuola fa ritornare il numero di alunni della scuola.

Il peso di primo stadio viene ottenuto con una procedura di calibrazione che assicura che:

- la somma del peso di primo stadio su tutti i plessi campione di una determinata tipologia riproduce il numero dei plessi scolastici di quella determinata tipologia nella regione;
- la somma su tutti i plessi campione di una determinata tipologia dei prodotti del peso di primo stadio per il numero di scolari del plesso che frequentano il grado scolare r coincide il numero di alunni nello strato che frequentano il grado scolare r .

Si riporta a titolo esemplificativo per un grado scolastico e una sola materia (tab. 10) il controllo attuato in fase di calcolo dei pesi campionari al fine di verificare che il prodotto tra i coefficienti di riporto alla popolazione studentesca di indagine $wijk$ moltiplicati per il numero di studenti campione per regione equivalga al numero di studenti dell'intera popolazione.

Tab. 10 – Tavola differenza studenti pesati e popolazione – Italiano grado 8

Ripartizione geografica	Numero studenti pesati (ottenuti dal prodotto tra i coefficienti di riporto w_{ijk} moltiplicati per il numero di studenti campione)	Numero studenti popolazione
Valle D'Aosta	1.150	1.150
Piemonte	38.392	38.392
Liguria	12.205	12.205
Lombardia	94.762	94.762
Prov. Aut. Bolzano (l. it.)	1.377	1.377
Prov. Aut. Trento	5.418	5.418
Veneto	46.671	46.671
Friuli-Venezia Giulia	10.256	10.256
Emilia-Romagna	40.344	40.344
Toscana	32.811	32.811
Umbria	7.868	7.868
Marche	13.731	13.731
Lazio	52.380	52.380
Abruzzo	11.502	11.502
Molise	2.536	2.536
Campania	61.723	61.723
Puglia	39.138	39.138
Basilicata	4.826	4.826
Calabria	18.016	18.016
Sicilia	50.133	50.133
Sardegna	13.683	13.683

6.2. Peso di secondo stadio

Il peso di secondo stadio per la sezione j selezionata nella scuola i è dato da

$$w_{(h)j|i} = \frac{1}{\pi_{(h)j|i}} \frac{tr A_{(h)i}}{\sum_{j=1}^{S(h)i} \frac{1}{\pi_{(h)j|i}} tr A^*_{(h)ij}}, \quad [6.1]$$

in cui $tr A^*_{(h)ij}$ denota il numero di alunni rispondenti al campione per l'anno scolastico nella sezione j della scuola i di tipologia t nella regione h .

6.3. Peso di primo stadio

Il peso finale di primo stadio, $tr^W(h)_i$, viene ottenuto risolvendo il seguente problema di minimo vincolato

$$tr^W(h)_i = \begin{cases} \sum_{i=1}^{r^S(h)} D\left(\frac{1}{tr^{\pi(h)}_i}, tr^W(h)_i\right) = \min \\ \sum_{i=1}^{r^S(h)} tr^W(h)_i = tr^S(h) \\ \sum_{i=1}^{r^S(h)} tr^W(h)_i tr^A(h)_i = tr^A(h) \end{cases} \quad [6.2]$$

in cui $D\left(\frac{1}{tr^{\pi(h)}_i}, tr^W(h)_i\right) = tr^W(h)_i \ln\left(tr^W(h)_i tr^{\pi(h)}_i\right) - tr^W(h)_i + \frac{1}{tr^{\pi(h)}_i}$ è la distanza logaritmica tra il peso iniziale, $\frac{1}{tr^{\pi(h)}_i}$, e quello finale $tr^W(h)_i$.

L' algoritmo per la soluzione del problema [6.2] è illustrato in appendice A.4.

6.4. Peso alunno

Il peso finale dell' alunno ijk è dato da:

$$tr^W(h)_{ijk} = tr^W(h)_i tr^W(h)_{jk}.$$

6.5 Stima dei punteggi medi

Le stime sono ottenute mediante lo stimatore *rapporto separato* per strato. I valori medi regionali sono calcolati come

$$tr^{\hat{Y}}(h) = \frac{tr^{\hat{Y}}(h)}{tr^{\hat{A}}(h)} \quad [6.3]$$

dove

$$tr^{\hat{Y}}(h) = \sum_{i=1}^{r^S(h)} \sum_{j=1}^{r^S(h)_i} \sum_{k=1}^{r^A(h)^*_{ij}} tr^Y(h)_{ijk} tr^W(h)_{ijk}$$

$$tr^{\hat{A}}(h) = \sum_{i=1}^{r^S(h)} \sum_{j=1}^{r^S(h)_i} \sum_{k=1}^{r^A(h)^*_{ij}} tr^W(h)_{ijk}.$$

I valori medi nazionali sono quindi stimati da:

$$tr \hat{Y} = \frac{tr \hat{Y}}{tr \hat{A}} \quad [6.4]$$

essendo

$$tr \hat{Y} = \sum_{h=1}^{21} \sum_{i=1}^{tr S(h)} \sum_{j=1}^{tr S(h)i} \sum_{k=1}^{tr A^*(h)ij} tr Y(h)ijk \quad tr W(h)ijk$$

$$tr \hat{A} = \sum_{h=1}^{21} \sum_{i=1}^{tr S(h)} \sum_{j=1}^{tr S(h)i} \sum_{k=1}^{tr A^*(h)ij} tr W(h)ijk.$$

Sotto la [6.2] si ha $tr \hat{A}(h) = tr A(h)$ e $tr \hat{A} = tr A$.

Le stime dei punteggi medi relative all'anno 2019 per ripartizione geografica, grado scolare e materia indagata sono riportate nelle successive tabelle dalla 11 alla 18.

6.6. Aggregazioni di tipologie di indirizzo differenziate rispetto all'ambito indagato

La scuola secondaria di secondo grado italiana si caratterizza per la sua articolazione in indirizzi, anche profondamente differenti rispetto alle competenze *accademiche* (Italiano, Matematica e Inglese) misurate dalle prove INVALSI. Inoltre, mentre la classe II della scuola secondaria di secondo grado ricade ancora all'interno dell'obbligo scolastico, ciò non è più vero per l'ultimo anno della scuola secondaria. D'altra parte, le prove INVALSI si caratterizzano per essere finalizzate principalmente alla valutazione delle competenze di base che dovrebbero essere possedute da tutti gli studenti delle scuole italiane. Per questa ragione, le prove sono indifferenziate per il grado 10, mentre sono stati introdotti moderati elementi di differenziazione nelle prove di Matematica destinate al grado 13. Tuttavia, ciò non preclude la possibilità di esprimere gli esiti delle prove su una stessa scala di misura.

Ai fini comparativi, gli indirizzi sono così suddivisi:

- Italiano e Inglese: licei classici e scientifici, altri licei, istituti tecnici, istituti professionali;
- Matematica: licei scientifici, altri licei, istituti tecnici, istituti professionali.

Fig. 1 – Risultati per tipologia di istituto – Italiano grado 13

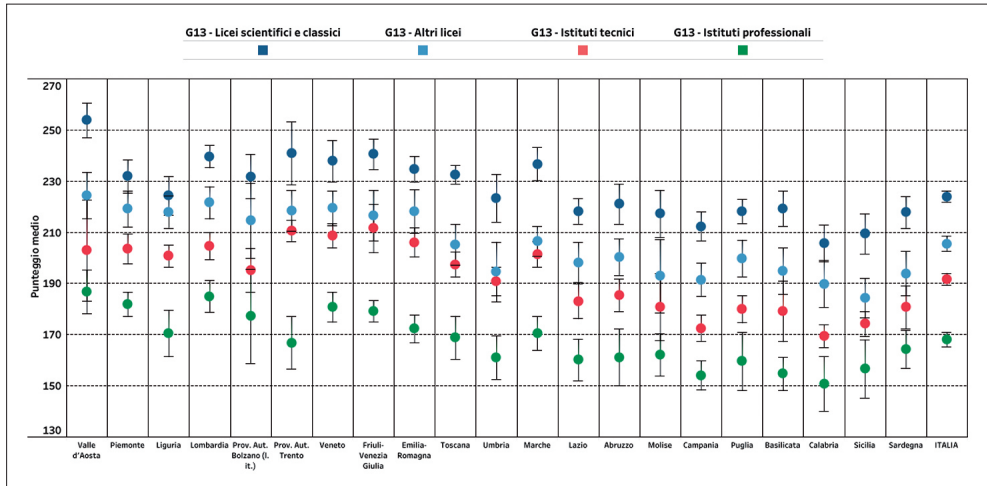
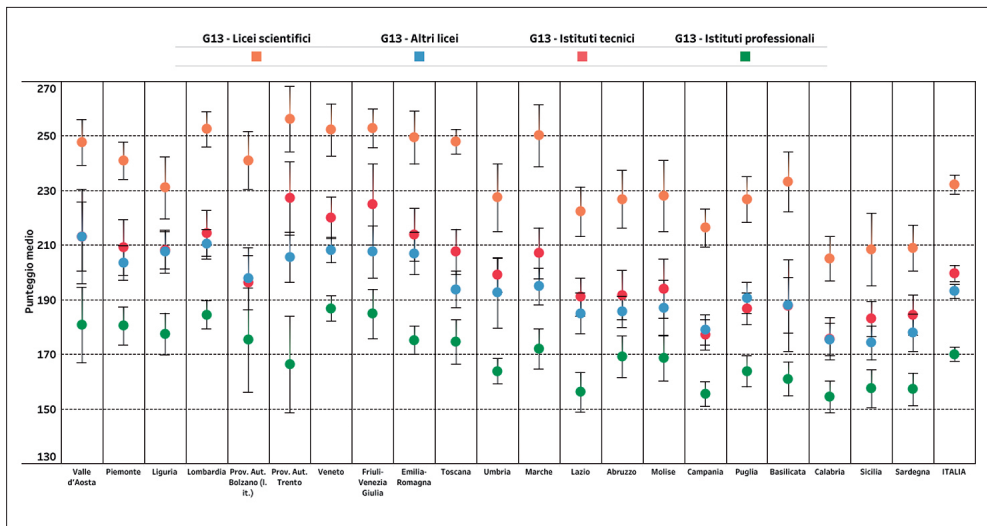


Fig. 2 – Risultati per tipologia di istituto – Matematica grado 13



La stima per una tipologia di indirizzo differenziata, indicata con t^* (per esempio i licei classici), viene ottenuta semplicemente come

$$t^*_r \hat{y} = \frac{t^*_r \hat{Y}}{t^*_r \hat{A}}$$

essendo

$${}_{t^*r}\hat{Y} = \sum_{h=1}^{21} \sum_{i \in {}_{t^*r}S(h)} \sum_{j=1}^{{}_{t^*r}S(h)i} \sum_{k=1}^{{}_{t^*r}A^*(h)ij} {}_{t^*r}Y(h)ijk \quad {}_{t^*r}W(h)ijk,$$

$${}_{t^*r}\hat{A} = \sum_{h=1}^{21} \sum_{i \in {}_{t^*r}S(h)} \sum_{j=1}^{{}_{t^*r}S(h)i} \sum_{k=1}^{{}_{t^*r}A^*(h)ij} {}_{t^*r}W(h)ijk,$$

in cui ${}_{t^*r}S(h)$ denota l'insieme di istituti campione nello strato h nella tipologia differenziata t^* .

Tab. 11 – Punteggi medi per Regione, per grado scolastico del I ciclo e materia indagata

Ripartizione geografica	Grado 2		Grado 5				Grado 8			
	Ita	Mat	Ita	Mat	Ere	Eli	Ita	Mat	Ere	Eli
Nord Ovest	200,5	201,1	203,7	203,7	203,2	201,2	203,3	206,6	209,6	209,4
Valle d'Aosta	208,3	206,6	208,6	202,6	208,6	215,1	205,7	208,3	214,0	217,4
Piemonte	204,3	202,0	204,3	203,3	200,3	196,9	203,0	204,4	207,7	205,7
Liguria	200,8	202,1	202,8	200,9	200,4	199,9	199,2	199,9	205,3	204,6
Lombardia	198,8	200,5	203,5	204,3	204,7	202,9	203,9	208,4	210,9	211,5
Nord Est	200,4	200,6	200,8	201,6	200,4	203,0	205,4	209,4	212,8	213,4
Prov. Aut. Bolzano (l. it.)	194,2	197,1	194,7	197,5	209,7	215,8	193,1	202,2	213,2	219,2
Prov. Aut. Trento	201,4	206,7	200,5	202,6	203,5	204,5	204,0	213,4	214,0	216,4
Veneto	200,3	202,4	200,5	202,6	200,7	202,3	205,5	209,3	212,7	212,4
Friuli-Venezia Giulia	203,5	203,4	201,1	202,5	202,4	203,9	206,8	212,4	214,9	217,0
Emilia-Romagna	199,8	197,3	201,3	200,4	198,9	202,9	205,5	208,4	212,2	213,1
Centro	201,1	201,5	202,9	203,2	202,0	204,9	201,5	203,4	207,1	205,9
Toscana	197,5	198,8	200,2	202,3	196,5	200,4	203,4	209,0	209,9	210,1
Umbria	211,5	211,7	210,0	207,6	205,9	206,8	200,2	203,4	203,9	203,2
Marche	208,6	210,6	210,0	208,3	206,4	207,4	207,7	209,2	213,8	211,1
Lazio	199,9	199,4	201,7	201,8	203,7	206,8	198,8	198,4	204,1	202,3
Sud	199,1	199,2	197,1	197,1	198,8	197,6	194,3	191,8	195,3	190,1
Abruzzo	204,2	200,8	203,4	199,7	198,3	198,5	201,3	200,6	201,9	199,3
Molise	209,2	211,6	205,3	206,8	202,3	203,9	197,7	194,1	197,8	193,3
Campania	195,7	196,7	193,4	193,4	196,6	195,6	190,2	186,5	191,5	185,7
Puglia	202,3	201,8	200,5	201,5	202,4	200,0	198,6	197,3	199,1	194,0
Sud e Isole	198,3	196,4	192,3	190,7	192,5	190,7	188,1	185,0	187,4	184,4
Basilicata	211,4	214,6	204,6	214,4	202,7	198,4	197,6	196,1	198,9	190,3
Calabria	198,2	192,2	190,2	185,9	190,5	188,8	185,8	180,9	187,9	184,4
Sicilia	197,6	196,9	190,8	190,3	193,7	193,2	186,4	184,5	185,2	182,5
Sardegna	196,7	193,9	196,2	190,1	187,1	181,5	193,8	188,3	190,8	189,5
Italia	200,0	200,0	200,0	200,0	200,0	200,0	199,1	200,1	203,3	201,6

Fig. 3 – Risultati per regione Italiano G8

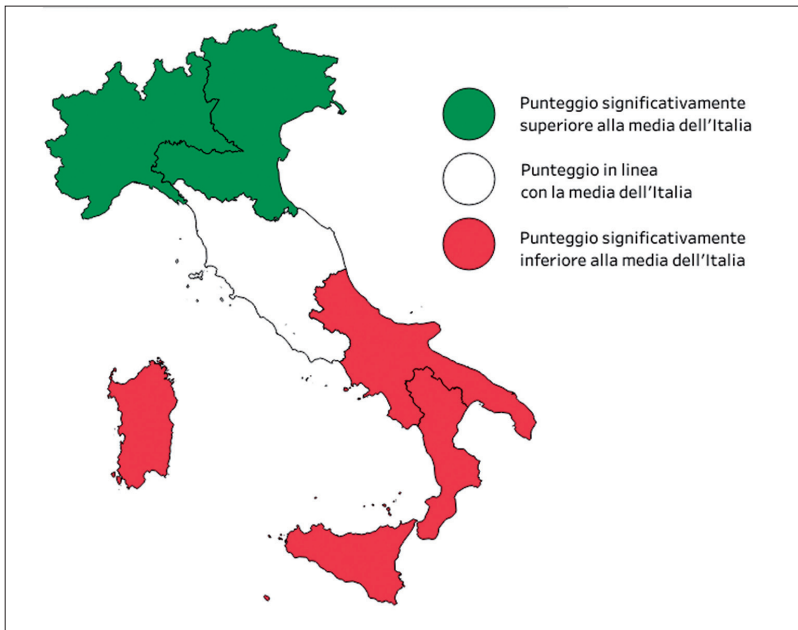


Fig. 4 – Risultati per regione Matematica G8

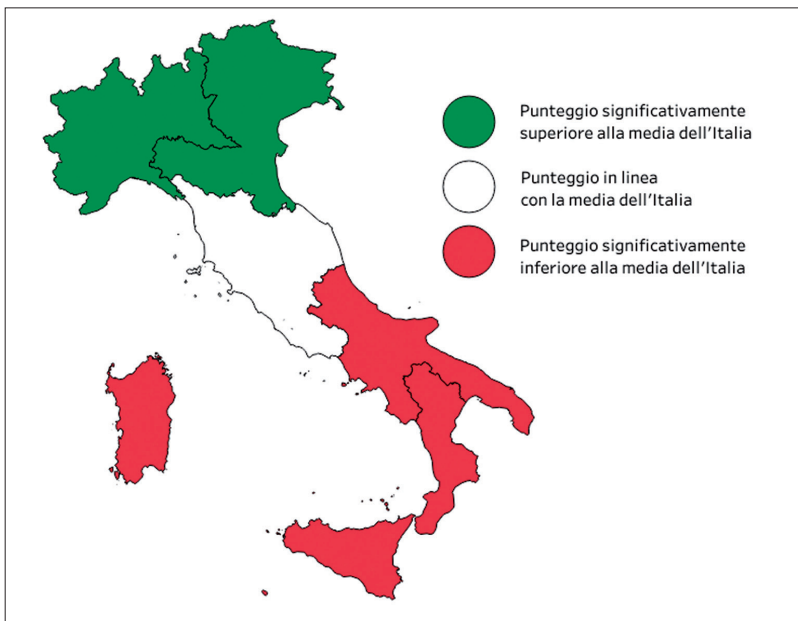


Fig. 5 – Risultati per regione Inglese Listening G8

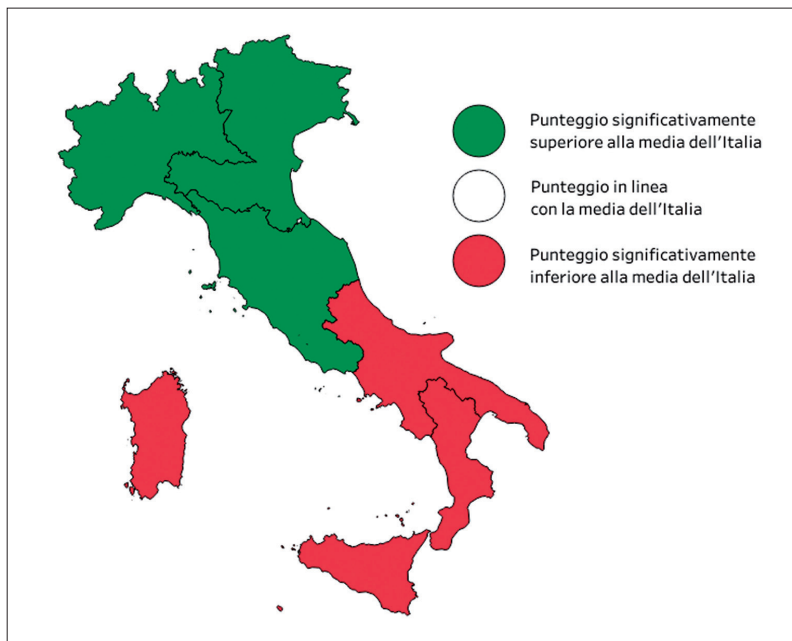
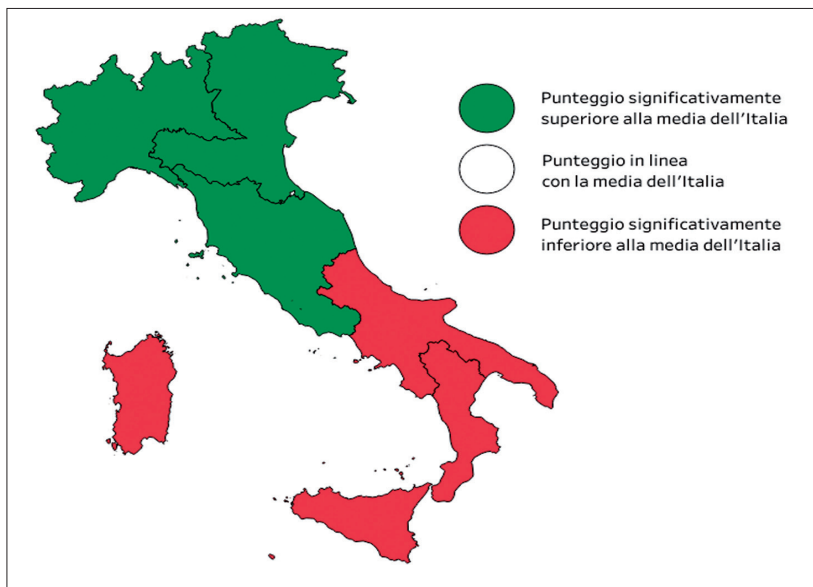


Fig. 6 – Risultati per regione Inglese Reading G8



Come si può vedere per il grado 8 e anche per il grado 10, i punteggi nazionali per materia sono ancorati alla metrica del 2018, quindi la media non è più 200, ma varia nel tempo, segnalando, quindi, un progresso o un regresso.

Di seguito un'esemplificazione del grado 8 che tiene in considerazione i punteggi sopra indicati e degli errori standard indicati nelle tabelle da 29 a 34 che permettono di calcolare le significatività statistiche delle differenze rispetto al Paese.

In Appendice è possibile osservare le immagini per tutti i gradi e le materie interessati dalle Rilevazioni.

Tab. 12 – Punteggi medi per Regione, per grado scolastico del II ciclo e materia indagata

<i>Ripartizione geografica</i>	<i>Grado 10</i>		<i>Grado 13</i>			
	<i>Ita</i>	<i>Mat</i>	<i>Ita</i>	<i>Mat</i>	<i>Ere</i>	<i>Eli</i>
<i>Nord Ovest</i>	213,9	214,7	212,8	213,8	212,6	215,4
Valle d'Aosta	217,7	212,3	216,9	212,8	212,1	218,8
Piemonte	210,3	211,5	210,5	209,2	207,7	210,8
Liguria	206,4	207,5	206,8	207,8	206,8	208,0
Lombardia	216,6	217,2	214,6	216,7	215,7	218,5
<i>Nord Est</i>	213,7	217,3	211,2	215,0	212,2	216,0
Prov. Aut. Bolzano (l. it.)	206,0	206,4	205,6	200,8	219,0	229,6
Prov. Aut. Trento	216,6	224,0	219,3	224,1	221,6	227,7
Veneto	216,5	220,1	211,8	216,5	213,2	215,8
Friuli-Venezia Giulia	213,0	218,1	214,1	219,1	215,5	220,3
Emilia-Romagna	210,8	213,5	208,8	211,5	208,7	213,1
<i>Centro</i>	205,1	203,4	199,8	198,7	200,2	202,0
Toscana	203,0	204,5	203,2	205,9	204,1	205,7
Umbria	205,5	207,4	197,3	198,4	195,8	194,6
Marche	209,8	212,2	205,4	205,1	202,7	204,6
Lazio	205,1	199,7	196,4	192,5	197,7	200,1
<i>Sud</i>	195,0	191,5	189,1	188,3	189,2	185,2
Abruzzo	203,6	202,8	197,2	196,4	193,9	191,3
Molise	199,1	198,4	191,5	196,0	189,1	185,0
Campania	192,5	187,7	185,9	183,6	187,7	183,0
Puglia	196,2	193,6	191,9	193,1	190,3	187,0
<i>Sud e Isole</i>	190,5	184,3	185,2	182,1	183,7	179,1
Basilicata	196,2	193,1	190,0	192,8	184,0	181,2
Calabria	188,8	183,7	181,8	179,5	180,4	177,2
Sicilia	191,7	184,2	184,1	181,4	184,9	178,0
Sardegna	187,0	181,8	192,2	183,3	184,4	184,8
Italia	204,1	202,8	200,0	200,0	200,0	200,0

Tab. 13 – Punteggi medi per regione, per grado 10 per tipologia di istituto – Matematica

<i>Ripartizione geografica</i>	<i>Licei scientifici</i>	<i>Altri licei</i>	<i>Istituti tecnici</i>	<i>Istituti professionali</i>
<i>Nord Ovest</i>	249,0	208,8	209,6	181,2
Valle d'Aosta	248,7	210,4	204,9	193,3
Piemonte	247,7	206,5	202,8	182,1
Liguria	241,0	202,1	205,0	172,8
Lombardia	250,9	211,0	212,9	182,1
<i>Nord Est</i>	250,7	211,4	215,9	184,2
Prov. Aut. Bolzano (l. it.)	241,7	199,9	205,2	175,3
Prov. Aut. Trento*	248,7	211,2	217,9	–
Veneto	253,3	215,6	219,4	191,1
Friuli-Venezia Giulia	243,6	208,9	220,4	187,0
Emilia-Romagna	250,5	208,1	210,7	174,0
<i>Centro</i>	233,7	195,7	199,4	170,2
Toscana	238,9	195,2	205,1	169,3
Umbria	241,8	194,1	203,4	177,1
Marche	248,7	203,7	213,9	175,2
Lazio	227,1	194,3	188,5	167,7
<i>Sud</i>	225,7	184,0	183,1	166,7
Abruzzo	235,3	190,3	195,0	173,8
Molise	234,8	196,2	187,0	169,8
Campania	221,0	178,5	179,1	165,1
Puglia	230,2	190,3	184,8	167,5
<i>Sud e Isole</i>	212,3	181,3	180,9	157,1
Basilicata	219,4	189,0	192,5	169,8
Calabria	215,8	182,8	175,9	150,9
Sicilia	210,7	179,8	183,7	157,4
Sardegna	209,4	182,0	174,2	158,1
Italia	234,6	196,5	199,4	171,7

* Dal 2013 la Provincia autonoma di Trento non ha più istituti professionali poiché tutti sono confluiti nella macro-categoria istituti di Istruzione e formazione professionale (IeFP).

Tab. 14 – Punteggi medi per regione, per grado 10 per tipologia di istituto – Italiano

<i>Ripartizione geografica</i>	<i>Licei scientifici e classici</i>	<i>Altri licei</i>	<i>Istituti tecnici</i>	<i>Istituti professionali</i>
<i>Nord Ovest</i>	236,9	218,8	203,9	182,3
Valle d'Aosta	242,8	223,3	207,2	193,6
Piemonte	236,1	214,2	196,8	183,9
Liguria	228,5	208,3	198,1	174,6
Lombardia	238,6	222,5	207,5	182,8
<i>Nord Est</i>	237,0	220,9	206,3	181,9
Prov. Aut. Bolzano (l. it.)	229,1	214,0	193,1	180,0
Prov. Aut. Trento*	234,7	213,6	204,7	–
Veneto	239,3	223,9	211,2	187,6
Friuli-Venezia Giulia	233,5	214,8	208,5	181,4
Emilia-Romagna	236,1	220,7	200,6	174,1
<i>Centro</i>	226,6	207,0	192,8	170,0
Toscana	226,3	205,5	196,2	168,3
Umbria	230,6	201,5	194,3	172,1
Marche	237,0	212,9	203,8	168,2
Lazio	224,3	207,2	185,7	171,8
<i>Sud</i>	220,7	196,2	181,7	167,2
Abruzzo	227,8	201,7	188,9	170,9
Molise	227,8	202,5	183,2	172,2
Campania	217,5	190,0	179,4	166,6
Puglia	223,4	203,9	182,6	167,0
<i>Sud e Isole</i>	212,0	195,3	182,3	158,6
Basilicata	222,6	195,4	187,4	168,8
Calabria	215,2	192,0	177,8	151,1
Sicilia	209,7	195,6	185,8	162,3
Sardegna	210,6	197,6	175,2	152,2
Italia	226,5	208,5	194,7	171,9

* Dal 2013 la Provincia autonoma di Trento non ha più istituti professionali poiché tutti sono confluiti nella macro-categoria istituti di Istruzione e formazione professionale (IeFP).

Tab. 15 – Punteggi medi per regione, per grado 13 per tipologia di istituto – Matematica

<i>Ripartizione geografica</i>	<i>Licei scientifici</i>	<i>Altri licei</i>	<i>Istituti tecnici</i>	<i>Istituti professionali</i>
<i>Nord Ovest</i>	247,1	208,3	212,5	182,5
Valle d'Aosta	247,6	213,2	213,2	180,7
Piemonte	241,0	203,5	209,2	180,4
Liguria	231,0	207,7	208,2	177,4
Lombardia	252,5	210,4	214,4	184,4
<i>Nord Est</i>	251,2	207,3	218,4	181,6
Prov. Aut. Bolzano (l. it.)	241,1	197,8	196,3	175,2
Prov. Aut. Trento	256,2	205,5	227,2	166,3
Veneto	252,2	208,3	220,0	186,8
Friuli-Venezia Giulia	252,8	207,5	224,9	184,8
Emilia-Romagna	249,4	206,9	213,8	175,2
<i>Centro</i>	232,7	189,2	199,6	165,5
Toscana	248,0	193,7	207,6	174,5
Umbria	227,4	192,5	199,0	163,8
Marche	250,1	194,8	207,0	172,0
Lazio	222,3	185,0	191,0	156,1
<i>Sud</i>	221,0	183,6	182,3	159,7
Abruzzo	226,9	185,6	191,7	169,0
Molise	228,0	187,0	194,0	168,7
Campania	216,3	178,9	177,1	155,4
Puglia	226,7	190,7	186,7	163,8
<i>Sud e Isole</i>	209,4	176,0	181,7	157,0
Basilicata	233,2	188,0	187,8	161,0
Calabria	205,0	175,5	175,7	154,5
Sicilia	208,4	174,2	183,0	157,4
Sardegna	208,9	177,8	184,4	157,2
Italia	232,2	193,0	199,7	169,9

Tab. 16 – Punteggi medi per regione, per grado 13 per tipologia di istituto – Italiano

<i>Ripartizione geografica</i>	<i>Licei scientifici e classici</i>	<i>Altri licei</i>	<i>Istituti tecnici</i>	<i>Istituti professionali</i>
<i>Nord Ovest</i>	235,9	220,6	204,1	182,6
Valle d'Aosta	253,8	224,4	202,9	186,8
Piemonte	231,9	219,1	203,5	181,8
Liguria	224,3	217,9	200,7	170,5
Lombardia	239,6	221,5	204,7	184,9
<i>Nord Est</i>	236,9	218,5	207,9	177,0
Prov. Aut. Bolzano (l. it.)	231,8	214,5	195,1	177,2
Prov. Aut. Trento	240,9	218,4	210,6	166,8
Veneto	237,9	219,5	208,7	180,7
Friuli-Venezia Giulia	240,5	216,6	211,5	179,1
Emilia-Romagna	234,5	218,1	206,0	172,3
<i>Centro</i>	224,2	201,2	191,3	164,8
Toscana	232,5	205,2	197,4	168,8
Umbria	223,3	194,5	190,7	161,0
Marche	236,7	206,5	201,3	170,5
Lazio	218,2	198,1	183,0	160,1
<i>Sud</i>	215,2	195,1	176,5	156,7
Abruzzo	221,0	200,3	185,4	161,1
Molise	217,2	192,9	180,7	162,2
Campania	212,3	191,5	172,5	154,0
Puglia	218,1	199,8	179,9	159,5
<i>Sud e Isole</i>	210,5	187,6	174,3	156,3
Basilicata	219,3	194,9	179,1	154,7
Calabria	205,6	189,8	169,4	150,8
Sicilia	209,4	184,3	174,2	156,6
Sardegna	217,8	193,9	180,7	164,3
Italia	224,0	205,5	191,6	168,1

Tab. 17 – Punteggi medi per regione, per grado 13 per tipologia di istituto – Inglese Reading

<i>Ripartizione geografica</i>	<i>Licei scientifici e classici</i>	<i>Altri licei</i>	<i>Istituti tecnici</i>	<i>Istituti professionali</i>
<i>Nord Ovest</i>	232,5	222,7	206,0	180,8
Valle d'Aosta	235,5	222,2	207,6	178,6
Piemonte	228,8	216,4	203,1	175,7
Liguria	222,2	219,5	201,9	169,7
Lombardia	235,9	225,9	207,7	185,3
<i>Nord Est</i>	234,2	221,5	210,7	177,3
Prov. Aut. Bolzano (l. it.)	240,4	223,9	212,8	195,8
Prov. Aut. Trento	239,8	225,0	212,0	165,0
Veneto	235,9	224,2	212,0	179,2
Friuli-Venezia Giulia	238,3	222,7	215,6	176,0
Emilia-Romagna	230,2	217,7	207,5	174,9
<i>Centro</i>	222,4	203,7	192,9	164,3
Toscana	234,8	205,2	199,3	167,2
Umbria	212,6	204,8	189,3	162,2
Marche	233,2	204,2	198,4	168,9
Lazio	216,2	202,5	186,8	160,4
<i>Sud</i>	213,1	197,0	178,0	156,0
Abruzzo	215,7	197,2	183,0	160,4
Molise	213,7	189,4	179,2	161,7
Campania	211,6	197,0	176,2	152,9
Puglia	214,8	197,6	179,2	159,7
<i>Sud e Isole</i>	207,3	185,3	174,6	155,8
Basilicata	212,5	195,1	169,5	149,2
Calabria	203,3	188,0	167,7	152,5
Sicilia	208,2	184,1	177,8	157,5
Sardegna	207,9	183,2	176,4	158,4
Italia	221,4	207,2	193,3	167,4

Tab. 18 – Punteggi medi per regione, per grado 13 per tipologia di istituto – Inglese Listening

<i>Ripartizione geografica</i>	<i>Licei scientifici e classici</i>	<i>Altri licei</i>	<i>Istituti tecnici</i>	<i>Istituti professionali</i>
<i>Nord Ovest</i>	235,3	225,1	208,4	184,7
Valle d'Aosta	242,7	229,5	211,8	188,4
Piemonte	232,0	220,2	204,8	180,3
Liguria	224,2	219,6	200,2	175,3
Lombardia	238,7	228,0	210,8	188,5
<i>Nord Est</i>	236,6	226,2	212,6	185,1
Prov. Aut. Bolzano (l. it.)	246,7	235,5	227,1	205,6
Prov. Aut. Trento	243,5	236,3	215,1	177,8
Veneto	237,2	225,5	213,9	185,9
Friuli-Venezia Giulia	241,9	227,8	218,1	186,2
Emilia-Romagna	233,1	224,5	208,8	183,4
<i>Centro</i>	222,3	208,1	192,6	169,5
Toscana	232,7	209,7	199,8	171,9
Umbria	213,9	204,2	182,7	163,0
Marche	233,3	209,1	198,4	172,4
Lazio	216,7	207,3	186,7	167,6
<i>Sud</i>	206,9	195,1	173,1	155,0
Abruzzo	210,9	196,2	180,9	159,6
Molise	206,0	189,2	173,8	161,8
Campania	202,9	193,6	171,3	154,1
Puglia	212,6	197,6	173,7	155,0
<i>Sud e Isole</i>	201,0	181,6	168,7	155,2
Basilicata	208,1	192,8	167,3	147,5
Calabria	197,6	185,8	165,3	152,1
Sicilia	200,1	179,7	167,7	155,1
Sardegna	206,0	179,6	179,0	164,3
Italia	219,6	208,6	192,2	170,5

Le tabelle precedenti (dalla 12 alla 18) illustrano i punteggi derivanti dalle Rilevazioni INVALSI in tutti i gradi e per tutte le materie indagate.

La robustezza del campione è mostrata dal fatto che l'ordinamento delle regioni nel tempo rimane quello noto anche da altre rilevazioni effettuate negli anni passati, nonché con gli esiti della popolazione (i cui grafici sono visibili nel paragrafo 8). Le regioni del Sud mostrano risultati più deboli, con delle differenze modeste nella scuola primaria che si accentuano via via che si va avanti con il grado scolastico.

6.7. Stima delle componenti della devianza

Devianza totale D

Una stima approssimativamente corretta della devianza totale $tr \hat{D}_{cor}$ è data da (vedi appendice A.5):

$$tr \hat{D}_{cor} = \sum_{h=1}^{21} \sum_{i=1}^{tr s(h)} tr w(h)_{ijk} \sum_{j=1}^{tr s(h)_i} \sum_{k=1}^{tr A_{ij}} y_{ijk}^2 - tr \hat{A} (tr \hat{y}^2 - \hat{V}(tr \hat{y})), \quad [6.5]$$

in cui la stima della varianza campionaria $\hat{V}(tr \hat{y})$ è definita dall'espressione [7.7].

Componente D3: devianza tra gli scolari della classe

È immediato dimostrare che una stima corretta della devianza tra gli alunni della classe è data da

$$tr \hat{D}3 = \sum_{h=1}^{21} \sum_{i=1}^{s(h)} \sum_{j=1}^{s(h)_i} \sum_{k=1}^{tr A_{ij}} tr w(h)_{ijk} (tr y(h)_{ijk} - tr \bar{y}(h)_{ij})^2.$$

Componente D1: devianza tra scuole

La stima approssimata della devianza tra scuole è data da (vedi appendice A.5)

$$tr \hat{D}1_{cor} = tr \hat{D}1 - \sum_{h=1}^{21} \sum_{i=1}^{tr s(h)} \ddot{V}(tr \hat{y}(h)_i) + tr A \hat{V}(tr \hat{y}), \quad [6.6]$$

dove

$$tr \hat{D}1 = \sum_{h=1}^{21} \sum_{i=1}^{s(h)} tr w(h)_i (tr \hat{A}(h)_i (tr \hat{y}(h)_i - tr \hat{y})^2),$$

in cui $\ddot{V}(\hat{y}_i)$ indica la stima della variabilità a livello d'istituto la cui espressione è riportata nell'appendice A.5.

Componente D2: devianza tra le classi

La componente D2 può essere stimata per differenza:

$$tr \hat{D}2 = tr \hat{D}_{cor} - (tr \hat{D}1_{cor} + tr \hat{D}3).$$

Le stime delle componenti della devianza relative all'anno 2019 per ripartizione geografica, grado scolare e materia sono riportate nelle seguenti tabelle dalla 19 alla 28.

Stime regionali

Le stime della devianza e delle relative componenti a livello regionale sono determinate mediante le espressioni precedenti, limitando le sommatorie alla specifica regione.

Tab. 19 – Componenti della devianza dei punteggi medi complessivi di Italiano per ripartizione geografica e grado scolare¹

Ripartizione	Grado 2			Grado 5			Grado 8			Grado 10		
	Perc. D1	Perc. D2	Perc. D3	Perc. D1	Perc. D2	Perc. D3	Perc. D1	Perc. D2	Perc. D3	Perc. D1	Perc. D2	Perc. D3
Nord Ovest	10,1	3,8	86,1	9,3	5,6	85,1	8,1	5,0	86,9	31,2	15,5	53,4
Nord Est	9,8	6,6	83,6	10,7	5,0	84,3	5,9	5,6	88,4	13,9	32,9	53,2
Centro	8,3	5,6	86,0	10,3	8,1	81,6	8,7	4,7	86,6	35,0	5,9	59,2
Sud	20,2	7,0	72,8	17,4	7,9	74,6	12,7	7,3	80,0	22,8	21,4	55,7
Sud e Isole	16,7	9,0	74,3	16,7	12,9	70,5	16,4	15,4	68,2	33,9	10,9	55,1
Italia	12,7	6,0	81,3	13,4	6,6	80,0	12,9	4,7	82,4	18,6	26,9	54,5

Tab. 20 – Componenti della devianza dei punteggi medi complessivi di Italiano per ripartizione geografica e tipologia di indirizzo per il grado 13

Ripartizione	Licei classici e scientifici			Altri licei			Istituti tecnici			Istituti professionali		
	Perc. D1	Perc. D2	Perc. D3	Perc. D1	Perc. D2	Perc. D3	Perc. D1	Perc. D2	Perc. D3	Perc. D1	Perc. D2	Perc. D3
Nord Ovest	7,6	59,2	33,2	13,2	31,8	55,0	17,1	7,7	75,2	10,7	26,3	62,9
Nord Est	7,6	58,2	34,3	14,4	26,6	59,1	14,2	12,1	73,6	8,3	36,4	55,4
Centro	12,1	37,1	50,8	16,7	7,4	75,9	19,7	13,5	66,9	8,3	54,7	37,0
Sud	12,7	21,1	66,2	13,3	9,3	77,4	12,6	37,7	49,6	6,9	67,5	25,7
Sud e Isole	21,2	13,9	64,9	14,9	17,7	67,4	9,3	43,0	47,7	10,8	66,4	22,9
Italia	17,1	35,5	47,4	26,2	7,5	66,3	28,8	10,4	60,8	13,9	48,9	37,2

¹ Perc. D1 = Perc. devianza tra scuole su devianza totale, Perc. D2 = Perc. devianza tra classi su devianza totale, Perc. D3 = Perc. devianza dentro le classi su devianza totale.

Tab. 21 – Componenti della devianza dei punteggi medi complessivi di Italiano per ripartizione geografica e tipologia di indirizzo per il grado 10

Ripartizione	Licei classici e scientifici			Altri licei			Istituti tecnici			Istituti professionali		
	Perc. D1	Perc. D2	Perc. D3	Perc. D1	Perc. D2	Perc. D3	Perc. D1	Perc. D2	Perc. D3	Perc. D1	Perc. D2	Perc. D3
Nord Ovest	6,9	60,1	33,0	11,7	27,6	60,7	14,4	6,0	79,6	6,8	28,2	65,0
Nord Est	5,1	61,7	33,2	8,6	34,6	56,8	9,9	8,9	81,3	8,5	32,6	58,9
Centro	9,2	41,8	49,0	17,1	9,5	73,4	11,4	10,1	78,5	5,0	52,3	42,8
Sud	8,6	30,5	60,9	15,9	8,6	75,5	8,2	30,2	61,6	4,7	55,9	39,4
Sud e Isole	13,9	17,5	68,6	12,4	9,6	78,0	13,5	27,6	58,9	4,4	68,6	27,0
Italia	12,7	40,7	46,6	21,8	10,6	67,6	20,7	7,5	71,9	9,8	47,2	43,0

Tab. 22 – Componenti della devianza dei punteggi medi complessivi di Matematica per ripartizione geografica e grado scolare

Ripartizione	Grado 2			Grado 5			Grado 8			Grado 10		
	Perc. D1	Perc. D2	Perc. D3	Perc. D1	Perc. D2	Perc. D3	Perc. D1	Perc. D2	Perc. D3	Perc. D1	Perc. D2	Perc. D3
Nord Ovest	11,0	6,4	82,6	10,0	8,4	81,6	8,9	7,4	83,7	40,6	16,6	42,9
Nord Est	11,7	7,8	80,5	12,6	6,9	80,4	7,5	9,7	82,8	33,9	21,0	45,2
Centro	12,0	8,3	79,7	12,0	11,5	76,5	10,4	6,4	83,2	43,3	5,4	51,3
Sud	31,2	7,8	60,9	30,8	9,3	59,9	15,3	11,0	73,7	41,5	8,7	49,8
Sud e Isole	22,9	11,0	66,1	27,0	14,0	58,9	16,7	20,2	63,1	32,9	21,2	46,0
Italia	17,6	7,9	74,6	19,0	8,5	72,5	16,7	5,2	78,1	48,7	4,4	46,9

Tab. 23 – Componenti della devianza dei punteggi medi complessivi di Matematica per ripartizione geografica e tipologia di Indirizzo per il grado 13

Ripartizione	Licei scientifici			Altri licei			Istituti tecnici			Istituti professionali		
	Perc. D1	Perc. D2	Perc. D3	Perc. D1	Perc. D2	Perc. D3	Perc. D1	Perc. D2	Perc. D3	Perc. D1	Perc. D2	Perc. D3
Nord Ovest	7,5	71,0	21,6	21,2	11,3	67,4	35,0	19,7	45,3	12,1	33,5	54,4
Nord Est	7,1	72,0	21,0	21,9	10,9	67,2	27,9	30,7	41,4	11,6	33,2	55,2
Centro	18,2	47,0	34,9	28,3	14,8	56,9	34,7	7,7	57,6	11,0	61,7	27,3
Sud	14,4	32,3	53,3	17,6	27,9	54,5	19,8	31,8	48,5	4,8	72,5	22,7
Sud e Isole	32,7	14,2	53,1	16,1	43,8	40,1	19,8	31,2	49,0	4,8	75,9	19,3
Italia	21,9	45,7	32,4	34,4	9,3	56,3	45,4	7,0	47,7	14,4	53,8	31,8

Tab. 24 – Componenti della devianza dei punteggi medi complessivi di Matematica per ripartizione geografica e tipologia di indirizzo per il grado 10

Ripartizione	Licei scientifici			Altri licei			Istituti tecnici			Istituti professionali		
	Perc. D1	Perc. D2	Perc. D3	Perc. D1	Perc. D2	Perc. D3	Perc. D1	Perc. D2	Perc. D3	Perc. D1	Perc. D2	Perc. D3
Nord Ovest	4,0	73,2	22,8	25,8	9,9	64,4	23,0	15,2	61,8	8,3	36,6	55,1
Nord Est	5,0	73,0	22,0	12,7	17,8	69,5	13,1	23,4	63,6	14,1	27,3	58,6
Centro	9,3	53,8	36,9	25,9	6,7	67,4	21,4	5,2	73,4	3,6	59,0	37,4
Sud	9,9	36,9	53,2	22,6	25,0	52,5	14,7	28,5	56,7	4,9	61,2	33,8
Sud e Isole	14,9	17,1	68,0	14,3	33,8	51,9	13,0	34,6	52,4	4,9	74,8	20,3
Italia	14,6	50,5	35,0	33,5	5,6	60,9	33,4	5,2	61,4	11,8	51,6	36,7

Tab. 25 – Componenti della devianza dei punteggi medi complessivi di Inglese Reading per ripartizione geografica e grado scolastico

Ripartizione	Grado 5			Grado 8		
	Perc. D1	Perc. D2	Perc. D3	Perc. D1	Perc. D2	Perc. D3
Nord Ovest	10,6	6,0	83,3	8,5	11,2	80,2
Nord Est	12,0	6,1	82,0	6,0	16,5	77,6
Centro	9,1	7,2	83,6	10,4	9,0	80,6
Sud	18,7	9,4	71,9	16,6	7,6	75,9
Sud e Isole	20,2	12,4	67,4	19,0	14,4	66,6
Italia	14,7	7,3	78,0	18,1	5,6	76,3

Tab. 26 – Componenti della devianza dei punteggi medi complessivi di Inglese Reading per ripartizione geografica e tipologia di indirizzo per il grado 13

Ripartizione	Licei scientifici			Altri licei			Istituti tecnici			Istituti professionali		
	Perc. D1	Perc. D2	Perc. D3	Perc. D1	Perc. D2	Perc. D3	Perc. D1	Perc. D2	Perc. D3	Perc. D1	Perc. D2	Perc. D3
Nord Ovest	10,1	57,9	32,0	20,4	37,0	42,6	14,6	10,5	74,9	10,9	31,7	57,4
Nord Est	7,6	62,2	30,3	16,8	36,2	47,0	17,6	17,1	65,3	10,2	38,2	51,6
Centro	13,4	36,0	50,6	22,5	8,4	69,1	21,1	13,4	65,5	9,1	57,9	33,1
Sud	13,2	19,7	67,1	23,4	12,0	64,6	15,0	36,2	48,8	6,7	72,6	20,7
Sud e Isole	20,7	13,2	66,2	28,7	20,0	51,3	14,3	41,8	44,0	8,4	69,8	21,8
Italia	18,8	33,3	47,9	34,7	10,8	54,5	31,1	10,5	58,5	13,9	53,0	33,1

Tab. 27 – Componenti della devianza dei punteggi medi complessivi di Inglese Listening per ripartizione geografica e grado scolastico

Ripartizione	Grado 5			Grado 8		
	Perc. D1	Perc. D2	Perc. D3	Perc. D1	Perc. D2	Perc. D3
Nord Ovest	17,9	8,1	74,0	12,5	12,6	74,9
Nord Est	17,4	10,7	71,9	7,0	19,1	73,9
Centro	17,9	12,9	69,2	11,8	8,4	79,8
Sud	34,7	12,8	52,5	14,9	13,3	71,7
Sud e Isole	36,6	20,2	43,2	16,9	21,8	61,3
Italia	26,1	11,5	62,4	22,3	5,2	72,5

Tab. 28 – Componenti della devianza dei punteggi medi complessivi di Inglese Listening per ripartizione geografica e tipologia di indirizzo per il grado 13

Ripartizione	Licei scientifici			Altri licei			Istituti tecnici			Istituti professionali		
	Perc. D1	Perc. D2	Perc. D3	Perc. D1	Perc. D2	Perc. D3	Perc. D1	Perc. D2	Perc. D3	Perc. D1	Perc. D2	Perc. D3
Nord Ovest	9,6	60,4	30,0	24,2	39,0	36,8	18,0	12,6	69,5	11,3	26,1	62,6
Nord Est	8,8	62,2	29,0	16,6	46,0	37,4	19,2	19,7	61,2	13,5	24,7	61,8
Centro	16,6	33,8	49,6	26,4	11,7	62,0	24,7	11,5	63,9	9,2	50,1	40,7
Sud	12,7	9,4	77,9	28,6	14,8	56,6	14,1	47,1	38,8	3,7	78,7	17,7
Sud e Isole	25,9	5,7	68,4	29,7	29,0	41,3	10,6	53,8	35,6	4,1	77,6	18,3
Italia	25,6	26,8	47,7	41,6	12,2	46,2	38,5	9,9	51,6	16,7	49,3	34,0

La variabilità tra scuole e tra classi (che vediamo rappresentata nelle tabelle dalla 19 alla 28 e nelle figure dalla 7 alla 10) è un indicatore di equità che riflette la capacità del sistema d'istruzione di assicurare agli alunni un eguale trattamento almeno nel tronco comune dell'itinerario scolastico, che in Italia corrisponde alla scuola primaria e alla secondaria di primo grado. In base al principio dell'eguaglianza di trattamento, le differenze di risultati tra gli alunni sono accettabili solo se essi hanno potuto usufruire di condizioni di insegnamento-apprendimento di qualità equivalente e dunque se non sono all'opera meccanismi, espliciti o impliciti, che determinano un'ineguale distribuzione degli studenti tra le scuole e tra le classi dal punto di vista del loro retroterra socio-culturale e delle abilità possedute. Un'alta variabilità tra scuole e tra classi è infatti il segno di un disequilibrio nella ripartizione degli studenti e della tendenza di quelli più capaci e più favoriti socialmente a raggrupparsi in certe scuole e classi e di quelli più deboli economicamente e culturalmente a concentrarsi in altre. Ciò non è senza conseguenze sui livelli di apprendimento che ogni alunno può raggiungere. La ricerca in campo educativo ha infatti constatato l'esistenza di quello che, con termine tecnico, si definisce "effetto di contesto" o di composizione del gruppo, per cui l'apprendimento di uno studente non è influenzato solo dalle sue caratteristiche personali ma anche, in una qualche misura, da quelle dei compagni con i quali si trova a interagire. Questo implica che il progresso di un alunno sarà

tanto minore quanto più il livello medio del gruppo del quale fa parte è basso e omogeneo. Inoltre, se l'effetto di contesto è, in termini diretti, un effetto dei compagni, esso comporta anche tutta una serie di effetti indiretti: per esempio, gli insegnanti agiscono in modo diverso a seconda degli alunni che hanno di fronte, adeguando a essi, in maniera più o meno consapevole, i propri comportamenti e le proprie metodologie d'insegnamento, così come i criteri di valutazione.

Fig. 7 – Variabilità dei risultati in Italiano – I ciclo di istruzione

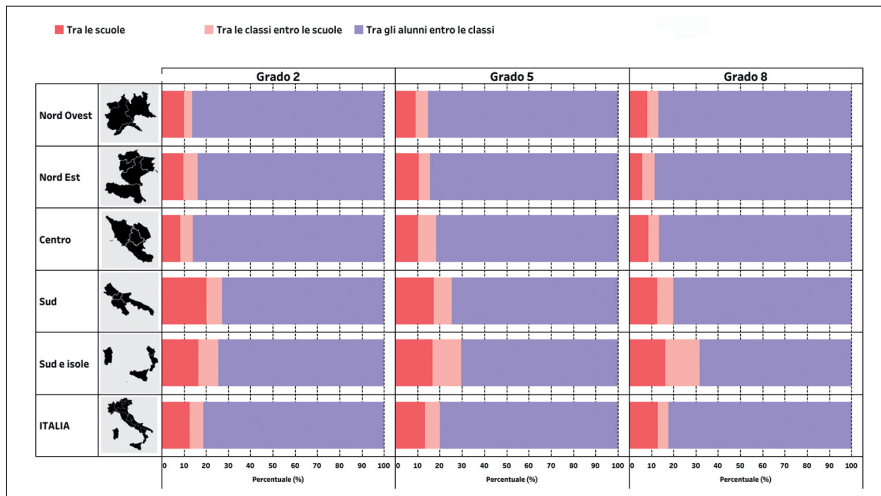


Fig. 8 – Variabilità dei risultati in Matematica – I ciclo di istruzione

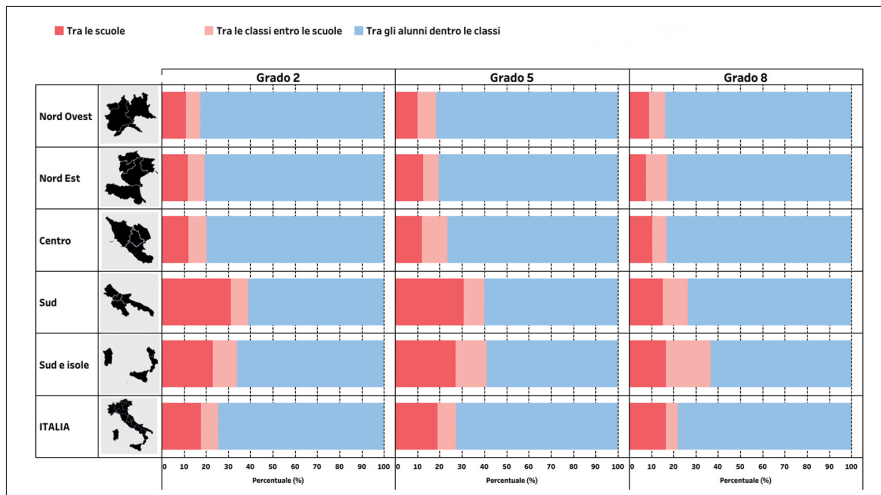


Fig. 9 – Variabilità dei risultati in Inglese Listening – I ciclo di istruzione

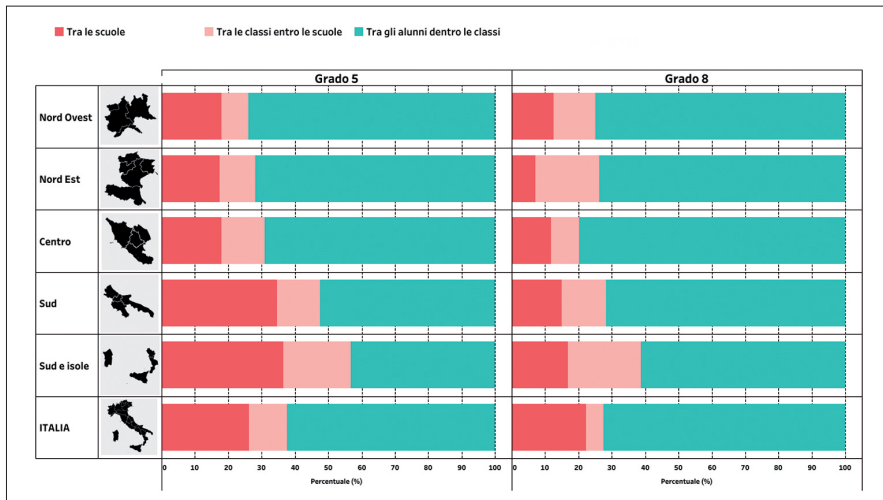
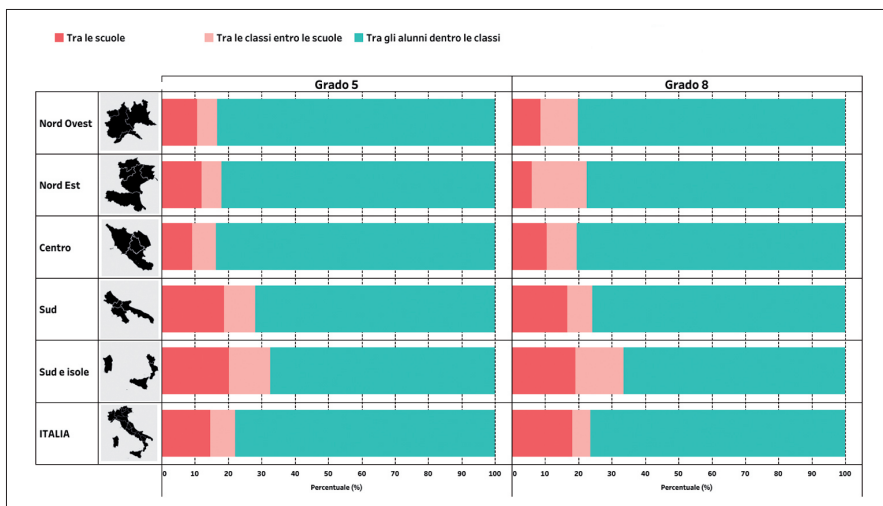


Fig. 10 – Variabilità dei risultati in Inglese Reading – I ciclo di istruzione



Infine si nota che, mentre la devianza totale è una quantità assoluta ottenuta come somma delle sue componenti, le quantità riportate nelle tabelle sono percentuali per cui può essere che la composizione di queste componenti possa variare a seconda dell'aggregazione considerata, per cui può verificarsi il caso che il valore percentuale a livello nazionale possa essere superiore della medesima percentuale calcolata a livello di una specifica area.

7. Variabilità campionaria

7.1. Espressione della varianza campionaria

Un'espressione approssimata della varianza campionaria dello stimatore definito nel capitolo 6 può essere derivata sulla base dei seguenti risultati:

- il teorema 11.2 (Cochran, 1977);
- il metodo di approssimazione lineare (Woodruff, 1971);
- il criterio d'approssimazione proposto da Deville e Tillè (2005), che permette di trascurare le probabilità di inclusione congiunte.

Su tali basi la varianza della stima regionale può essere espressa come:

$$\begin{aligned}
 V(\hat{y}_{(h)}) &= \frac{1}{tr A_{(h)}^2} V(\hat{Y}_{(h)}) \\
 &\cong \frac{1}{tr A_{(h)}^2} \left[\sum_{i=1}^{rS_{(h)}} \left(\frac{1}{tr \pi_{(h)i}} - 1 \right) tr Z_{(h)i}^2 + \sum_{i=1}^{rS_{(h)}} \frac{tr S_{(h)i} (tr S_{(h)i} - tr S_{(h)i})}{tr \pi_{(h)i} tr S_{(h)i}} tr \sigma_{z(h)i}^2 \right] \quad [7.1]
 \end{aligned}$$

essendo

$$tr Z_{(h)i} = \sum_{j=1}^{rS_{(h)i}} tr Z_{(h)ij}, \quad tr \sigma_{z(h)i}^2 = \frac{1}{tr S_{(h)i} - 1} \sum_{j=1}^{rS_{(h)i}} \left(tr Z_{(h)ij} - \frac{tr Z_{(h)i}}{tr S_{(h)i}} \right)^2,$$

$$tr Z_{(h)ij} = \sum_{k=1}^{tr A_{(h)ij}} tr z_{(h)ijk}$$

in cui

$$tr z_{(h)ijk} = tr y_{(h)ijk} - tr \bar{y}_{(h)}$$

rappresenta la variabile linearizzata riferita al singolo alunno.

La variabilità della stima nazionale può essere ottenuta come

$$V_{(tr, \hat{y})} \cong \frac{1}{tr A^2} \sum_{h=1}^{21} V_{(tr, \hat{Y}_{(h)})} \quad [7.2]$$

In appendice A.6 è illustrata una riscrittura della [7.1] utile al metterne in evidenza il termine $\rho_{(h)}$, semplificando in tal modo i calcoli necessari per la determinazione della numerosità campionaria.

7.2. Stima della varianza campionaria

La stima della varianza campionaria può essere determinata in modo semplice, adottando l'ipotesi che in primo stadio si sia effettuato un campionamento senza re-immissione.

Regioni NAR

Nel contesto appena introdotto, la stima della varianza campionaria della stima regionale, $tr \hat{y}_{(h)}$, è data da:

$$\hat{V}_{(tr, \hat{y}_{(h)})} = \frac{1}{tr A_{(h)}^2} \hat{V}_{(tr, \hat{Y}_{(h)})}, \quad [7.3]$$

essendo

$$\hat{V}_{(tr, \hat{Y}_{(h)})} = \frac{t^{S(h)}}{t^{S(h)} - 1} \sum_{i=1}^{t^{S(h)}} (tr \hat{Z}_{(h)i} - tr \hat{\bar{Z}}_{(h)})^2, \quad [7.4]$$

in cui

$$tr \hat{Z}_{(h)i} = \sum_{j=1}^{t^{S(h)i}} \sum_{k=1}^{tr A_{(h)ij}^*} tr z_{(h)ijk} \quad tr w_{(h)ijk} \quad tr \hat{\bar{Z}}_{(h)} = \frac{1}{t^{S(h)}} \sum_{i=1}^{t^{S(h)}} tr \hat{Z}_{(h)i}. \quad [7.5]$$

Regioni AR

Nelle regioni di tipo Auto Rappresentativo, la formula da applicare è la seguente:

$$\hat{V}(\text{tr} \hat{Y}(h)) = \frac{1}{\text{tr} A^2(h)} \sum_{i=1}^{t^S} \hat{V}(\text{tr} \hat{Y}_{(hi)}), \quad [7.6]$$

essendo

$$\hat{V}(\text{tr} \hat{Y}_{(hi)}) = \frac{t^S(h)i}{t^S(h)i-1} \sum_{j=1}^{t^S(h)i} (\text{tr} \hat{Z}_{(h)ij} - \text{tr} \hat{Z}_{(h)i})^2$$

in cui

$$\text{tr} \hat{Z}_{(h)ij} = \sum_{k=1}^{\text{tr} A^*(h)ij} \text{tr} z_{(h)ijk} \text{tr} w_{(h)ijk}, \quad \text{tr} \hat{Z}_{(h)i} = \frac{1}{t^S(h)i} \sum_{i=1}^{t^S(h)i} \text{tr} \hat{Z}_{(h)ij}.$$

Stima nazionale

La variabilità della stima nazionale può essere ottenuta come

$$\hat{V}(\text{tr} \hat{Y}) \cong \frac{1}{\text{tr} A^2} \sum_{h=1}^{21} \hat{V}(\text{tr} \hat{Y}_{(h)}). \quad [7.7]$$

Gli errori di campionamento stimati, espressi nei termini di coefficiente di variazione percentuale, presenti nelle tabelle dalla 4 alla 9 sono definiti da:

$$\hat{CV}(\text{tr} \hat{Y}_{(h)}) = 100 \sqrt{\hat{V}(\text{tr} \hat{Y}_{(h)})} / \text{tr} \hat{Y}_{(h)}$$

e

$$\hat{CV}(\text{tr} \hat{Y}) = 100 \sqrt{\hat{V}(\text{tr} \hat{Y})} / \text{tr} \hat{Y}.$$

La stima dei coefficienti di variazione costituisce la base per la determinazione degli intervalli di confidenza. Con probabilità approssimativamente pari a 0,95 i valori incogniti $\text{tr} \bar{Y}$ e $\text{tr} \bar{Y}_{(h)}$ dei parametri di interesse sono compresi nei seguenti intervalli:

$$\text{Prob} \left\{ \text{tr} \hat{Y} - 2 \text{tr} \hat{Y}_{(h)} [\hat{CV}\%(\text{tr} \hat{Y})/100] \leq \text{tr} \bar{Y} \leq \text{tr} \hat{Y} + 2 \text{tr} \hat{Y}_{(h)} [\hat{CV}\%(\text{tr} \hat{Y})/100] \right\} = 0,95;$$

$$\text{Prob} \left\{ \text{tr} \hat{Y}_{(h)} - 2 \text{tr} \hat{Y}_{(h)} [\hat{CV}\%(\text{tr} \hat{Y}_{(h)})/100] \leq \text{tr} \bar{Y}_{(h)} \leq \text{tr} \hat{Y}_{(h)} + 2 \text{tr} \hat{Y}_{(h)} [\hat{CV}\%(\text{tr} \hat{Y}_{(h)})/100] \right\} = 0,95.$$

Come si può notare dai valori riportati nelle tabelle dalla 29 alla 34, gli errori standard ottenuti sono piuttosto modesti, permettendo quindi di cogliere in modo adeguato le differenze, laddove esse esistono, tra una regione e un'altra. Ovviamente, come noto in letteratura, l'entità dell'errore standard dipende anche dalla consistenza numerica della popolazione di riferimento. Infatti, si nota che quest'ultimo tende a decrescere, all'aumentare della numerosità della popolazione scolastica dell'area territoriale considerata.

Tab. 29 – Errori standard per ripartizione geografica, grado scolastico – Italiano

Ripartizione geografica	Grado 2	Grado 5	Grado 8	Grado 10	G10 – Licei classici e scientifici	G10 – Altri licei	G10 – Istituti tecnici	G10 – Istituti profess.
<i>Nord Ovest</i>	1,2	1,2	1,0	1,7	1,8	2,0	1,9	1,8
Valle d'Aosta	2,0	1,9	1,3	4,8	7,5	5,1	4,0	7,1
Piemonte	2,6	2,2	1,5	2,9	2,8	3,3	2,1	2,6
Liguria	2,6	2,8	1,3	3,3	4,8	2,9	3,8	3,0
Lombardia	1,5	1,5	1,4	2,3	2,4	2,7	2,6	2,6
<i>Nord Est</i>	1,3	1,3	0,8	1,5	1,5	1,7	1,6	1,8
Prov. Aut. Bolzano (l. it.)	0,0	0,0	2,8	5,6	7,2	7,0	3,2	6,7
Prov. Aut. Trento	2,3	2,0	1,4	3,2	4,1	3,8	4,3	–
Veneto	2,2	1,9	1,1	2,3	2,3	2,5	1,9	1,8
Friuli-Venezia Giulia	2,5	3,0	1,6	3,0	4,0	4,7	2,8	4,1
Emilia-Romagna	2,0	2,5	1,6	2,8	2,5	2,9	2,6	2,5
<i>Centro</i>	1,1	1,2	1,0	1,6	1,8	2,2	1,6	1,9
Toscana	1,7	2,0	1,4	2,8	3,3	3,6	2,8	3,8
Umbria	1,9	2,8	2,0	3,6	1,5	4,5	2,4	3,5
Marche	2,6	2,1	1,8	3,5	3,0	4,9	2,6	2,6
Lazio	1,6	1,9	1,6	2,6	2,5	3,8	1,9	3,0
<i>Sud</i>	1,7	1,5	1,2	1,8	1,6	2,4	1,5	2,0
Abruzzo	2,5	2,3	1,8	3,6	2,6	5,5	3,3	2,9
Molise	2,1	2,3	2,5	5,2	6,3	3,0	3,2	3,7
Campania	2,6	2,3	1,8	2,6	2,1	3,4	2,0	3,4
Puglia	2,7	2,5	1,8	3,0	2,9	2,4	2,7	1,9
<i>Sud e Isole</i>	1,7	1,7	1,5	1,8	1,8	2,0	2,2	2,0
Basilicata	2,6	2,5	2,1	4,4	4,4	6,5	5,1	6,7
Calabria	3,0	2,3	2,5	3,9	2,9	4,0	2,5	4,4
Sicilia	2,7	2,7	2,3	2,5	2,8	3,2	3,5	2,3
Sardegna	1,3	2,7	2,0	3,9	3,5	2,7	4,2	1,8
Italia	0,6	0,6	0,5	0,8	1,0	1,3	1,1	1,2

Tab. 30 – Errori standard per ripartizione geografica, grado scolastico – Italiano

<i>Ripartizione geografica</i>	<i>Grado 13</i>	<i>G13 – Licei classici e scientifici</i>	<i>G13 – Altri licei</i>	<i>G13 – Istituti tecnici</i>	<i>G13 – Istituti professionali</i>
<i>Nord Ovest</i>	1,7	1,8	2,2	1,9	2,1
Valle d'Aosta	8,0	3,5	4,7	10,1	4,4
Piemonte	2,6	3,3	3,6	3,0	2,4
Liguria	3,3	3,9	3,2	2,2	4,7
Lombardia	2,5	2,2	3,2	2,7	3,2
<i>Nord Est</i>	1,7	2,1	2,3	1,7	1,9
Prov. Aut. Bolzano (l. it.)	6,0	4,5	7,5	4,4	9,4
Prov. Aut. Trento	3,4	6,3	4,1	2,2	5,2
Veneto	2,8	4,1	3,5	2,5	3,0
Friuli-Venezia Giulia	3,8	3,1	5,0	4,9	2,1
Emilia-Romagna	2,9	2,6	4,4	2,9	2,7
<i>Centro</i>	1,7	1,9	2,4	2,0	2,4
Toscana	3,1	1,9	4,0	2,5	4,3
Umbria	4,2	4,8	6,0	2,8	4,3
Marche	3,5	3,3	3,0	2,5	3,3
Lazio	2,8	2,6	4,0	3,5	4,2
<i>Sud</i>	2,0	1,9	2,3	1,8	2,6
Abruzzo	3,5	4,0	3,7	3,2	5,7
Molise	5,0	4,7	7,3	6,7	4,2
Campania	2,9	2,9	3,3	2,6	2,9
Puglia	3,3	2,4	3,7	2,7	5,8
<i>Sud e Isole</i>	1,9	2,4	2,6	1,7	3,4
Basilicata	4,7	3,5	4,6	6,0	3,3
Calabria	3,4	3,7	4,7	2,2	5,5
Sicilia	2,9	4,0	3,9	2,5	5,8
Sardegna	3,3	3,2	4,4	4,3	3,8
Italia	0,9	1,1	1,5	1,2	1,5

Tab. 31 – Errori standard per ripartizione geografica, grado scolastico – Matematica

Ripartizione geografica	Grado 2	Grado 5	Grado 8	Grado 10	G10 – Licei scientifici	G10 – Altri licei	G10 – Istituti tecnici	G10 – Istituti profess.
<i>Nord Ovest</i>	1,3	1,2	1,1	2,1	1,7	2,3	2,5	1,8
Valle d'Aosta	2,8	3,6	1,8	5,9	6,1	6,9	9,1	9,9
Piemonte	2,5	2,3	1,9	3,6	2,9	4,8	3,1	2,9
Liguria	2,7	2,4	2,0	4,1	7,0	3,2	5,1	2,7
Lombardia	1,7	1,6	1,4	2,8	2,2	3,1	3,5	2,6
<i>Nord Est</i>	1,3	1,4	1,0	1,8	2,2	1,5	1,9	2,3
Prov. Aut. Bolzano (l. it.)	0,0	0,0	2,9	6,7	7,5	7,3	5,4	5,0
Prov. Aut. Trento	2,3	2,1	1,9	3,9	6,1	3,8	4,2	–
Veneto	2,3	1,8	1,2	2,7	2,6	2,3	2,4	2,7
Friuli-Venezia Giulia	2,6	2,9	1,9	3,7	5,2	3,7	5,3	4,7
Emilia-Romagna	2,1	2,9	2,0	3,2	4,3	2,1	3,3	2,5
<i>Centro</i>	1,3	1,3	1,1	1,9	2,4	2,1	2,1	1,4
Toscana	2,0	2,2	1,6	3,4	3,2	2,8	3,0	2,9
Umbria	1,9	3,1	2,1	4,3	2,2	3,7	3,1	2,7
Marche	2,1	2,1	2,0	3,9	4,1	4,2	4,1	2,0
Lazio	2,2	2,1	1,7	3,1	3,5	3,6	2,4	2,0
<i>Sud</i>	2,3	2,1	1,4	2,1	2,3	2,4	1,9	1,9
Abruzzo	3,0	3,1	1,9	4,0	3,2	3,7	3,5	3,0
Molise	3,9	2,8	2,1	5,5	6,7	4,9	4,4	2,7
Campania	3,7	2,9	2,1	3,1	2,7	3,7	2,3	2,7
Puglia	3,2	3,9	1,9	3,4	4,8	2,4	3,5	3,1
<i>Sud e Isole</i>	2,0	2,2	1,5	1,8	2,3	1,7	2,0	2,0
Basilicata	2,5	3,5	2,0	4,2	4,0	4,8	5,8	7,4
Calabria	2,9	2,7	2,5	3,8	4,8	3,4	2,7	3,2
Sicilia	3,2	3,6	2,4	2,7	3,7	2,6	3,1	3,0
Sardegna	2,3	2,3	2,0	3,4	3,2	2,9	3,7	4,4
Italia	0,7	0,7	0,7	1,0	1,4	1,3	1,4	1,2

Tab. 32 – Errori standard per ripartizione geografica, grado scolastico – Matematica

Ripartizione geografica	Grado 13	G13 – Licei scientifici	G13 – Altri licei	G13 – Istituti tecnici	G13 – Istituti profess.
<i>Nord Ovest</i>	2,1	2,5	2,0	3,1	1,9
Valle d'Aosta	6,7	4,3	6,5	8,8	7,1
Piemonte	3,3	3,5	3,3	5,2	3,6
Liguria	3,4	5,8	4,0	3,5	3,9
Lombardia	3,0	3,3	2,8	4,3	2,6
<i>Nord Est</i>	2,1	2,9	1,9	2,7	1,9
Prov. Aut. Bolzano (l. it.)	6,8	5,4	5,8	5,1	9,7
Prov. Aut. Trento	5,7	6,1	4,7	6,8	9,0
Veneto	3,2	4,9	2,4	3,9	2,4
Friuli-Venezia Giulia	4,7	3,6	4,9	7,6	4,6
Emilia-Romagna	3,7	4,9	3,9	4,9	2,7
<i>Centro</i>	2,1	3,3	2,4	2,4	2,6
Toscana	3,8	2,3	3,4	4,2	4,2
Umbria	4,6	6,3	6,6	3,2	2,3
Marche	4,4	5,8	3,4	4,8	3,7
Lazio	3,3	4,7	3,8	3,6	3,7
<i>Sud</i>	2,0	2,6	2,1	2,0	1,8
Abruzzo	3,9	5,4	2,9	4,6	3,9
Molise	5,3	6,6	5,2	5,5	4,3
Campania	3,0	3,6	2,8	2,8	2,3
Puglia	3,3	4,2	2,9	3,0	2,9
<i>Sud e Isole</i>	1,9	3,8	2,1	2,2	2,1
Basilicata	5,3	5,6	5,2	8,6	3,2
Calabria	3,1	4,1	3,0	3,9	3,0
Sicilia	3,0	6,7	3,2	3,2	3,6
Sardegna	3,2	4,3	3,5	3,7	3,0
Italia	1,1	1,8	1,3	1,5	1,3

Tab. 33 – Errori standard per ripartizione geografica, grado scolastico – Inglese Reading

<i>Ripartizione geografica</i>	<i>Grado 5</i>	<i>Grado 8</i>	<i>Grado 13</i>	<i>G13 – Licei classici e scientifici</i>	<i>G13 – Altri licei</i>	<i>G13 – Istituti tecnici</i>	<i>G13 – Istituti profess.</i>
<i>Nord Ovest</i>	1,2	1,0	1,7	2,0	2,9	1,9	2,0
Valle d'Aosta	2,0	2,4	6,3	6,9	9,1	10,2	5,6
Piemonte	2,1	1,6	2,9	2,8	5,8	2,9	2,9
Liguria	2,9	1,7	3,5	4,4	4,9	2,6	4,5
Lombardia	1,6	1,3	2,4	2,8	3,8	2,5	2,6
<i>Nord Est</i>	1,3	0,8	1,8	1,9	2,6	2,0	2,0
Prov. Aut. Bolzano (l. it.)	0,0	3,4	6,6	4,6	14,7	5,8	11,2
Prov. Aut. Trento	2,5	1,8	3,9	3,6	7,9	1,9	8,0
Veneto	2,1	1,1	2,9	3,2	3,8	3,1	3,3
Friuli-Venezia Giulia	2,7	1,8	3,9	3,0	4,0	4,9	4,3
Emilia-Romagna	2,4	1,5	2,8	2,9	4,8	3,5	2,8
<i>Centro</i>	1,1	1,1	1,8	2,0	3,0	2,2	2,5
Toscana	1,8	1,7	3,3	1,5	4,6	2,8	4,4
Umbria	2,6	2,1	3,6	4,9	3,5	3,0	1,9
Marche	2,0	2,0	3,8	2,6	5,4	4,0	4,8
Lazio	1,7	1,7	2,9	2,7	5,3	4,0	4,3
<i>Sud</i>	1,7	1,4	2,0	2,0	3,2	2,1	2,4
Abruzzo	2,4	1,6	3,4	3,6	5,2	3,4	5,7
Molise	2,1	2,8	4,8	4,6	8,1	6,7	4,9
Campania	2,5	2,2	3,1	3,2	5,1	3,2	3,3
Puglia	2,9	2,1	3,1	2,2	4,9	3,2	4,2
<i>Sud e Isole</i>	1,9	1,7	2,0	2,3	4,1	2,3	2,9
Basilicata	3,3	2,3	4,6	2,6	5,1	6,4	3,8
Calabria	2,5	3,0	3,5	4,3	6,4	3,0	4,0
Sicilia	3,1	2,7	3,1	3,7	6,4	3,5	4,7
Sardegna	3,6	2,3	3,6	4,2	7,5	4,8	6,2
Italia	0,7	0,7	0,9	1,1	1,8	1,3	1,4

Tab. 34 – Errori standard per ripartizione geografica, grado scolastico – Inglese Listening

Ripartizione geografica	Grado 5	Grado 8	Grado 13	G13 – Licei classici e scientifici	G13 – Altri licei	G13 – Istituti tecnici	G13 – Istituti profess.
<i>Nord Ovest</i>	1,5	1,1	1,8	2,1	3,3	2,1	1,8
Valle d'Aosta	3,7	2,5	6,1	2,3	9,9	8,6	7,3
Piemonte	2,4	2,1	3,0	3,3	6,6	3,0	2,9
Liguria	3,5	1,5	3,5	4,3	7,0	2,6	3,7
Lombardia	2,1	1,4	2,5	2,8	4,3	2,9	2,4
<i>Nord Est</i>	1,6	0,9	1,7	2,2	2,8	2,2	2,1
Prov. Aut. Bolzano (I. it.)	0,0	2,7	6,5	3,6	13,6	6,2	15,6
Prov. Aut. Trento	2,6	1,9	4,6	5,2	9,2	3,7	5,7
Veneto	2,3	1,3	2,8	3,7	4,0	3,7	3,5
Friuli-Venezia Giulia	2,9	1,7	3,5	3,9	4,3	4,0	2,5
Emilia-Romagna	3,1	1,6	2,8	3,4	5,2	3,3	3,1
<i>Centro</i>	1,6	1,1	1,8	2,2	3,3	2,4	2,3
Toscana	2,8	1,9	3,1	2,7	4,3	3,0	4,3
Umbria	3,4	1,9	3,9	5,0	5,4	3,6	2,6
Marche	2,9	1,6	3,6	2,5	5,6	3,3	4,3
Lazio	2,6	1,6	3,0	3,2	6,1	4,4	4,1
<i>Sud</i>	2,4	1,3	1,9	1,7	3,7	2,0	1,7
Abruzzo	3,1	1,7	3,3	4,4	5,9	3,5	3,1
Molise	2,6	3,1	4,6	5,1	9,2	6,5	3,9
Campania	3,7	1,9	2,8	2,4	5,9	3,2	1,4
Puglia	3,6	1,8	3,2	1,9	5,0	2,8	4,4
<i>Sud e Isole</i>	2,8	1,5	1,9	2,6	4,1	2,0	1,7
Basilicata	4,5	2,0	4,4	3,0	6,4	4,6	3,3
Calabria	3,2	2,8	3,3	4,3	6,3	3,5	3,4
Sicilia	4,6	2,2	2,9	4,2	6,0	3,1	2,3
Sardegna	3,8	2,1	3,6	4,7	9,4	4,2	4,4
Italia	0,9	0,7	1,0	1,3	2,0	1,5	1,4

7.3. Un modello per la presentazione sintetica degli errori campionari

In questo paragrafo riportiamo alcuni modelli per la presentazione sintetica degli errori campionari. Essi consentono di determinare in modo semplice le varianze campionarie delle stime dei punteggi medi che possono essere calcolati per specifici sottoinsiemi di studenti, esempio, le femmine o i maschi e gli alunni delle classi paritarie.

I maschi e le femmine costituiscono un esempio di sottoinsiemi che si distribuiscono con proporzioni approssimativamente uguali nel territorio; mentre gli alunni delle scuole paritarie o pubbliche costituiscono un esempio di sottoinsiemi che si

distribuiscono in modo diseguale. I modelli utilizzati per la presentazione sintetica degli errori campionari differiscono tra i due tipi di sottoinsiemi in quanto quelli che si distribuiscono in modo disuguale presentano una componente aggiuntiva di varianza (dovuta appunto alla disomogeneità della distribuzione) di cui bisogna tenere conto nella modellizzazione della varianza.

7.3.1. Statistica deff e varianza campionaria

Nel caso si voglia conoscere l'errore campionario per una statistica differente da quelle riportate nelle tabelle precedenti, si può fare riferimento ai modelli per la presentazione sintetica degli errori di campionamento basati sulla statistica *deff* (Kish, 1965) espressa come:

$${}_{tr} deff(h) = \frac{\hat{V}({}_{tr}\hat{y}(h))}{\hat{V}_{srs}({}_{tr}\hat{y}(h))}, \quad [7.8]$$

in cui $\hat{V}_{srs}({}_{tr}\hat{y}(h))$ denota la varianza della stima ottenuta sotto il campionamento casuale semplice, espressa come

$$\hat{V}_{srs}({}_{tr}\hat{y}(h)) = \frac{{}_{tr}\hat{\sigma}^2(h)}{{}_{tr}a(h)},$$

in cui

$${}_{tr}\hat{\sigma}^2(h) = \frac{1}{{}_{tr}a(h)} \sum_{i=1}^{{}_{tr}s(h)} \sum_{j=1}^{{}_{tr}s(h)} \sum_{k=1}^{{}_{tr}A^*(h)ij} ({}_{tr}y(h)_{ijk} - {}_{tr}\bar{y}(h))^2$$

$${}_{tr}\bar{y}(h) = \frac{1}{{}_{tr}a(h)} \sum_{i=1}^{{}_{tr}s(h)} \sum_{j=1}^{{}_{tr}s(h)} \sum_{k=1}^{{}_{tr}A^*(h)ij} {}_{tr}y(h)_{ijk}$$

essendo ${}_{tr}a(h)$ il numero di alunni campione rispondenti nella regione h per l'anno r e la tipologia di scuola t .

La statistica *deff* può anche essere espressa come:

$${}_{tr} deff(h) = 1 + {}_{tr}\rho(h) ({}_{tr}\bar{a}(h) - 1),$$

essendo ${}_{tr}\rho(h)$ il coefficiente di correlazione intraclasse e

$${}_{tr}\bar{a}(h) = \frac{{}_{tr}a(h)}{{}_{tr}s(h)}$$

il numero medio di alunni campione per istituto.

Disponendo di una stima del $deff$ e avendo calcolato $tr a(h)$, la correlazione intra-classe può essere stimata come:

$$tr \rho(h) = \frac{tr deff(h) - 1}{tr \bar{a}(h) - 1}.$$

Box 7.1 – Formula per le regioni AR (Valle d’Aosta e Bolzano)

La precedente relazione vale per le regioni NAR, nelle rimanenti regioni AR la formula da applicare è la seguente:

$$tr deff(h) = 1 + tr \rho(h) (tr \bar{a}(h)[ij] - 1)$$

essendo $tr \rho(h)$ il coefficiente di correlazione intraclasse e

$$tr \bar{a}(h)[ij] = \frac{tr a(h)}{\sum_{i=1}^{tr S(h)} tr S(h)}$$

il numero medio di alunni campione per classe.

Disponendo di una stima del $deff$ e avendo calcolato $tr \bar{a}(h)[ij]$, la correlazione intraclasse può essere stimata come:

$$tr \rho(h) = \frac{tr deff(h) - 1}{tr \bar{a}(h)[ij] - 1}$$

La varianza della stima può quindi essere espressa come:

$$\hat{V}(tr \hat{y}(h)) = \frac{1}{tr a(h)} \frac{(tr A(h) - tr a(h))}{tr A(h)} tr \hat{\sigma}^2(h) [1 + tr \rho(h) (tr \bar{a}(h) - 1)]. \quad [7.9]$$

La varianza di una stima nazionale, può quindi essere determinata mediante semplice aggregazione delle stime regionali:

$$\hat{V}(tr \hat{y}) = \sum_{h=1}^{21} \hat{V}(tr \hat{y}(h)). \quad [7.10]$$

Le precedenti espressioni [7.9] e [7.10] costituiscono la base per la determinazione di modelli per la presentazione sintetica degli errori per sottoclassi. La teoria per i sottoinsiemi distribuiti in modo approssimativamente uguale è riportata nel successivo paragrafo 7.3.2; mentre quella relativa a sottoinsiemi distribuiti in modo diseguale è illustrata nel paragrafo 7.3.3. Infine, nel paragrafo 7.3.4 viene descritto come un utente che disponga dei micro-dati dell’indagine possa calcolare autonomamente le varianze campionarie delle stime.

7.3.2. Modelli per sottopopolazioni distribuite in modo approssimativamente uguale

Sia b una particolare sottoclasse distribuita in modo approssimativamente uguale nel territorio, e sia ${}_{tr}m_{b(h)}$ il numero di alunni campione rispondenti appartenenti alla sottoclasse nella regione h per l'anno r e la tipologia di scuola t . Sia ancora:

$${}_{tr}f_{b(h)} = \frac{{}_{tr}m_{b(h)}}{{}_{tr}a(h)}$$

la proporzione di alunni nella sottoclasse, con ${}_{tr}f_{b(h)} < 1$.

Sotto l'ipotesi che il coefficiente di variazione

$${}_{tr}c_{(h)}^2 = {}_{tr}\hat{\sigma}_{(h)}^2 / {}_{tr}\hat{y}_{(h)}^2, \quad [7.11]$$

nella sottopopolazione sia approssimativamente uguale a quello della popolazione nel suo complesso, la stima della varianza della stima ${}_{tr}\hat{y}_{b(h)}$ riferita alla sottopopolazione b può essere ottenuta come:

$$\hat{V}({}_{tr}\hat{y}_{b(h)}) \cong \frac{1}{{}_{tr}f_{b(h)} \cdot {}_{tr}a(h)} \frac{({}_{tr}A(h) - {}_{tr}a(h))}{{}_{tr}A(h)} {}_{tr}c_{(h)}^2 {}_{tr}\hat{y}_{b(h)}^2 [1 + {}_{tr}\rho(h) ({}_{tr}f_{b(h)} \cdot {}_{tr}\bar{a}(h) - 1)]. \quad [7.12]$$

La varianza di una stima nazionale (o di aggregazione di regioni) viene quindi ottenuta come somma delle varianze regionali determinate secondo l'espressione [7.11].

Box 7.2 – Formula per le regioni AR (Valle d'Aosta e Bolzano)

La precedente relazione vale per le regioni NAR, nelle rimanenti regioni AR la formula da applicare è la seguente:

$$\hat{V}({}_{tr}\hat{y}_{b(h)}) \cong \frac{1}{{}_{tr}f_{b(h)} \cdot {}_{tr}a(h)} \frac{({}_{tr}A(h) - {}_{tr}a(h))}{{}_{tr}A(h)} {}_{tr}c_{(h)}^2 {}_{tr}\hat{y}_{b(h)}^2 [1 + {}_{tr}\rho(h) ({}_{tr}f_{b(h)} \cdot {}_{tr}\bar{a}(h)_{[ij]} - 1)]$$

7.3.3. Modelli per sottoclassi distribuite in modo diseguale

Per tali sottoclassi si può partire dalla formula (8.3) di Cicchitelli (Cicchitelli, Herzel e Montanari, 1992).

$$\hat{V}(\text{tr} \hat{y}_{b(h)}) \cong \frac{1}{\text{tr} f_b \text{tr} a(h)} \frac{(\text{tr} A(h) - \text{tr} a(h))}{\text{tr} A(h)} \text{tr} c(h)^2 \text{tr} \hat{y}_b^2 \times \\ \times \{[(1 + \text{tr} \rho(h)) (\text{tr} f_{b(h)} \text{tr} \bar{a}(h)) - 1] + [(1 - \text{tr} f_{b(h)}) \text{tr} \hat{y}_{b(h)}^2]\} \quad [7.13]$$

Per le regioni AR si adotta la formula precedente previa sostituzione di $\text{tr} \bar{a}(h)$ con $\text{tr} \bar{a}(h)[ij]$.

7.3.4. Calcolo delle varianze campionarie

Gli utenti e i ricercatori che dispongano dei micro-dati possono calcolare autonomamente le varianze campionarie delle proprie stime, adottando la procedura articolata nei seguenti passi:

- calcolo della proporzione della sottoclasse $\text{tr} f_b$;
- calcolo della stima per la specifica sottoclasse $\text{tr} \hat{y}_{bh}$;
- calcolo della varianza $\text{tr} \hat{\sigma}_h^2$ mediante un software standard statistico e calcolo del coefficiente di variazione mediante semplice rapporto $\text{tr} c(h)^2 = \text{tr} \hat{\sigma}_h^2 / \text{tr} \hat{y}_b^2$;
- calcolo della varianza campionaria mediante la [7.12] per le classi che si distribuiscono in modo approssimativamente uguale o la [7.13] per le altre sottoclassi.

Tab. 35 – Statistiche per il calcolo delle varianze campionarie per stime di punteggi medi riferiti a specifiche sottopopolazioni – Italiano

Ripartizione geografica	Grado 2			Grado 5			Grado 8		
	$v_r A_{(h)}$	$v_r a_{(h)}$	$v_r \rho_{(h)}$	$v_r A_{(h)}$	$v_r a_{(h)}$	$v_r \rho_{(h)}$	$v_r A_{(h)}$	$v_r a_{(h)}$	$v_r \rho_{(h)}$
Valle d'Aosta	1.177	564	0,1	1.172	561	0,1	1.150	599	0,1
Piemonte	36.214	1.292	0,3	37.817	1.347	0,2	38.392	1.583	0,1
Liguria	11.156	943	0,2	12.083	943	0,3	12.205	1.296	0,1
Lombardia	90.745	1.906	0,1	94.952	1.982	0,1	94.762	2.576	0,1
Prov. Aut. Bolzano (l. it.)	–	–	–	–	–	–	1.377	565	0,3
Prov. Aut. Trento	5.027	859	0,1	5.311	951	0,1	5.418	1.144	0,1
Veneto	43.595	1.347	0,2	46.353	1.391	0,1	46.671	1.699	0,0
Friuli-Venezia Giulia	9.424	914	0,2	9.845	926	0,3	10.256	1.117	0,1
Emilia-Romagna	39.188	1.266	0,1	41.026	1.340	0,2	40.344	1.760	0,1
Toscana	31.011	1.302	0,1	32.525	1.357	0,1	32.811	1.598	0,1
Umbria	7.397	839	0,1	7.866	922	0,3	7.868	1.144	0,2
Marche	13.255	979	0,2	13.634	926	0,1	13.731	1.319	0,1
Lazio	51.235	1.470	0,1	52.483	1.599	0,2	52.380	2.020	0,2
Abruzzo	10.871	819	0,2	11.172	933	0,1	11.502	1.188	0,1
Molise	2.198	794	0,2	2.346	832	0,2	2.536	688	0,2
Campania	54.993	1.682	0,4	58.149	1.798	0,3	61.723	2.191	0,2
Puglia	34.919	1.343	0,3	35.599	1.433	0,3	39.138	1.886	0,2
Basilicata	4.243	815	0,2	4.542	904	0,2	4.826	1.070	0,2
Calabria	16.617	1.108	0,4	17.232	1.160	0,2	18.016	1.258	0,3
Sicilia	44.720	1.478	0,4	45.419	1.497	0,4	50.133	2.050	0,4
Sardegna	12.645	600	0,0	12.805	676	0,1	13.683	1.122	0,2

Tab. 36 – Statistiche per il calcolo delle varianze campionarie per stime di punteggi medi riferiti a specifiche sottopopolazioni – Italiano G10

Ripartizione geografica	Licei classici e scientifici			Altri licei			Istituti tecnici			Istituti professionali		
	$trA_{(h)}$	$tr\alpha_{(h)}$	$tr\rho_{(h)}$	$trA_{(h)}$	$tr\alpha_{(h)}$	$tr\rho_{(h)}$	$trA_{(h)}$	$tr\alpha_{(h)}$	$tr\rho_{(h)}$	$trA_{(h)}$	$tr\alpha_{(h)}$	$tr\rho_{(h)}$
Valle d'Aosta	227	116	1,0	308	143	0,3	340	126	0,3	161	123	1,2
Piemonte	9.732	628	0,2	8.180	423	0,2	11.214	525	0,1	4.996	359	0,1
Liguria	3.585	588	0,6	3.123	358	0,1	3.142	443	0,3	1.868	264	0,1
Lombardia	21.404	919	0,3	18.796	858	0,3	27.006	771	0,2	9.988	394	0,1
Prov. Aut. Bolzano (l. it.)	347	220	1,8	387	117	0,3	394	188	0,2	233	58	0,2
Prov. Aut. Trento	1.251	306	0,4	1.205	197	0,2	1.612	230	0,2	–	–	–
Veneto	9.721	739	0,2	8.789	685	0,2	15.228	680	0,1	7.128	381	0,1
Friuli-Venezia Giulia	2.516	446	0,4	1.999	319	0,4	3.527	381	0,2	1.258	222	0,3
Emilia-Romagna	9.181	796	0,2	8.531	535	0,2	13.413	631	0,2	4.955	478	0,2
Toscana	8.742	944	0,5	8.631	616	0,4	9.489	595	0,2	4.630	258	0,2
Umbria	2.605	518	0,0	1.795	265	0,3	2.161	329	0,1	1.012	181	0,1
Marche	3.600	716	0,4	3.480	407	0,5	4.330	464	0,1	1.981	235	0,1
Lazio	19.523	1.139	0,3	12.968	469	0,3	11.579	620	0,1	5.316	315	0,1
Abruzzo	4.110	654	0,2	2.408	357	0,5	3.328	417	0,2	1.397	157	0,1
Molise	820	319	0,8	525	164	0,1	929	194	0,1	393	113	0,1
Campania	20.802	987	0,1	12.751	574	0,3	16.079	639	0,1	10.839	327	0,3
Puglia	11.543	877	0,3	8.089	498	0,1	12.419	624	0,2	7.136	325	0,0
Basilicata	1.800	413	0,5	1.098	270	0,5	1.568	326	0,5	1.208	162	0,5
Calabria	6.496	652	0,2	3.128	367	0,3	5.622	413	0,1	3.209	178	0,3
Sicilia	14.570	944	0,3	10.529	515	0,2	13.047	519	0,3	7.685	249	0,1
Sardegna	4.461	630	0,3	3.314	317	0,1	4.145	371	0,4	2.649	112	-0,1

Tab. 37 – Statistiche per il calcolo delle varianze campionarie per stime di punteggi medi riferiti a specifiche sottopopolazioni – Italiano G13

Ripartizione geografica	Licei classici e scientifici			Altri licei			Istituti tecnici			Istituti professionali		
	$trA_{(h)}$	$tr a_{(h)}$	$tr \rho_{(h)}$	$trA_{(h)}$	$tr a_{(h)}$	$tr \rho_{(h)}$	$trA_{(h)}$	$tr a_{(h)}$	$tr \rho_{(h)}$	$trA_{(h)}$	$tr a_{(h)}$	$tr \rho_{(h)}$
Valle d'Aosta	211	51	0,0	241	145	0,7	300	98	0,9	180	118	0,4
Piemonte	8.424	721	0,3	6.746	480	0,3	9.385	655	0,3	6.041	450	0,1
Liguria	3.151	526	0,4	2.657	277	0,1	2.636	509	0,1	1.878	233	0,3
Lombardia	18.315	881	0,2	16.210	801	0,4	22.455	974	0,4	11.628	561	0,2
Prov. Aut. Bolzano (l. it.)	260	187	0,8	330	111	0,6	307	116	0,2	232	59	0,5
Prov. Aut. Trento	986	263	0,8	1.176	161	0,1	1.505	229	0,0	134	63	0,1
Veneto	8.590	691	0,6	7.537	634	0,3	13.563	767	0,3	7.664	466	0,2
Friuli-Venezia Giulia	2.245	456	0,3	1.762	299	0,5	2.975	366	0,5	1.612	241	0,0
Emilia-Romagna	7.766	748	0,2	7.219	502	0,5	10.628	582	0,3	6.476	507	0,2
Toscana	7.076	852	0,1	7.188	564	0,4	8.039	752	0,2	5.053	325	0,3
Umbria	2.307	473	0,5	1.548	201	0,3	1.786	328	0,1	1.211	171	0,2
Marche	3.317	643	0,4	3.102	353	0,1	4.115	485	0,2	2.596	223	0,1
Lazio	16.280	1.124	0,3	11.208	389	0,3	10.807	658	0,4	6.336	335	0,3
Abruzzo	3.623	659	0,6	2.248	271	0,1	3.131	380	0,3	1.559	234	0,5
Molise	846	343	0,5	516	122	0,4	803	177	0,6	473	139	0,2
Campania	19.109	950	0,3	12.224	499	0,2	17.187	863	0,4	10.761	471	0,3
Puglia	10.952	813	0,2	7.247	431	0,2	11.125	661	0,3	6.514	208	0,4
Basilicata	1.725	430	0,3	1.046	223	0,2	1.510	293	0,7	1.115	189	0,1
Calabria	5.845	596	0,3	2.722	342	0,4	5.804	485	0,1	2.899	224	0,5
Sicilia	12.696	847	0,6	9.540	404	0,3	12.905	494	0,2	7.134	308	1,0
Sardegna	3.628	539	0,3	2.815	318	0,3	3.534	425	0,5	2.045	193	0,2

Tab. 38 – Statistiche per il calcolo delle varianze campionarie per stime di punteggi medi riferiti a specifiche sottopopolazioni – Matematica

Ripartizione geografica	Grado 2			Grado 5			Grado 8		
	$trA_{(h)}$	$tr\alpha_{(h)}$	$tr\rho_{(h)}$	$trA_{(h)}$	$tr\alpha_{(h)}$	$tr\rho_{(h)}$	$trA_{(h)}$	$tr\alpha_{(h)}$	$tr\rho_{(h)}$
Valle d'Aosta	1.177	575	0,4	1.172	567	0,6	1.150	595	0,1
Piemonte	36.214	1.306	0,3	37.817	1.362	0,2	38.392	1.576	0,2
Liguria	11.156	952	0,3	12.083	941	0,2	12.205	1.288	0,2
Lombardia	90.745	1.927	0,1	94.952	2.012	0,1	94.762	2.572	0,1
Prov. Aut. Bolzano (l. it.)	–	–	–	–	–	–	1.377	496	0,2
Prov. Aut. Trento	5.027	876	0,2	5.311	955	0,2	5.418	1.143	0,1
Veneto	43.595	1.377	0,2	46.353	1.389	0,1	46.671	1.697	0,1
Friuli-Venezia Giulia	9.424	916	0,2	9.845	936	0,3	10.256	1.117	0,1
Emilia-Romagna	39.188	1.287	0,2	41.026	1.354	0,3	40.344	1.756	0,2
Toscana	31.011	1.320	0,1	32.525	1.370	0,2	32.811	1.618	0,1
Umbria	7.397	853	0,1	7.866	917	0,3	7.868	1.107	0,2
Marche	13.255	986	0,1	13.634	936	0,1	13.731	1.316	0,1
Lazio	51.235	1.509	0,2	52.483	1.601	0,2	52.380	2.012	0,2
Abruzzo	10.871	853	0,3	11.172	952	0,3	11.502	1.193	0,1
Molise	2.198	794	0,7	2.346	847	0,4	2.536	687	0,1
Campania	54.993	1.730	0,9	58.149	1.818	0,6	61.723	2.185	0,3
Puglia	34.919	1.348	0,5	35.599	1.442	0,7	39.138	1.876	0,2
Basilicata	4.243	825	0,2	4.542	916	0,5	4.826	1.063	0,2
Calabria	16.617	1.138	0,3	17.232	1.178	0,4	18.016	1.255	0,3
Sicilia	44.720	1.516	0,6	45.419	1.495	0,7	50.133	2.020	0,4
Sardegna	12.645	608	0,1	12.805	637	0,1	13.683	1.103	0,2

Tab. 39 – Statistiche per il calcolo delle varianze campionarie per stime di punteggi medi riferiti a specifiche sottopopolazioni – Matematica G10

Ripartizione geografica	Licei scientifici			Altri licei			Istituti tecnici			Istituti professionali		
	$trA_{(h)}$	$tr a_{(h)}$	$tr \rho_{(h)}$	$trA_{(h)}$	$tr a_{(h)}$	$tr \rho_{(h)}$	$trA_{(h)}$	$tr a_{(h)}$	$tr \rho_{(h)}$	$trA_{(h)}$	$tr a_{(h)}$	$tr \rho_{(h)}$
Valle d'Aosta	173	83	0,4	362	175	0,9	340	124	1,1	161	118	2,2
Piemonte	8.111	521	0,2	9.801	531	0,5	11.214	506	0,3	4.996	350	0,2
Liguria	2.797	436	1,0	3.911	507	0,3	3.142	441	0,5	1.868	260	0,2
Lombardia	17.899	747	0,2	22.301	969	0,5	27.006	764	0,5	9.988	377	0,2
Prov. Aut. Bolzano (l. it.)	299	121	0,6	435	171	1,1	394	165	0,5	233	44	0,1
Prov. Aut. Trento	1.097	226	0,4	1.359	269	0,2	1.612	229	0,3	–	–	–
Veneto	7.976	542	0,2	10.534	899	0,2	15.228	690	0,2	7.128	378	0,2
Friuli-Venezia Giulia	2.091	327	0,4	2.424	434	0,4	3.527	379	0,5	1.258	220	0,4
Emilia-Romagna	7.778	568	0,5	9.934	745	0,1	13.413	620	0,3	4.955	468	0,2
Toscana	7.315	516	0,2	10.058	1.070	0,4	9.489	608	0,2	4.630	260	0,1
Umbria	2.054	358	0,0	2.346	414	0,3	2.161	330	0,2	1.012	180	0,1
Marche	2.800	414	0,3	4.280	701	0,6	4.330	467	0,4	1.981	219	0,0
Lazio	14.547	592	0,3	17.944	963	0,6	11.579	598	0,2	5.316	308	0,1
Abruzzo	3.283	401	0,2	3.235	595	0,4	3.328	405	0,3	1.397	151	0,1
Molise	640	209	0,5	705	275	0,6	929	189	0,3	393	107	0,1
Campania	16.214	650	0,1	17.339	876	0,5	16.079	636	0,2	10.839	307	0,2
Puglia	9.094	569	0,4	10.538	790	0,2	12.419	595	0,4	7.136	302	0,2
Basilicata	1.346	249	0,2	1.552	422	0,7	1.568	327	0,6	1.208	173	0,7
Calabria	4.813	478	0,5	4.811	540	0,4	5.622	401	0,2	3.209	171	0,1
Sicilia	10.456	595	0,3	14.643	840	0,3	13.047	493	0,3	7.685	272	0,2
Sardegna	3.395	369	0,2	4.380	547	0,3	4.145	353	0,3	2.649	113	0,2

Tab. 40 – Statistiche per il calcolo delle varianze campionarie per stime di punteggi medi riferiti a specifiche sottopopolazioni – Matematica G13

Ripartizione geografica	Licei scientifici			Altri licei			Istituti tecnici			Istituti professionali		
	$trA_{(h)}$	$tr a_{(h)}$	$tr \rho_{(h)}$	$trA_{(h)}$	$tr a_{(h)}$	$tr \rho_{(h)}$	$trA_{(h)}$	$tr a_{(h)}$	$tr \rho_{(h)}$	$trA_{(h)}$	$tr a_{(h)}$	$tr \rho_{(h)}$
Valle d'Aosta	160	55	0,1	292	170	1,3	300	119	1,2	180	117	1,2
Piemonte	6.956	530	0,3	8.214	665	0,4	9.385	643	0,9	6.041	450	0,3
Liguria	2.442	354	0,7	3.366	443	0,4	2.636	489	0,4	1.878	249	0,3
Lombardia	15.379	701	0,4	19.146	983	0,4	22.455	982	0,9	11.628	561	0,3
Prov. Aut. Bolzano (l. it.)	210	149	0,9	380	116	0,3	307	100	0,2	232	27	0,5
Prov. Aut. Trento	852	215	0,5	1.310	213	0,3	1.505	232	0,7	134	64	0,4
Veneto	7.114	482	0,5	9.013	852	0,3	13.563	760	0,6	7.664	450	0,2
Friuli-Venezia Giulia	1.862	350	0,3	2.145	405	0,6	2.975	365	1,1	1.612	232	0,4
Emilia-Romagna	6.558	500	0,6	8.427	732	0,7	10.628	582	0,8	6.476	515	0,2
Toscana	5.895	454	0,1	8.369	984	0,5	8.039	771	0,8	5.053	326	0,4
Umbria	1.824	324	0,6	2.031	357	0,7	1.786	363	0,3	1.211	172	0,1
Marche	2.611	387	0,7	3.808	579	0,3	4.115	503	0,7	2.596	216	0,2
Lazio	12.129	640	0,5	15.359	822	0,6	10.807	637	0,5	6.336	344	0,4
Abruzzo	2.933	378	0,5	2.938	549	0,3	3.131	385	0,7	1.559	253	0,3
Molise	655	245	0,6	707	247	0,6	803	181	0,4	473	153	0,3
Campania	15.024	609	0,3	16.309	879	0,4	17.187	853	0,5	10.761	445	0,2
Puglia	8.536	540	0,4	9.663	691	0,3	11.125	702	0,4	6.514	262	0,2
Basilicata	1.246	290	0,5	1.525	410	0,6	1.510	319	1,3	1.115	245	0,3
Calabria	4.367	440	0,3	4.200	513	0,4	5.804	473	0,5	2.899	223	0,2
Sicilia	9.108	425	1,0	13.128	831	0,5	12.905	506	0,4	7.134	341	0,5
Sardegna	2.742	319	0,3	3.701	553	0,4	3.534	414	0,5	2.045	184	0,2

Tab. 41 – Statistiche per il calcolo delle varianze campionarie per stime di punteggi medi riferiti a specifiche sottopopolazioni – Inglese L

Ripartizione geografica	Grado 5			Grado 8		
	$v_r A_{(h)}$	$v_r a_{(h)}$	$v_r \rho_{(h)}$	$v_r A_{(h)}$	$v_r a_{(h)}$	$v_r \rho_{(h)}$
Valle d'Aosta	1.172	561	0,7	1.150	581	0,4
Piemonte	37.817	1.342	0,3	38.392	1.528	0,3
Liguria	12.083	907	0,4	12.205	1.286	0,1
Lombardia	94.952	1.960	0,3	94.762	2.536	0,2
Prov. Aut. Bolzano (l. it.)	–	–	–	1.377	583	0,3
Prov. Aut. Trento	5.311	956	0,3	5.418	1.142	0,2
Veneto	46.353	1.393	0,2	46.671	1.713	0,1
Friuli-Venezia Giulia	9.845	930	0,3	10.256	1.083	0,1
Emilia-Romagna	41.026	1.328	0,4	40.344	1.708	0,2
Toscana	32.525	1.372	0,4	32.811	1.565	0,2
Umbria	7.866	896	0,4	7.868	1.112	0,2
Marche	13.634	933	0,3	13.731	1.311	0,1
Lazio	52.483	1.587	0,3	52.380	2.026	0,2
Abruzzo	11.172	946	0,3	11.502	1.181	0,1
Molise	2.346	843	0,3	2.536	664	0,4
Campania	58.149	1.799	0,8	61.723	2.148	0,3
Puglia	35.599	1.449	0,6	39.138	1.872	0,2
Basilicata	4.542	913	0,8	4.826	1.061	0,2
Calabria	17.232	1.154	0,5	18.016	1.194	0,4
Sicilia	45.419	1.481	0,9	50.133	2.009	0,4
Sardegna	12.805	696	0,4	13.683	1.111	0,2

Tab. 42 – Statistiche per il calcolo delle varianze campionarie per stime di punteggi medi riferiti a specifiche sottopopolazioni – Inglese L G13

Ripartizione geografica	Licei classici e scientifici			Altri licei			Istituti tecnici			Istituti professionali		
	$trA_{(h)}$	$tr\alpha_{(h)}$	$tr\rho_{(h)}$	$trA_{(h)}$	$tr\alpha_{(h)}$	$tr\rho_{(h)}$	$trA_{(h)}$	$tr\alpha_{(h)}$	$tr\rho_{(h)}$	$trA_{(h)}$	$tr\alpha_{(h)}$	$tr\rho_{(h)}$
Valle d'Aosta	211	69	0,0	241	104	1,5	300	100	0,6	180	116	1,1
Piemonte	8.424	648	0,4	6.746	472	0,8	9.385	639	0,3	6.041	425	0,2
Liguria	3.151	502	0,5	2.657	269	0,6	2.636	478	0,2	1.878	246	0,2
Lombardia	18.315	865	0,4	16.210	781	0,7	22.455	964	0,4	11.628	548	0,2
Prov. Aut. Bolzano (l. it.)	260	164	0,5	330	74	1,4	307	103	0,3	232	38	0,7
Prov. Aut. Trento	986	268	0,7	1.176	162	0,9	1.505	213	0,2	134	61	0,1
Veneto	8.590	641	0,5	7.537	563	0,5	13.563	752	0,5	7.664	445	0,3
Friuli-Venezia Giulia	2.245	457	0,6	1.762	308	0,4	2.975	354	0,4	1.612	238	0,1
Emilia-Romagna	7.766	736	0,6	7.219	493	0,6	10.628	540	0,4	6.476	496	0,3
Toscana	7.076	799	0,4	7.188	556	0,4	8.039	712	0,4	5.053	303	0,3
Umbria	2.307	481	0,6	1.548	192	0,2	1.786	318	0,2	1.211	156	0,0
Marche	3.317	609	0,3	3.102	313	0,4	4.115	479	0,2	2.596	206	0,3
Lazio	16.280	1.030	0,5	11.208	389	0,7	10.807	556	0,5	6.336	330	0,3
Abruzzo	3.623	642	0,7	2.248	259	0,4	3.131	372	0,3	1.559	252	0,3
Molise	846	337	0,8	516	142	0,8	803	181	0,6	473	133	0,2
Campania	19.109	812	0,2	12.224	534	0,7	17.187	815	0,5	10.761	434	0,1
Puglia	10.952	805	0,1	7.247	426	0,5	11.125	689	0,4	6.514	257	0,4
Basilicata	1.725	417	0,2	1.046	256	0,5	1.510	310	0,5	1.115	238	0,4
Calabria	5.845	575	0,5	2.722	355	0,6	5.804	466	0,4	2.899	213	0,3
Sicilia	12.696	874	0,7	9.540	380	0,7	12.905	481	0,3	7.134	325	0,2
Sardegna	3.628	494	0,5	2.815	252	0,9	3.534	409	0,5	2.045	169	0,3

Tab. 43 – Statistiche per il calcolo delle varianze campionarie per stime di punteggi medi riferiti a specifiche sottopopolazioni – Inglese R

Ripartizione geografica	Grado 5			Grado 8		
	$v \cdot A_{(h)}$	$r \cdot a_{(h)}$	$u \cdot \rho_{(h)}$	$v \cdot A_{(h)}$	$r \cdot a_{(h)}$	$u \cdot \rho_{(h)}$
Valle d'Aosta	1.172	562	0,2	1.150	581	0,3
Piemonte	37.817	1.344	0,2	38.392	1.582	0,2
Liguria	12.083	907	0,3	12.205	1.292	0,1
Lombardia	94.952	1.961	0,2	94.762	2.570	0,1
Prov. Aut. Bolzano (l. it.)	–	–	–	1.377	589	0,4
Prov. Aut. Trento	5.311	956	0,2	5.418	1.145	0,2
Veneto	46.353	1.394	0,2	46.671	1.718	0,1
Friuli-Venezia Giulia	9.845	932	0,2	10.256	1.109	0,2
Emilia-Romagna	41.026	1.328	0,2	40.344	1.755	0,1
Toscana	32.525	1.371	0,1	32.811	1.607	0,2
Umbria	7.866	897	0,2	7.868	1.139	0,2
Marche	13.634	928	0,1	13.731	1.296	0,2
Lazio	52.483	1.587	0,1	52.380	2.038	0,2
Abruzzo	11.172	945	0,2	11.502	1.191	0,1
Molise	2.346	842	0,2	2.536	688	0,3
Campania	58.149	1.799	0,4	61.723	2.161	0,3
Puglia	35.599	1.449	0,4	39.138	1.875	0,3
Basilicata	4.542	913	0,4	4.826	1.060	0,2
Calabria	17.232	1.154	0,2	18.016	1.221	0,4
Sicilia	45.419	1.481	0,4	50.133	2.013	0,5
Sardegna	12.805	710	0,3	13.683	1.110	0,2

Tab. 44 – Statistiche per il calcolo delle varianze campionarie per stime di punteggi medi riferiti a specifiche sottopopolazioni – Inglese R G13

Ripartizione geografica	Licei classici e scientifici			Altri licei			Istituti tecnici			Istituti professionali		
	$trA_{(h)}$	$tr\alpha_{(h)}$	$tr\rho_{(h)}$	$trA_{(h)}$	$tr\alpha_{(h)}$	$tr\rho_{(h)}$	$trA_{(h)}$	$tr\alpha_{(h)}$	$tr\rho_{(h)}$	$trA_{(h)}$	$tr\alpha_{(h)}$	$tr\rho_{(h)}$
Valle d'Aosta	211	70	0,4	241	108	1,0	300	99	0,9	180	116	0,6
Piemonte	8.424	675	0,3	6.746	470	0,7	9.385	643	0,3	6.041	445	0,2
Liguria	3.151	497	0,5	2.657	270	0,3	2.636	481	0,2	1.878	244	0,3
Lombardia	18.315	869	0,5	16.210	776	0,6	22.455	968	0,3	11.628	548	0,2
Prov. Aut. Bolzano (l. it.)	260	162	0,9	330	77	1,2	307	101	0,3	232	38	0,5
Prov. Aut. Trento	986	267	0,4	1.176	162	0,6	1.505	215	0,0	134	61	0,2
Veneto	8.590	642	0,5	7.537	561	0,4	13.563	769	0,4	7.664	444	0,3
Friuli-Venezia Giulia	2.245	454	0,4	1.762	308	0,3	2.975	355	0,5	1.612	238	0,3
Emilia-Romagna	7.766	735	0,4	7.219	494	0,5	10.628	538	0,4	6.476	499	0,2
Toscana	7.076	797	0,1	7.188	558	0,5	8.039	726	0,3	5.053	310	0,4
Umbria	2.307	480	0,5	1.548	193	0,1	1.786	356	0,2	1.211	158	0,0
Marche	3.317	610	0,3	3.102	313	0,4	4.115	479	0,4	2.596	207	0,4
Lazio	16.280	1.056	0,3	11.208	387	0,5	10.807	595	0,4	6.336	338	0,4
Abruzzo	3.623	647	0,5	2.248	264	0,3	3.131	382	0,3	1.559	253	0,7
Molise	846	334	0,6	516	140	0,6	803	180	0,6	473	134	0,2
Campania	19.109	883	0,4	12.224	538	0,5	17.187	817	0,4	10.761	457	0,4
Puglia	10.952	807	0,2	7.247	428	0,4	11.125	702	0,4	6.514	264	0,3
Basilicata	1.725	431	0,1	1.046	256	0,3	1.510	310	0,8	1.115	240	0,4
Calabria	5.845	575	0,5	2.722	359	0,6	5.804	471	0,2	2.899	211	0,3
Sicilia	12.696	870	0,5	9.540	406	0,8	12.905	484	0,4	7.134	327	0,6
Sardegna	3.628	508	0,4	2.815	277	0,7	3.534	407	0,6	2.045	172	0,6

8. *Il cheating e l'affidabilità dei dati*

L'importanza crescente che a livello nazionale e internazionale rivestono le valutazioni standardizzate delle competenze degli studenti si accompagna a una crescente esigenza di disporre di dati qualitativamente affidabili e attendibili. In particolar modo, in Italia è molto forte l'attenzione dell'opinione pubblica verso le rilevazioni effettuate dall'INVALSI nell'ambito del Servizio nazionale di valutazione (SNV). Di conseguenza la qualità dei dati raccolti è diventata un requisito indispensabile per renderli pienamente fruibili e "apprezzabili" da tutti gli *stakeholders* del sistema educativo: *policy makers*, insegnanti, ricercatori, famiglie e studenti. In questa prospettiva risulta necessario disporre di strumenti metodologici tecnicamente validi e condivisi in grado di massimizzare la qualità dei dati e di limitare l'effetto di componenti "distorsive" come quelle legate al fenomeno del *cheating*.

Il *cheating* in ambito educativo indica le pratiche messe in atto dagli studenti (*student cheating*) o degli insegnanti (*teacher cheating*) per alterare, sovrastimandoli, i risultati delle prove standardizzate di valutazione. La presenza del *cheating* tende a invalidare i risultati delle rilevazioni compromettendone fortemente l'accuratezza e di conseguenza l'attendibilità.

Il *cheating* è un argomento molto complesso e, dal punto di vista metodologico, difficile da identificare e correggere in quanto può essere attuato da molteplici attori, in diversi step del processo di rilevazione dei dati e con differenti modalità.

8.1. L'individuazione e correzione del *cheating* mediante l'approccio *fuzzy clustering*

L'INVALSI, sin dalle prime rilevazioni condotte nell'a.s. 2004/05 nell'ambito del SNV, ha messo in luce la presenza del fenomeno e ha cercato di correggerne e limitarne l'impatto adottando una procedura statistica di correzione costruita *ad hoc* per i dati del SNV (Quintano, Castellano e Longobardi, 2009), procedura resa più robusta con un'evoluzione della stessa a partire dall'anno 2014 (Longobardi, Falzetti e Pagliuca, 2018). Dall'anno scolastico 2017/18, con l'introduzione delle prove al

computer per la scuola secondaria di primo e secondo grado, è stato verificato che il fenomeno del *cheating* è circoscritto solo alle prove svolte in formato cartaceo ovvero solo per i gradi interessati della scuola primaria.

I dati rilevati con l'*Indagine Campionaria di Controllo* (ICC) risultano statisticamente accurati in virtù del ricorso a osservatori esterni in grado di garantire un corretto svolgimento dell'intero processo di somministrazione. La presenza degli osservatori si traduce nel produrre una base dati di elevata affidabilità caratterizzata dalla pressoché assenza di anomalie in termini di *cheating*.

La metodologia adottata quindi si applica principalmente sui dati provenienti dall'*Indagine Standard*, ricorrendo a tecniche statistiche di riduzione dimensionale e di raggruppamento sfocato (fattorizzazione e *fuzzy clustering*), permette di stimare la probabilità di *cheating* di ogni classe di studenti mediante l'analisi di un insieme di indicatori: punteggio medio, variabilità interna, livello di eterogeneità nelle risposte e tasso di presenza di dati mancanti.

Sulla base di questa probabilità si costruisce un fattore di correzione che “penalizza” i punteggi medi di classe al crescere del grado di sospetto di *cheating*.

Si parte dal presupposto che se in una classe vengono messi in atto dei comportamenti opportunistici, determinando *patterns* anomali di risposta, da parte degli insegnanti oppure degli studenti, i risultati tenderanno a perdere la variabilità che naturalmente dovrebbe contraddistinguerli e, di conseguenza, si osserverà un elevato punteggio medio della classe (che tenderà verso il valore massimo), una variabilità interna minima o nulla e, inoltre, valori minimi dei tassi di mancata risposta, in quanto tutti gli studenti forniranno risposte al test che tenderanno a essere uguali per tutta la classe (eterogeneità minima o assente nelle risposte).

8.2. Le prove CBT e l'assenza di *cheating*

Fino al 2017, l'INVALSI doveva distribuire due milioni e mezzo di prove sotto forma di fascicoli cartacei in oltre 46.000 sedi scolastiche, sparse su tutto il territorio nazionale. Dal 2018, invece, le prove hanno cominciato a trasferirsi dalla carta al computer. Questo cambiamento ha permesso di dispensare gli insegnanti delle scuole secondarie di primo e secondo grado dall'inserimento manuale delle risposte in apposite maschere elettroniche. Ancora, nelle prove su carta, a tutti gli studenti erano presentate le stesse domande, in quelle al computer invece, gli studenti svolgono prove differenti, equivalenti per abilità misurata e livello di difficoltà, tratti da un'apposita *banca delle domande*. In questo modo i risultati sono più affidabili e attendibili (cfr. capitolo 1). Inoltre, in una prova al computer i risultati sono basati su un maggior numero di dati, grazie alle prove e ai pretest effettuati negli ultimi anni precedenti su milioni di studenti. In questo modo, l'INVALSI dispone di una grande base di dati sulla capacità degli studenti di rispondere ai quesiti contenuti nella *banca delle domande*: analizzandoli statisticamente, è possibile calcolare la probabilità con la quale, data una risposta corretta a una delle domande della prova, uno studente

avrebbe potuto rispondere correttamente ad altri quesiti. In altre parole, è come se lo studente avesse risposto a molte più domande di quelle presenti nella sua prova. Il risultato è una valutazione ancora più accurata rispetto al passato, e che lo diventerà sempre di più in futuro quando l'INVALSI disporrà di ancora più dati.

La sostanziale coincidenza dei dati delle classi campione e di quelle non campione conferma l'efficacia della modalità CBT nel ridurre fortemente i fenomeni di *cheating*. Questo è molto importante non solo dal punto di vista statistico, ma anche rispetto al più generale piano educativo e formativo. Regolarità e correttezza nell'esecuzione delle prove è in primo luogo un esempio importante di rispetto delle regole e quindi della legalità, che va ben oltre l'importanza dell'attendibilità delle misure statistiche.

A titolo esemplificativo si riportano i grafici che rappresentano la differenza tra i risultati degli studenti campione e popolazione nel grado 8, in Appendice i grafici per il grado 13.

Fig. 11 – Andamento campione vs popolazione – G8 Italiano

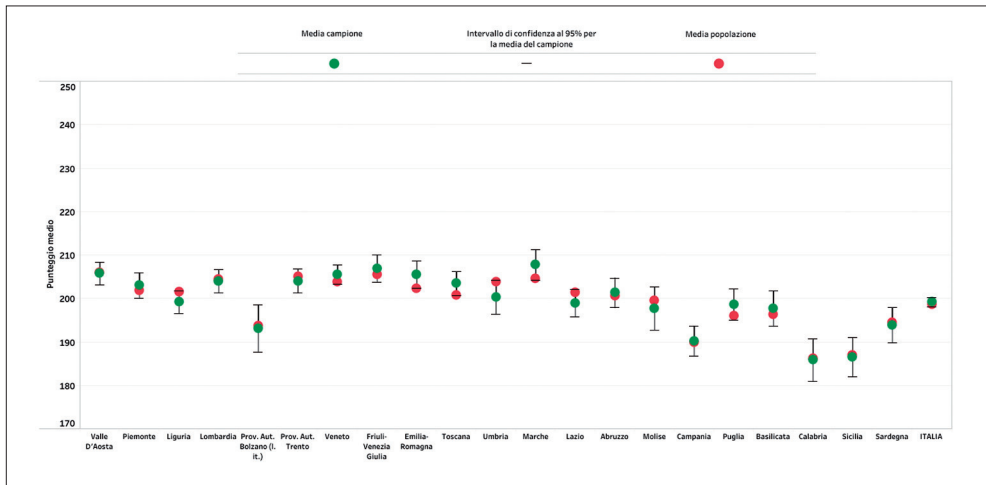


Fig. 12 – Andamento campione vs popolazione – G8 Matematica

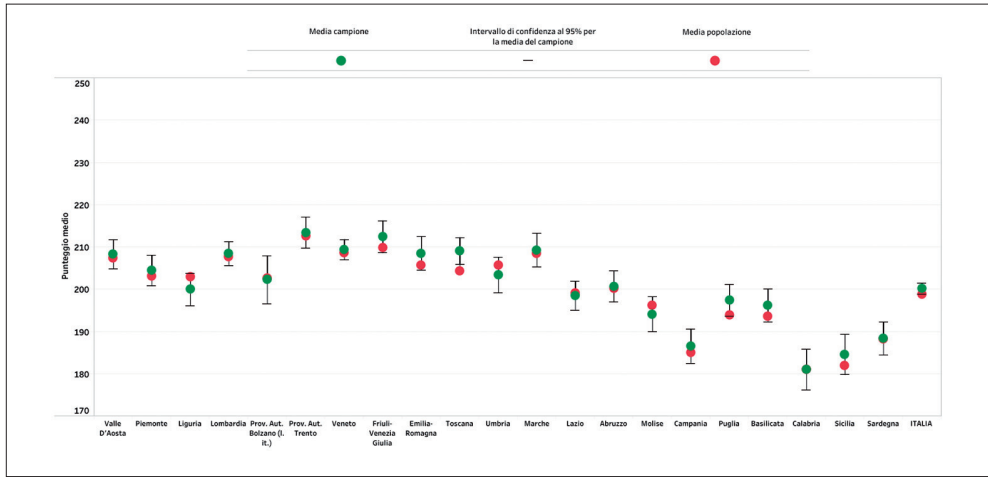


Fig. 13 – Andamento campione vs popolazione – G8 Inglese Listening

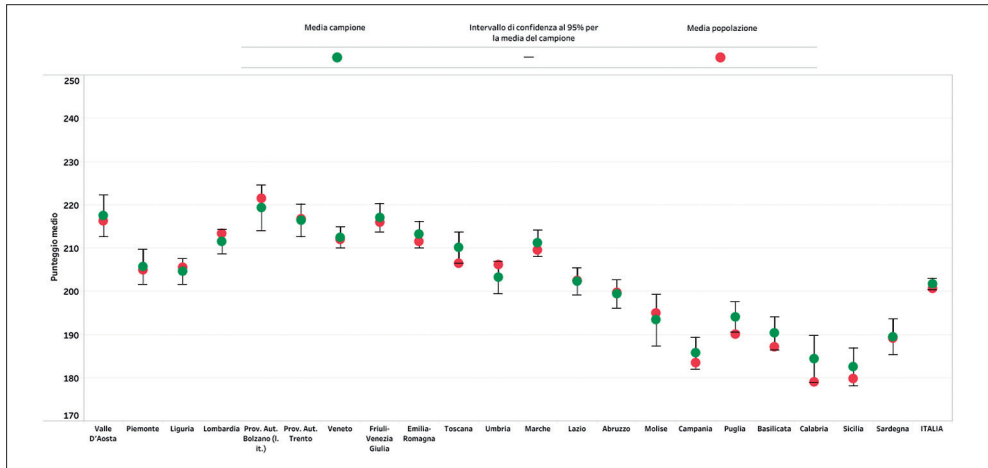
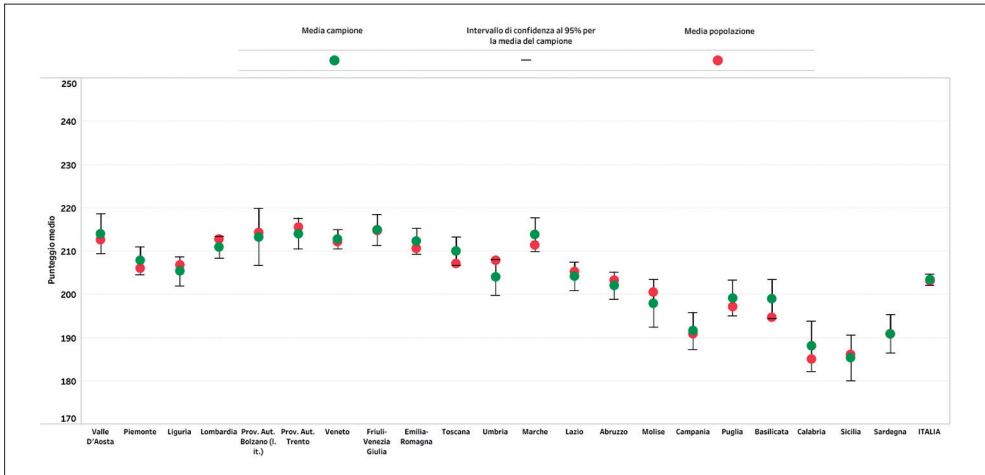


Fig. 14 – Andamento campione vs popolazione – G8 Inglese Reading



Come si può osservare dai grafici, i risultati di popolazione cadono all'interno degli intervalli di confidenza dei risultati degli studenti campione: per questo motivo le differenze sono imputabili al caso e non risultano statisticamente significative.

Bibliografia

- Cicchitelli G., Herzal A., Montanari G.E. (1992), *Il campionamento statistico*, il Mulino, Bologna.
- Cochran W.G. (1977), *Sampling Techniques*, Wiley, New York.
- Deville J.-C., Tillé Y. (2005), “Variance approximation under balanced sampling”, *Journal of Statistical Planning and Inference*, 128, pp. 569-591.
- Kish L. (1965), *Survey Sampling*, Wiley, New York.
- Longobardi S., Falzetti P., Pagliuca M.M. (2018), “Quis custodiet ipsos custodes? How to detect and correct teacher cheating in Italian student data”, *Statistical Methods and Applications*, doi: 10.1007/s10260-018-0426-2.
- Quintano C., Castellano R., Longobardi S. (2009), “A fuzzy clustering approach to improve the accuracy of Italian student data. An experimental procedure to correct the impact of outliers on assessment test scores”, *Statistica&Applicazioni*, VII (2).
- Tandalla L. (2012), *Scoring Short Answers Essays. Automated Student Assessment Prize (ASAP)*, <https://storage.googleapis.com/kaggle-competitions/kaggle/2959/media/TechnicalMethodsPaper.pdf>, p. 13.
- Woodruff R.S. (1971), “A Simple Method for Approximating the Variance of a Complicated Estimate”, *Journal of the American Statistical Association*, 66 (334), pp. 411-414.

A.1. Riformulazione dei punteggi medi

I valori medi regionali possono essere riformulati nel modo seguente

$$\begin{aligned} tr \bar{Y}(h) &= \frac{1}{tr A(h)} \sum_{i=1}^{tr S(h)} \sum_{j=1}^{tr S(h)} tr Y(h)_{ij} \\ &= \frac{1}{tr A(h)} \sum_{i=1}^{tr S(h)} tr Y(h)_i \\ &= \frac{tr Y(h)}{tr A(h)}. \end{aligned}$$

Una formulazione alternativa è data da

$$\begin{aligned} tr \bar{Y}(h) &= \frac{1}{tr A(h)} \sum_{i=1}^{tr S(h)} \sum_{j=1}^{tr S(h)} tr A(h)_{ij} tr \bar{Y}(h)_{ij} \\ &= \frac{1}{tr A(h)} \sum_{i=1}^{tr S(h)} tr A(h)_i tr \bar{Y}(h)_i. \end{aligned}$$

Analogamente, il valore medio nazionale è dato da

$$\begin{aligned} tr \bar{Y} &= \frac{1}{tr A} \sum_{h=1}^{21} tr Y(h) \\ &= \frac{tr Y}{tr A}, \end{aligned}$$

$$= \frac{1}{tr A} \sum_{h=1}^{21} tr A^{(h)} tr \bar{Y}^{(h)}$$

essendo $tr Y = \sum_{h=1}^{21} tr Y^{(h)}$.

A.2. Differenti componenti della devianza

Essendo i doppi prodotti nulli:

$$tr D = \sum_{h=1}^{21} \sum_{i=1}^{tr S^{(h)}} \sum_{j=1}^{tr S^{(h)i}} \sum_{k=1}^{tr A^{(h)ij}} (tr Y^{(h)ijk} - tr \bar{Y} + tr \bar{Y}^{(h)i} - tr \bar{Y}^{(h)i} + tr \bar{Y}^{(h)ij} - tr \bar{Y}^{(h)ij})^2 =$$

$$tr D = \sum_{h=1}^{21} \sum_{i=1}^{tr S^{(h)}} \sum_{j=1}^{tr S^{(h)i}} \sum_{k=1}^{tr A^{(h)ij}} (tr \bar{Y}^{(h)i} - tr \bar{Y})^2 +$$

$$+ \sum_{h=1}^{21} \sum_{i=1}^{tr S^{(h)}} \sum_{j=1}^{tr S^{(h)i}} \sum_{k=1}^{tr A^{(h)ij}} (tr \bar{Y}^{(h)ij} - tr \bar{Y}^{(h)i})^2 +$$

$$+ \sum_{h=1}^{21} \sum_{i=1}^{tr S^{(h)}} \sum_{j=1}^{tr S^{(h)i}} \sum_{k=1}^{tr A^{(h)ij}} (tr Y^{(h)ijk} - tr \bar{Y}^{(h)ij})^2.$$

Si ottiene il risultato, riportato nel paragrafo 4.3, considerando che

- $\sum_{h=1}^{21} \sum_{i=1}^{tr S^{(h)}} \sum_{j=1}^{tr S^{(h)i}} \sum_{k=1}^{tr A^{(h)ij}} (tr \bar{Y}^{(h)i} - tr \bar{Y})^2 = \sum_{h=1}^{21} \sum_{i=1}^{tr S^{(h)}} tr A^{(h)i} (tr \bar{Y}^{(h)i} - tr \bar{Y})^2,$
- $\sum_{h=1}^{21} \sum_{i=1}^{tr S^{(h)}} \sum_{j=1}^{tr S^{(h)i}} \sum_{k=1}^{tr A^{(h)ij}} (tr \bar{Y}^{(h)ij} - tr \bar{Y}^{(h)i})^2$
- $= \sum_{h=1}^{21} \sum_{i=1}^{tr S^{(h)}} \sum_{j=1}^{tr S^{(h)i}} tr A^{(h)ij} (tr \bar{Y}^{(h)ij} - tr \bar{Y}^{(h)i})^2.$

A.3. Condizione di autoponderazione

Nell'ipotesi che la dimensione delle classi in termini di alunni sia pressappoco costante nello strato e che ogni classe sia approssimativamente costituita da $t \bar{A}^{(h)}$ alunni, valgono le seguenti condizioni:

a) $tr A^{(h)ij} \cong t \bar{A}^{(h)}$

[a]

b) $tr A^{(h)i} \cong t S^{(h)i} t \bar{A}^{(h)}$

Valendo la [a], si ha

- ${}_t \ddot{A}(h)_i \cong {}_t \bar{A}(h) {}_t S(h)_i \frac{2}{{}_t S(h)_i}$ [b]
- $\sum_{i=1} {}_t S(h) {}_t \ddot{A}(h)_i \cong {}_t \bar{A}(h) {}_t \bar{S}(h)$.

Sostituendo le precedenti espressioni [a] e [b] nella (5.4) si ha:

$${}_{tr} \pi(h)_{ijk} = {}_t S(h) \frac{{}_t \ddot{A}(h)_i}{{}_t S(h)_i} \frac{{}_t S(h)_i}{{}_t S(h)_i} \cong {}_t S(h) \frac{{}_t \bar{A}(h) {}_t S(h)_i}{{}_t \bar{A}(h) {}_t \bar{S}(h)} \frac{2}{{}_t S(h)_i} \frac{{}_t S(h)_i}{{}_t S(h)_i} = \frac{2 {}_t S(h)}{{}_t \bar{S}(h)} \quad [c]$$

A.4. Algoritmo per il peso calibrato

1. Prima iterazione

$${}_{tr}^1 W_{(h)i} = \frac{1}{{}_t \pi_{(h)i}} \frac{{}_t S(h)}{\sum_{i=1} {}_t S(h) \frac{1}{{}_t \pi_{(h)i}}}$$

2. Iterazione sul numero di alunni

$${}_{tr}^2 W_{(h)i} = {}_{tr}^1 W_{(h)i} \frac{{}_{tr} A(h)}{\sum_{i=1} {}_t S(h) {}_{tr}^1 W_{(h)i} {}_{tr} A(h)_i}$$

3. Iterazione sul numero di scuole

$${}_{tr}^3 W_{(h)i} = {}_{tr}^2 W_{(h)i} \frac{{}_t S(h)}{\sum_{i=1} {}_t S(h) {}_{tr}^2 W_{(h)i}}$$

4. Generica iterazione (v) sul numero di alunni

$${}_{tr}^v W_{(h)i} = {}_{tr}^{v-1} W_{(h)i} \frac{{}_{tr} A(h)}{\sum_{i=1} {}_t S(h) {}_{tr}^{v-1} W_{(h)i} {}_{tr} A(h)_i}$$

5. Generica iterazione sul numero di scuole

$${}_{tr}^v W_{(h)i} = {}_{tr}^{v-1} W_{(h)i} \frac{{}_t S(h)}{\sum_{i=1} {}_t S(h) {}_{tr}^{v-1} W_{(h)i}}$$

6. Interruzione della procedura

La procedura si interrompe dopo 4/5 iterazioni allo stabilizzarsi, definendo il peso di primo stadio $tr w_{(h)i}$.

A.5. Stima delle componenti della devianza

Stima della devianza totale D

Per la dimostrazione della [6.5], si considerino i seguenti passaggi in cui, per non appesantire eccessivamente la dimostrazione si sono omessi gli indici t , r e (h) :

$$\begin{aligned} \hat{D} &= \sum_{i=1}^s w_i \sum_{j=1}^{s_i} \sum_{k=1}^{A_{ij}} w_{j|i} (y_{ijk} - \hat{y})^2 = \\ &= \sum_{i=1}^s w_i \sum_{j=1}^{s_i} \sum_{k=1}^{A_{ij}} w_{j|i} y_{ijk}^2 + \sum_{i=1}^s w_i \sum_{j=1}^{s_i} \sum_{k=1}^{A_{ij}} w_{j|i} \hat{y}^2 + \\ &- 2 \hat{y} \sum_{i=1}^s w_i \sum_{j=1}^{s_i} \sum_{k=1}^{A_{ij}} w_{j|i} y_{ijk} \\ &= \sum_{i=1}^s w_i \sum_{j=1}^{s_i} \sum_{k=1}^{A_{ij}} w_{j|i} y_{ijk}^2 + \hat{A} \hat{y}^2 - 2 \hat{y} \hat{A} \hat{y} \\ &= \sum_{i=1}^s w_i \sum_{j=1}^{s_i} \sum_{k=1}^{A_{ij}} w_{j|i} y_{ijk}^2 - \hat{A} \hat{y}^2. \end{aligned}$$

Si considerino i seguenti risultati:

$$\begin{aligned} E_p \left(\sum_{i=1}^s w_i \sum_{j=1}^{s_i} \sum_{k=1}^{A_{ij}} w_{j|i} y_{ijk}^2 \right) &= \sum_{i=1}^s \sum_{j=1}^{s_i} \sum_{k=1}^{A_{ij}} w_{ijk}^2 \\ E_p(\hat{A} \hat{y}^2) &\cong E_p(\hat{A}) E_p(\hat{y}^2) = A(V(\hat{y}) + \bar{y}^2). \end{aligned}$$

Di conseguenza, una stima approssimativamente corretta di D è data da:

$$\hat{D}_{cor} = \sum_{i=1}^s w_i \sum_{j=1}^{s_i} \sum_{k=1}^{A_{ij}} w_{j|i} y_{ijk}^2 - \hat{A} (\hat{y}^2 - \hat{V}(\hat{y})).$$

Stima della componente D_I

Per la dimostrazione della [6.6], si considerino i seguenti passaggi in cui, per non appesantire eccessivamente la dimostrazione si sono omessi gli indici t , r e (h) :

$$\begin{aligned}
\hat{D}1 &= \sum_{i=1}^s w_i \hat{A}_i (\hat{y}_i - \hat{y})^2 = \\
\hat{D}1 &= \sum_{i=1}^s w_i \hat{A}_i \hat{y}_i^2 + \hat{y}^2 \sum_{i=1}^s w_i \hat{A}_i - 2\hat{y} \sum_{i=1}^s w_i \hat{A}_i \hat{y}_i \\
&= \sum_{i=1}^s w_i \hat{A}_i \hat{y}_i^2 + \hat{A} \hat{y}^2 - 2\hat{y} \hat{A} \hat{y} \\
&= \sum_{i=1}^s w_i \hat{A}_i \hat{y}_i^2 - \hat{A} \hat{y}^2.
\end{aligned}$$

In base ai risultati precedenti, una stima approssimativamente corretta di $D1$ è data da:

$$\begin{aligned}
\hat{D}1_{cor} &= [\sum_{i=1}^s w_i \hat{A}_i (\hat{y}_i^2 - \check{V}(\hat{y}_i))] - [A(\hat{y}^2 - \hat{V}(\hat{y}))] \\
\hat{D}1_{cor} &= \hat{D}1 - \sum_{i=1}^s \check{V}(\hat{y}_i) + A\hat{V}(\hat{y})
\end{aligned}$$

essendo

$$\check{V}(\hat{y}_i) = \begin{cases} \hat{V}(\hat{y}_i) & \text{per } i \in (>1) \bar{S}(h) \\ \tilde{V}(\hat{y}_i) & \text{per } i \in (=1) \bar{S}(h) \end{cases}.$$

Negli istituti con almeno 2 sezioni campione tale variabilità viene stimata come:

$$\hat{V}(tr \hat{y}_i) = \sum_{j=1}^{t^S(h)i} \frac{t^S(h)i}{t^S(h)i - 1} \left(tr \hat{R}_{ij} - tr \hat{R}(h)_i \right)^2,$$

essendo

$$\begin{aligned}
tr \hat{R}(h)_{ij} &= \frac{1}{t \pi(h)_{j|i}} \frac{1}{tr \hat{A}(h)_i} (tr Y_{(h)ij} - \frac{tr \hat{Y}(h)_i}{tr \hat{A}(h)_i} tr A^*(h)_{ij}), & tr \hat{R}(h)_i &= \sum_{j=1}^{t^S(h)i} \frac{1}{t^S(h)i} tr \hat{R}(h)_{ij}, \\
tr \hat{Y}(h)_i &= \sum_{j=1}^{t^S(h)i} \frac{1}{t \pi(h)_{j|i}} tr A^*(h)_{ij}, & tr \hat{A}(h)_i &= \sum_{j=1}^{t^S(h)i} \frac{1}{t \pi(h)_{j|i}} tr A^*(h)_{ij}.
\end{aligned}$$

Negli istituti con un'unica sezione campione tale variabilità viene stimata come:

$$\hat{V}(tr \hat{y}_i) = \sum_{k=1}^{tr A_{hij}} \frac{1}{tr A_{hij} - 1} \left(tr z_{(h)ijk} - tr \bar{z}(h)_{ij} \right)^2.$$

A.6. Espressione alternativa della stima della varianza regionale

Riformulando, la [7.1] al fine di mettere in evidenza il termine ${}^iS(h)$ si ottiene:

$$V(\text{tr} \hat{y}(h)) = \frac{1}{\text{tr} A^2(h)} \left[\frac{\text{tr} K(h)}{\text{tr} S(h)} - \text{tr} G(h) \right], \quad [6.3]$$

in cui

$$\text{tr} K(h) = \sum_{i=1} {}^iS(h) \frac{1}{\text{tr} \alpha(h)_i} \left[\text{tr} Z^2(h)_i + \frac{\text{tr} S(h)_i (\text{tr} S(h)_i - \text{tr} S(h)_i)}{\text{tr} S(h)_i} \text{tr} \sigma^2_z(h)_i \right], \quad [6.4]$$

$$\text{tr} G(h) = \sum_{i=1} {}^iS(h) \text{tr} Z^2(h)_i. \quad [6.5]$$

A.7. Tavole di allocazione del campione per i gradi 2, 5 e 10

Tab. 45 – Allocazione del campione nella scuola primaria classi II e V: anno 2018/19

Regione e/o Provincia autonoma	Totale scuole	Totale alumni	Campione						
			Allocazione proporzionale tra le regioni			Allocazione uguale tra le regioni			Allocazione di compromesso $\beta=0,3$
			Scuole	CV%	Scuole	CV%	Scuole	CV%	
Valle d'Aosta	21	1.189	21	–	21	–	21	–	–
Piemonte	434	38.248	42	1,5	35	1,6	37	1,6	1,6
Liguria	167	12.248	13	2,6	35	1,5	28	1,7	1,7
Lombardia	1.010	94.872	96	0,9	35	1,5	53	1,2	1,2
Prov. Aut. di Bolzano (l. it.)	17	1.369	17	–	17	–	17	–	–
Prov. Aut. di Trento	58	5.353	5	3,4	35	1	26	1,3	1,3
Veneto	495	46.228	43	1,2	35	1,4	37	1,3	1,3
Friuli-Venezia Giulia	132	10.431	10	2,6	35	1,3	27	1,5	1,5
Emilia-Romagna	414	40.686	38	1,3	35	1,4	36	1,4	1,4
Toscana	388	33.086	37	1,5	35	1,6	35	1,6	1,6
Umbria	100	7.987	8	3	35	1,3	27	1,5	1,5
Marche	165	13.594	13	2,3	35	1,4	28	1,5	1,5
Lazio	681	54.538	63	1,3	35	1,7	43	1,5	1,5
Abruzzo	149	11.451	13	2,8	35	1,6	28	1,8	1,8
Molise	38	2.428	4	6,5	35	1,4	26	1,9	1,9
Campania	928	61.496	103	1,6	35	2,8	55	2,2	2,2
Puglia	437	38.328	51	1,6	35	2	40	1,9	1,9
Basilicata	80	4.885	7	4,6	35	1,7	26	2,1	2,1
Calabria	265	17.795	28	2,8	35	2,5	33	2,6	2,6
Sicilia	647	48.563	76	1,7	35	2,6	47	2,2	2,2
Sardegna	194	13.253	12	2,5	35	1,4	28	1,6	1,6
Italia	6.820	558.028	700	0,4	700	0,5	700	0,5	0,5

Tab. 46 – Allocations del campione nella scuola secondaria di secondo grado classi II – Licei scientifici: anno 2018/19

Regione e/o Provincia autonoma e tipologia di scuola	Totale scuole		Totale alunni		Campione					
					Allocazione proporzionale tra le regioni		Allocazione uguale tra le regioni		Allocazione di compromesso $\beta=0,3$	
	Scuole	CV%	Scuole	CV%	Scuole	CV%	Scuole	CV%	Scuole	CV%
Valle d'Aosta	3	167	3	—	3	—	3	—	3	—
Piemonte	96	7.864	13	1,6	11	1,7	11	1,7	11	1,8
Liguria	32	2.840	5	2,7	11	1,6	10	1,6	10	1,8
Lombardia	254	17.364	32	1,1	11	1,9	18	1,9	18	1,5
Prov. Aut. di Bolzano (l. it.)	7	238	7	—	7	—	7	—	7	—
Prov. Aut. di Trento	14	1.040	2	4,4	11	1,2	7	1,2	7	1,9
Veneto	94	7.769	11	1,5	11	1,5	13	1,5	13	1,3
Friuli-Venezia Giulia	29	2.152	4	3,3	11	1,7	8	1,7	8	2,2
Emilia-Romagna	87	7.667	12	1,5	11	1,6	13	1,6	13	1,5
Toscana	85	6.805	13	1,9	11	2	12	2	12	1,9
Umbria	19	2.085	3	2,9	11	1,2	8	1,2	8	1,5
Marche	28	2.780	5	2,6	11	1,3	10	1,3	10	1,5
Lazio	173	14.440	25	1,6	11	2,7	15	2,7	15	2,3
Abruzzo	31	3.092	5	2,5	11	1,5	10	1,5	10	1,6
Molise	12	728	2	2,1	11	0,5	7	0,5	7	0,9
Campania	172	16.031	25	1,3	11	1,9	16	1,9	16	1,6
Puglia	98	8.965	15	1,5	11	1,8	13	1,8	13	1,7
Basilicata	20	1.311	3	4,4	10	0,7	7	0,7	7	2,0
Catabria	57	4.606	8	2,3	11	1,8	11	1,8	11	1,9
Sicilia	120	10.203	19	1,5	11	2	14	2	14	1,9
Sardegna	46	3.148	5	2,8	11	1,7	10	1,7	10	2,0
Italia	1.477	121.295	216	0,4	226	0,6	223	0,6	223	0,5

Tab. 47 – Allocations del campione nella scuola secondaria di secondo grado classi II – Licei classici: anno 2018/19

Regione e/o Provincia autonoma e tipologia di scuola	Totale scuole		Totale alunni		Campione				
	Scuole	CV%	Scuole	CV%	Allocazione proporzionale tra le regioni		Allocazione uguale tra le regioni		Allocazione di compromesso $\beta=0,3$
					Scuole	CV%	Scuole	CV%	
Valle d'Aosta	1	43	1	—	1	—	1	—	—
Piemonte	36	1.510	3	4,6	5	3,5	4	3,8	3,8
Liguria	15	746	2	4,8	5	2,7	4	3,1	3,1
Lombardia	67	3.106	6	2,9	5	3,2	5	3,1	3,1
Prov. Aut. di Bolzano (l. it.)	3	64	3	—	3	—	3	—	—
Prov. Aut. di Trento	4	185	2	5,1	4	3	3	3,4	3,4
Veneto	42	1.685	5	5	5	5	5	5,0	5,0
Friuli-Venezia Giulia	11	419	2	4,4	5	1	4	1,9	1,9
Emilia-Romagna	23	1.368	5	5,7	5	5,3	5	5,4	5,4
Toscana	38	1.429	10	6,3	13	5,4	12	5,7	5,7
Umbria	11	463	2	7,1	5	4,4	3	5,7	5,7
Marche	18	763	6	6	7	5,4	8	4,9	4,9
Lazio	88	4.588	17	3,6	9	5,3	12	4,5	4,5
Abruzzo	13	791	4	7,3	7	5,4	6	5,9	5,9
Molise	5	203	3	8,9	5	6,5	3	8,9	8,9
Campania	82	4.297	10	3,5	5	5,1	9	3,7	3,7
Puglia	50	2.415	8	4,9	7	5,4	8	5,0	5,0
Basilicata	11	455	2	5,5	5	3	5	3,1	3,1
Catabria	34	1.637	4	5,3	5	4,4	5	4,4	4,4
Sicilia	78	4.100	14	3,9	10	4,8	10	4,8	4,8
Sardegna	20	1.014	7	5,5	8	5	8	5,2	5,2
Italia	651	3.1281	117	1,2	125	1,4	123	1,3	1,3

Tab. 48 – Allocations del campione nella scuola secondaria di secondo grado classi II – Altri licei: anno 2018/19

Regione e/o Provincia autonoma e tipologia di scuola	Totale scuole		Totale alunni		Campione					
					Allocazione proporzionale tra le regioni		Allocazione uguale tra le regioni		Allocazione di compromesso $\beta=0,3$	
	Scuole	CV%	Scuole	CV%	Scuole	CV%	Scuole	CV%	Scuole	CV%
Valle d'Aosta	5	289	5	—	5	—	5	—	5	—
Piemonte	85	7.614	13	1,1	11	1,2	12	1,2	12	1,2
Liguria	27	3.280	5	1,5	11	0,6	9	0,8	9	0,8
Lombardia	245	18.529	41	0,8	12	1,6	21	1,2	21	1,2
Prov. Aut. di Bolzano (l. it.)	6	361	6	—	6	—	6	—	6	—
Prov. Aut. di Trento	15	1.197	2	2,6	7	0,5	5	0,9	5	0,9
Veneto	107	8.526	23	1,2	12	1,8	16	1,5	16	1,5
Friuli-Venezia Giulia	24	1.992	3	2,1	11	0,8	9	1,0	9	1,0
Emilia-Romagna	70	8.362	15	1	11	1,2	12	1,2	12	1,2
Toscana	91	8.253	20	1,3	12	1,8	15	1,5	15	1,5
Umbria	24	1.716	2	2,2	11	0,6	7	1,0	7	1,0
Marche	41	3.360	5	1,5	11	0,7	10	0,8	10	0,8
Lazio	158	12.214	15	0,8	11	1	12	0,9	12	0,9
Abruzzo	25	2.414	3	2	11	0,9	9	1,0	9	1,0
Molise	8	530	2	2,3	6	0,5	5	0,9	5	0,9
Campania	175	1.2370	22	1	11	1,5	14	1,3	14	1,3
Puglia	89	8.117	10	1	11	0,9	11	1,0	11	1,0
Basilicata	17	1.156	3	3,5	10	0,6	8	1,2	8	1,2
Calabria	50	3.041	5	2,1	11	1,3	9	1,5	9	1,5
Sicilia	145	10.665	20	1,1	11	1,5	14	1,4	14	1,4
Sardegna	45	3.256	4	1,8	11	1	9	1,2	9	1,2
Italia	1.452	117.242	223	0,3	216	0,4	218	0,4	218	0,4

Tab. 49 – Allocations del campione nella scuola secondaria di secondo grado classi II – Istituti tecnici: anno 2018/19

Regione e/o Provincia autonoma e tipologia di scuola	Totale scuole		Totale alunni		Campione					
					Allocazione proporzionale tra le regioni		Allocazione uguale tra le regioni		Allocazione di compromesso $\beta=0,3$	
	Scuole	CV%	Scuole	CV%	Scuole	CV%	Scuole	CV%	Scuole	CV%
Valle d'Aosta	4	317	4	—	4	—	4	—	4	—
Piemonte	113	10.901	13	0,6	14	0,6	14	0,6	14	0,6
Liguria	31	3.131	4	1,4	14	0,5	11	0,6	11	0,6
Lombardia	272	27.050	34	0,4	14	0,7	20	0,6	20	0,6
Prov. Aut. di Bolzano (l. it.)	9	355	9	—	9	—	9	—	9	—
Prov. Aut. di Trento	19	1.675	4	1,8	8	0,4	7	0,8	7	0,8
Veneto	126	14.684	20	0,6	14	0,8	16	0,7	16	0,7
Friuli-Venezia Giulia	37	3.469	5	1,1	12	0,2	10	0,5	10	0,5
Emilia-Romagna	108	13.056	17	0,6	14	0,7	15	0,6	15	0,6
Toscana	92	9.512	21	0,8	14	1,1	16	1,0	16	1,0
Umbria	23	2.000	3	1,9	14	0,6	10	0,9	10	0,9
Marche	51	4.232	7	1,2	14	0,7	13	0,8	13	0,8
Lazio	155	11.849	22	0,8	14	1,1	17	1,0	17	1,0
Abruzzo	45	3.401	5	1,4	14	0,7	11	0,8	11	0,8
Molise	11	906	2	1,9	7	0,2	6	0,5	6	0,5
Campania	249	16.398	39	0,8	15	1,5	21	1,2	21	1,2
Puglia	128	12.620	27	0,8	14	1,2	18	1,0	18	1,0
Basilicata	29	1.649	2	0,8	13	0	10	0,2	10	0,2
Catabria	76	5.904	9	1,2	14	0,8	13	0,9	13	0,9
Sicilia	201	13.236	18	0,7	14	0,8	15	0,8	15	0,8
Sardegna	53	4.194	8	1,5	14	1	12	1,2	12	1,2
Italia	1.833	160.539	275	0,2	267	0,3	268	0,2	268	0,2

Tab. 50 – Allocations del campione nella scuola secondaria di secondo grado classi II – Istituti professionali: anno 2018/19

Regione e/o Provincia autonoma e tipologia di scuola	Totale scuole	Totale alunni	Campione				Allocazione di compromesso $\beta=0,3$	
			Allocazione proporzionale tra le regioni		Allocazione uguale tra le regioni			
			Scuole	CV%	Scuole	CV%		
Valle d'Aosta	3	213	3	–	3	–	3	–
Piemonte	68	6.053	14	2,6	9	3,4	11	3,0
Liguria	22	2.010	4	4,2	8	2,6	8	2,7
Lombardia	131	1.0514	22	1,9	9	3,1	13	2,5
Prov. Aut. di Bolzano (l. it.)	3	253	3	–	3	–	3	–
Prov. Aut. di Trento	2	136	2	5,3	2	5,3	2	5,3
Veneto	81	7.353	15	2,2	9	2,9	11	2,6
Friuli-Venezia Giulia	19	1.481	2	4,3	8	1,9	7	2,1
Emilia-Romagna	70	7.321	16	2,4	12	2,8	13	2,7
Toscana	62	5.511	10	2,5	8	2,8	9	2,7
Umbria	22	1.327	2	4,7	8	1,9	6	2,4
Marche	29	2.585	5	3,8	8	2,8	7	3,1
Lazio	77	6.518	12	2,4	8	2,8	9	2,7
Abruzzo	21	1.529	2	4,5	8	2	7	2,3
Molise	9	318	3	8	8	4,1	4	6,8
Campania	151	12.159	16	1,7	8	2,4	12	2,0
Puglia	80	7.920	11	1,9	8	2,2	10	2,0
Basilicata	22	1.267	2	5	8	1,9	6	2,4
Catabria	53	3.249	4	3,2	8	2,2	7	2,5
Sicilia	115	8.867	13	1,9	8	2,4	10	2,2
Sardegna	28	2.842	4	3,6	8	2,5	7	2,7
Italia	1.070	89.426	167	0,6	165	0,8	165	0,7

A.8. Confronto campione vs popolazione grado 13

Fig. 15 – Andamento campione vs popolazione – G13 Italiano

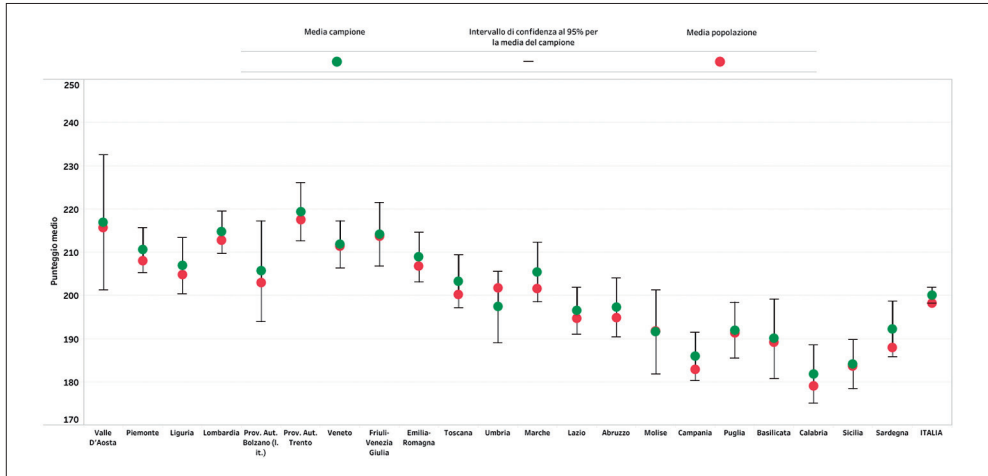


Fig. 16 – Andamento campione vs popolazione – G13 Matematica

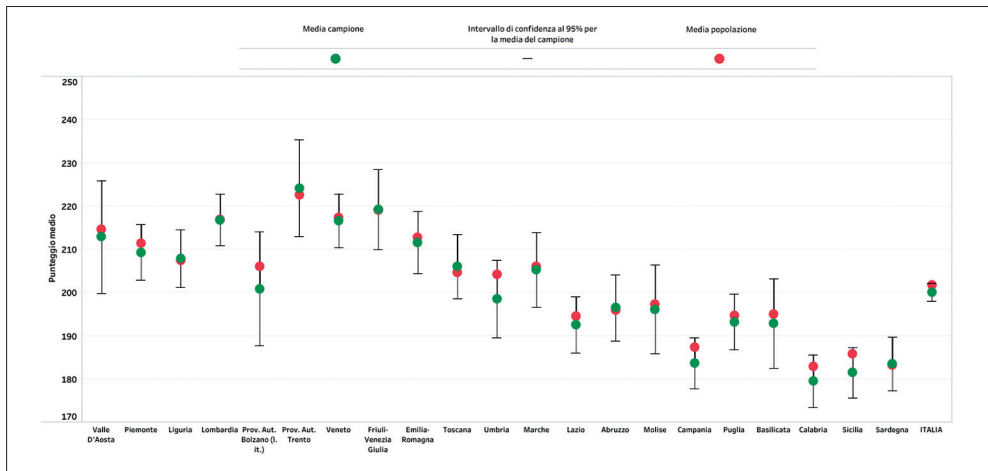


Fig. 17 – Andamento campione vs popolazione – G13 Inglese Listening

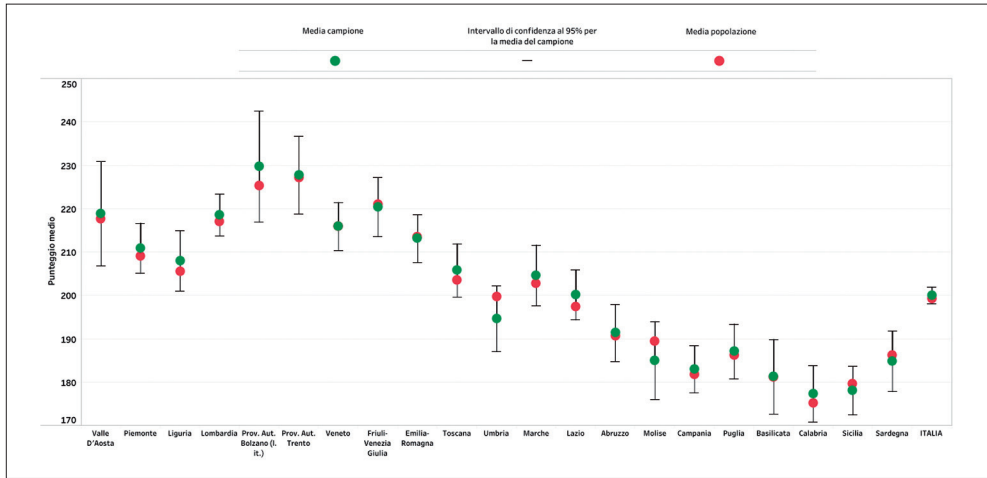
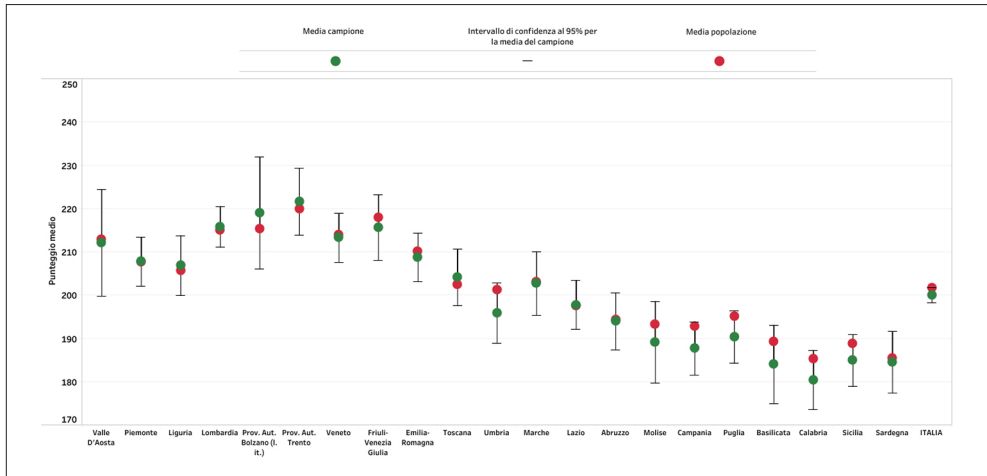


Fig. 18 – Andamento campione vs popolazione – G13 Inglese Reading



A.9. Risultati per regione

Fig. 19 – Risultati per regione Italiano G2

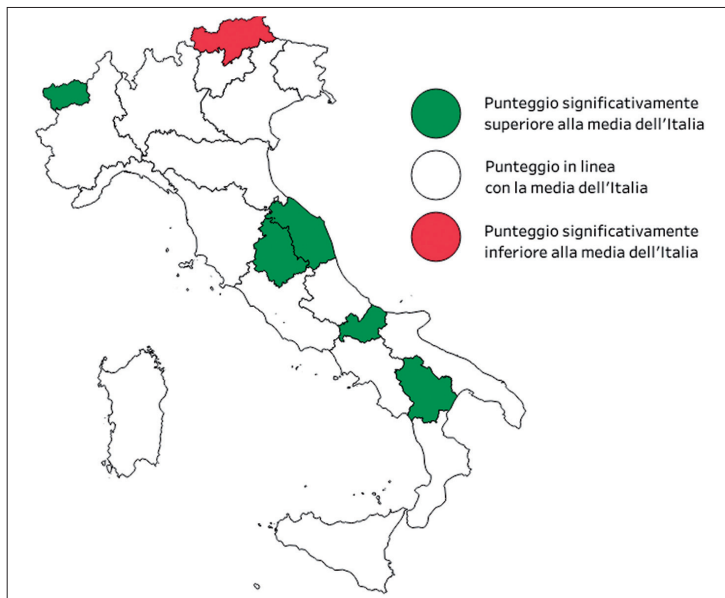


Fig. 20 – Risultati per regione Matematica G2

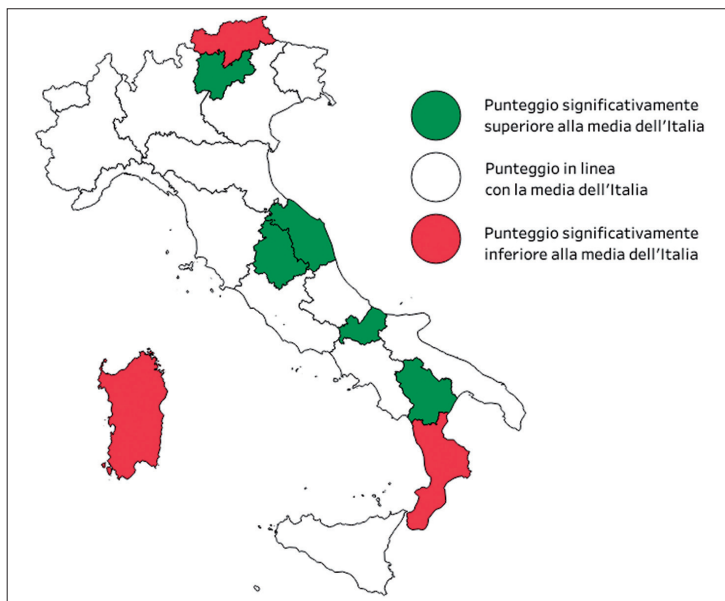


Fig. 21 – Risultati per regione Italiano G5

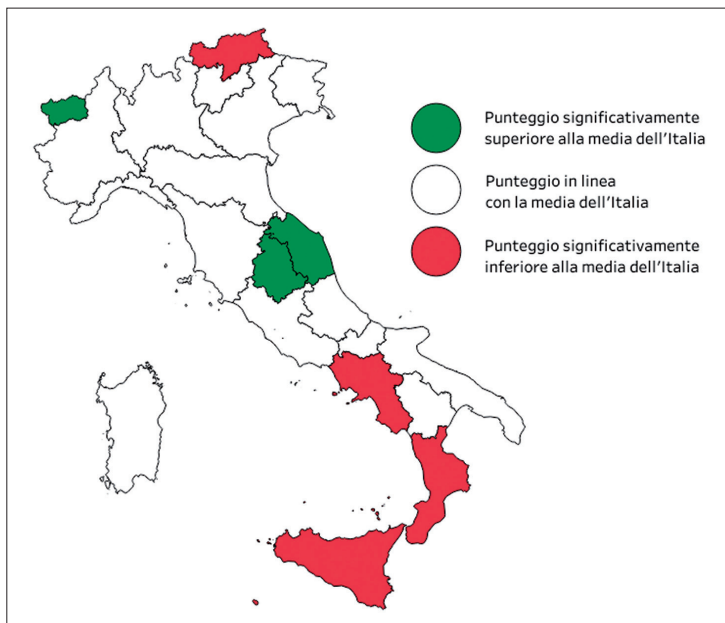


Fig. 22 – Risultati per regione Matematica G5

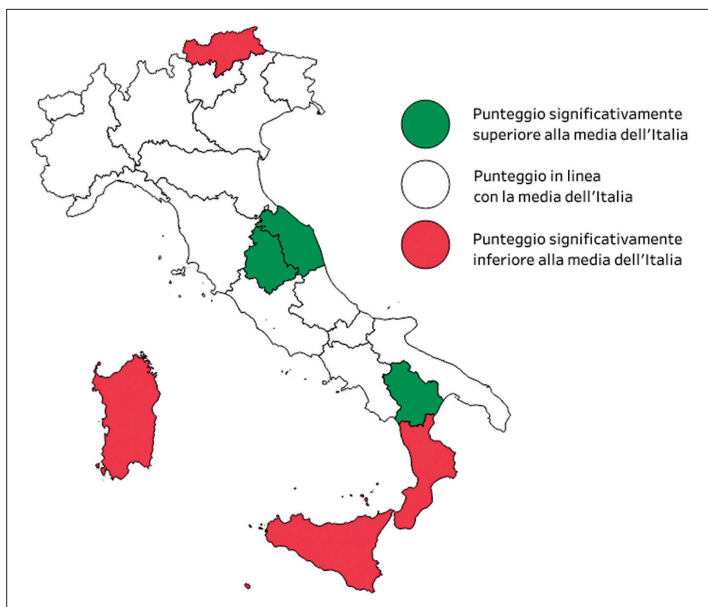


Fig. 23 – Risultati per regione Inglese Listening G5

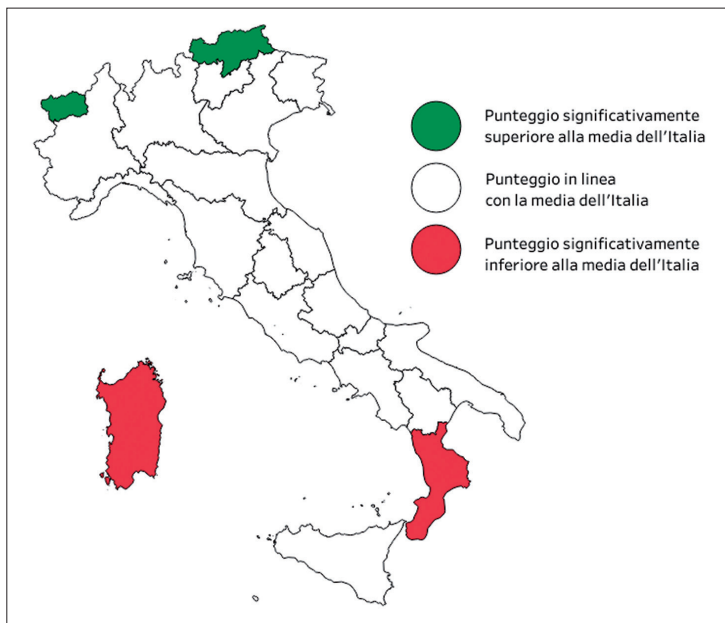


Fig. 24 – Risultati per regione Inglese Reading G5

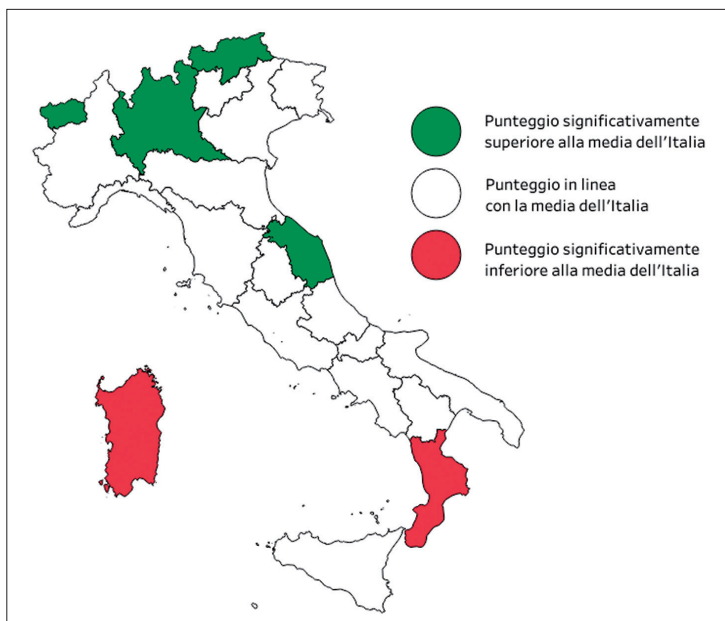


Fig. 25 – Risultati per regione Italiano G10

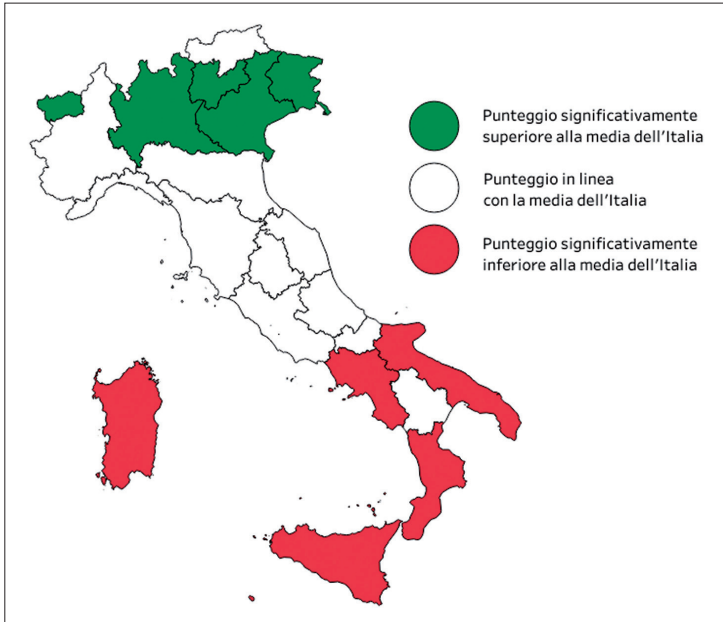


Fig. 26 – Risultati per regione Matematica G10

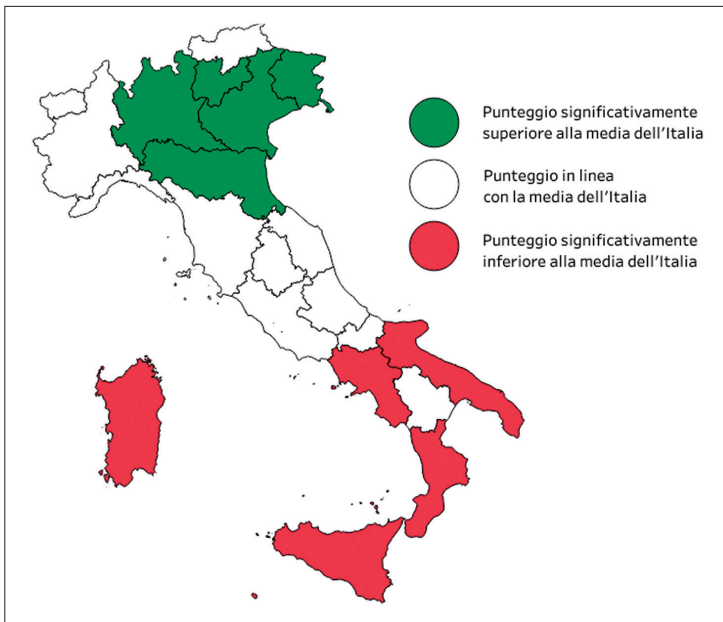


Fig. 27 – Risultati per regione Italiano G13

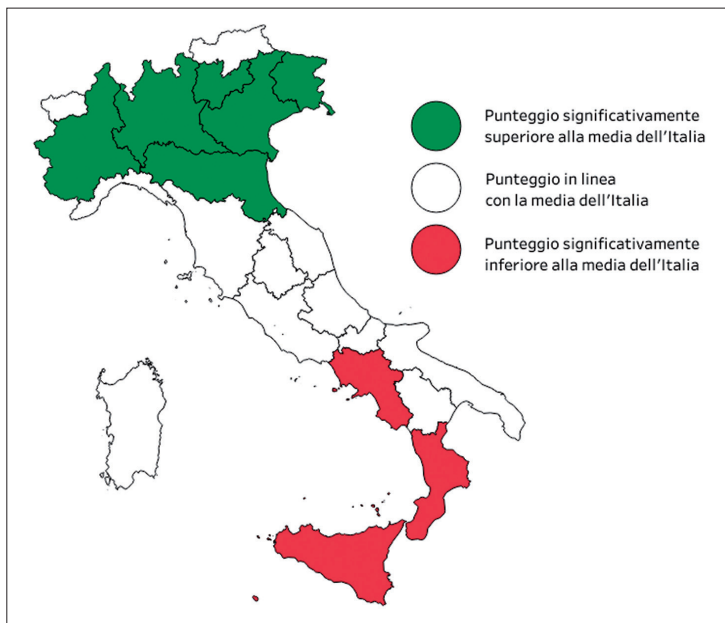


Fig. 28 – Risultati per regione Matematica G13

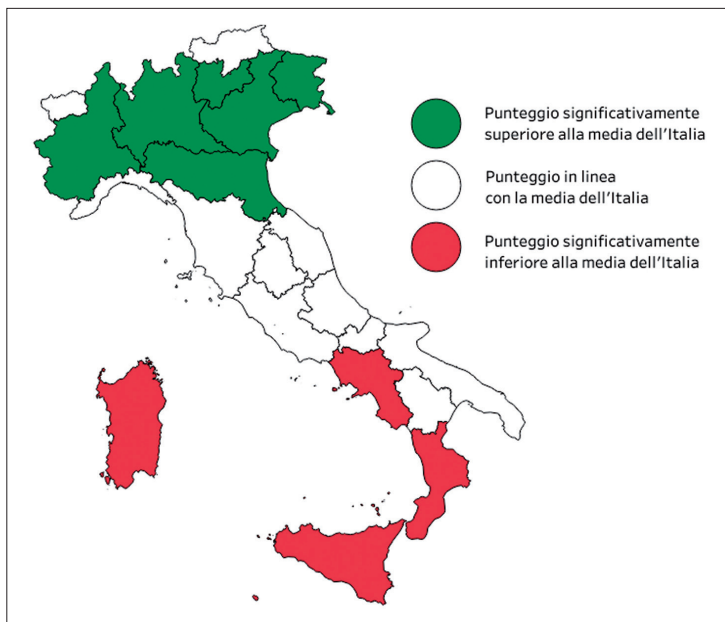


Fig. 29 – Risultati per regione Inglese Listening G13

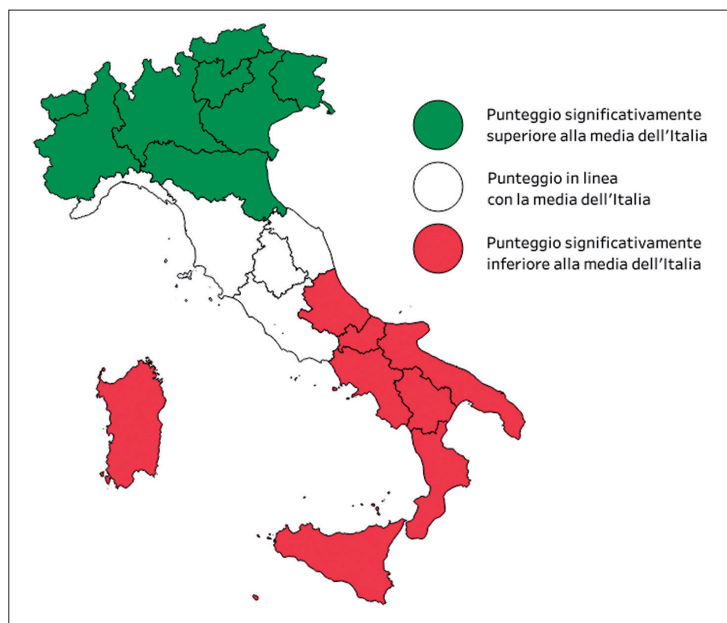
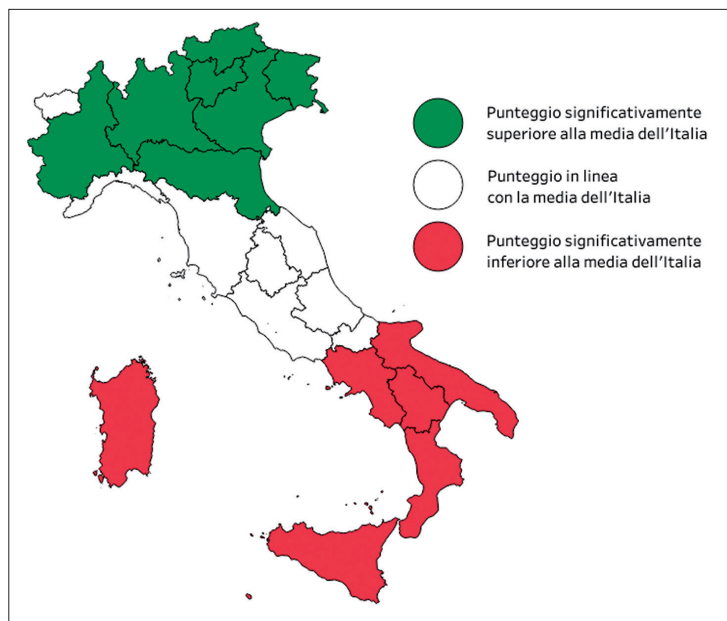


Fig. 30 – Risultati per regione Inglese Reading G13



Indice delle tabelle

Tab. 1 – Totale delle classi e degli studenti per grado scolastico coinvolti nella Rilevazione per l'anno scolastico 2018/2019	pag. 9
Tab. 2 – Tempi di somministrazione	» 11
Tab. 3 – Sottopopolazioni individuate dal grado scolastico e dalla tipologia di scuola	» 14
Tab. 4 – Allocazione del campione nella scuola secondaria di primo grado classi III: anno 2018-19	» 23
Tab. 5 – Allocazione del campione nella scuola secondaria di secondo grado classi V – Licei scientifici: anno 2018/19	» 25
Tab. 6 – Allocazione del campione nella scuola secondaria di secondo grado classi V – Licei classici: anno 2018/19	» 26
Tab. 7 – Allocazione del campione nella scuola secondaria di secondo grado classi V – Altri licei: anno 2018/19	» 27
Tab. 8 – Allocazione del campione nella scuola secondaria di secondo grado classi V – Istituti tecnici: anno 2018/19	» 28
Tab. 9 – Allocazione del campione nella scuola secondaria di secondo grado classi V – Istituti professionali: anno 2018/19	» 29
Tab. 10 – Tavola differenza studenti pesati e popolazione – Italiano grado 8	» 32
Tab. 11 – Punteggi medi per Regione, per grado scolastico del I ciclo e materia indagata	» 36
Tab. 12 – Punteggi medi per Regione, per grado scolastico del II ciclo e materia indagata	» 39
Tab. 13 – Punteggi medi per regione, per grado 10 per tipologia di istituto – Matematica	» 40
Tab. 14 – Punteggi medi per regione, per grado 10 per tipologia di istituto – Italiano	» 41

Tab. 15 – Punteggi medi per regione, per grado 13 per tipologia di istituto – Matematica	pag. 42
Tab. 16 – Punteggi medi per regione, per grado 13 per tipologia di istituto – Italiano	» 43
Tab. 17 – Punteggi medi per regione, per grado 13 per tipologia di istituto – Inglese <i>Reading</i>	» 44
Tab. 18 – Punteggi medi per regione, per grado 13 per tipologia di istituto – Inglese <i>Listening</i>	» 45
Tab. 19 – Componenti della devianza dei punteggi medi complessivi di Italiano per ripartizione geografica e grado scolastico	» 47
Tab. 20 – Componenti della devianza dei punteggi medi complessivi di Italiano per ripartizione geografica e tipologia di indirizzo per il grado 13	» 47
Tab. 21 – Componenti della devianza dei punteggi medi complessivi di Italiano per ripartizione geografica e tipologia di indirizzo per il grado 10	» 48
Tab. 22 – Componenti della devianza dei punteggi medi complessivi di Mate- matica per ripartizione geografica e grado scolastico	» 48
Tab. 23 – Componenti della devianza dei punteggi medi complessivi di Mate- matica per ripartizione geografica e tipologia di Indirizzo per il grado 13	» 48
Tab. 24 – Componenti della devianza dei punteggi medi complessivi di Mate- matica per ripartizione geografica e tipologia di indirizzo per il grado 10	» 49
Tab. 25 – Componenti della devianza dei punteggi medi complessivi di Inglese <i>Reading</i> per ripartizione geografica e grado scolastico	» 49
Tab. 26 – Componenti della devianza dei punteggi medi complessivi di Inglese <i>Reading</i> per ripartizione geografica e tipologia di indirizzo per il grado 13	» 49
Tab. 27 – Componenti della devianza dei punteggi medi complessivi di Inglese <i>Listening</i> per ripartizione geografica e grado scolastico	» 50
Tab. 28 – Componenti della devianza dei punteggi medi complessivi di Inglese <i>Listening</i> per ripartizione geografica e tipologia di indirizzo per il grado 13	» 50
Tab. 29 – Errori standard per ripartizione geografica, grado scolastico e Italiano	» 56
Tab. 30 – Errori standard per ripartizione geografica, grado scolastico e Italiano	» 57
Tab. 31 – Errori standard per ripartizione geografica, grado scolastico e Matematica	» 58
Tab. 32 – Errori standard per ripartizione geografica, grado scolastico e Matematica	» 59
Tab. 33 – Errori standard per ripartizione geografica, grado scolastico e Inglese <i>Reading</i>	» 60
Tab. 34 – Errori standard per ripartizione geografica, grado scolastico e Inglese <i>Listening</i>	» 61

Tab. 35 – Statistiche per il calcolo delle varianze campionarie per stime di punteggi medi riferiti a specifiche sottopopolazioni – Italiano	pag. 66
Tab. 36 – Statistiche per il calcolo delle varianze campionarie per stime di punteggi medi riferiti a specifiche sottopopolazioni – Italiano G10	» 67
Tab. 37 – Statistiche per il calcolo delle varianze campionarie per stime di punteggi medi riferiti a specifiche sottopopolazioni – Italiano G13	» 68
Tab. 38 – Statistiche per il calcolo delle varianze campionarie per stime di punteggi medi riferiti a specifiche sottopopolazioni – Matematica	» 69
Tab. 39 – Statistiche per il calcolo delle varianze campionarie per stime di punteggi medi riferiti a specifiche sottopopolazioni – Matematica G10	» 70
Tab. 40 – Statistiche per il calcolo delle varianze campionarie per stime di punteggi medi riferiti a specifiche sottopopolazioni – Matematica G13	» 71
Tab. 41 – Statistiche per il calcolo delle varianze campionarie per stime di punteggi medi riferiti a specifiche sottopopolazioni – Inglese L	» 72
Tab. 42 – Statistiche per il calcolo delle varianze campionarie per stime di punteggi medi riferiti a specifiche sottopopolazioni – Inglese L G13	» 73
Tab. 43 – Statistiche per il calcolo delle varianze campionarie per stime di punteggi medi riferiti a specifiche sottopopolazioni – Inglese R	» 74
Tab. 44 – Statistiche per il calcolo delle varianze campionarie per stime di punteggi medi riferiti a specifiche sottopopolazioni – Inglese R G13	» 75
Tab. 45 – Allocazione del campione nella scuola primaria classi II e V: anno 2018/19	» 89
Tab. 46 – Allocazione del campione nella scuola secondaria di secondo grado classi II – Licei scientifici: anno 2018/19	» 90
Tab. 47 – Allocazione del campione nella scuola secondaria di secondo grado classi II – Licei classici: anno 2018/19	» 91
Tab. 48 – Allocazione del campione nella scuola secondaria di secondo grado classi II – Altri licei: anno 2018/19	» 92
Tab. 49 – Allocazione del campione nella scuola secondaria di secondo grado classi II – Istituti tecnici: anno 2018/19	» 93
Tab. 50 – Allocazione del campione nella scuola secondaria di secondo grado classi II – Istituti professionali: anno 2018/19	» 94

Indice delle figure

Fig. 1 – Risultati per tipologia di istituto – Italiano grado 13	pag. 35
Fig. 2 – Risultati per tipologia di istituto – Matematica grado 13	» 35
Fig. 3 – Risultati per regione Italiano G8	» 37
Fig. 4 – Risultati per regione Matematica G8	» 37
Fig. 5 – Risultati per regione Inglese <i>Listening</i> G8	» 38
Fig. 6 – Risultati per regione Inglese <i>Reading</i> G8	» 38
Fig. 7 – Variabilità dei risultati in Italiano – I ciclo di istruzione	» 51
Fig. 8 – Variabilità dei risultati in Matematica – I ciclo di istruzione	» 51
Fig. 9 – Variabilità dei risultati in Inglese <i>Listening</i> – I ciclo di istruzione	» 52
Fig. 10 – Variabilità dei risultati in Inglese <i>Reading</i> – I ciclo di istruzione	» 52
Fig. 11 – Andamento campione vs popolazione – G8 Italiano	» 78
Fig. 12 – Andamento campione vs popolazione – G8 Matematica	» 79
Fig. 13 – Andamento campione vs popolazione – G8 Inglese <i>Listening</i>	» 79
Fig. 14 – Andamento campione vs popolazione – G8 Inglese <i>Reading</i>	» 80
Fig. 15 – Andamento campione vs popolazione – G13 Italiano	» 95
Fig. 16 – Andamento campione vs popolazione – G13 Matematica	» 95
Fig. 17 – Andamento campione vs popolazione – G13 Inglese <i>Listening</i>	» 96
Fig. 18 – Andamento campione vs popolazione – G13 Inglese <i>Reading</i>	» 96
Fig. 19 – Risultati per regione Italiano G2	» 97
Fig. 20 – Risultati per regione Matematica G2	» 97
Fig. 21 – Risultati per regione Italiano G5	» 98

Fig. 22 – Risultati per regione Matematica G5	pag. 98
Fig. 23 – Risultati per regione Inglese <i>Listening</i> G5	» 99
Fig. 24 – Risultati per regione Inglese <i>Reading</i> G5	» 99
Fig. 25 – Risultati per regione Italiano G10	» 100
Fig. 26 – Risultati per regione Matematica G10	» 100
Fig. 27 – Risultati per regione Italiano G13	» 101
Fig. 28 – Risultati per regione Matematica G13	» 101
Fig. 29 – Risultati per regione Inglese <i>Listening</i> G13	» 102
Fig. 30 – Risultati per regione Inglese <i>Reading</i> G13	» 102

Le prove INVALSI sugli apprendimenti sono rivolte a tutta la popolazione degli studenti di alcuni gradi scolastici: la seconda e la quinta primaria, la terza secondaria di primo, la seconda e la quinta secondaria di secondo grado. Per garantire la qualità dei dati raccolti è estratto un campione di scuole e di classi in cui le prove si svolgono alla presenza di un osservatore esterno. In base alle elaborazioni effettuate sugli esiti delle classi campione, INVALSI fornisce i riferimenti comparativi, i *benchmark*, rispetto alle diverse aree geografiche, alle regioni e agli indirizzi di studio per la scuola secondaria di secondo grado. L'importanza di questi termini di confronto richiede un piano di campionamento definito secondo le più avanzate metodologie statistiche. In questo manuale sono illustrati gli aspetti tecnico-metodologici del piano di campionamento delle prove INVALSI, realizzato in base a una pluriennale e consolidata collaborazione tra l'Istat e l'INVALSI stesso.

Piero Demetrio Falorsi è responsabile della Direzione centrale per la metodologia e il disegno dei processi statistici dell'Istat. Ha un'esperienza trentennale nella progettazione e realizzazione degli aspetti metodologici e architettonici connessi a indagini e registri statistici. Ha guidato gruppi nazionali e internazionali di ricerca e ha rivestito ruoli di gestione di progetti di ricerca alla FAO, a Eurostat e alle Nazioni Unite.

Patrizia Falzetti è dirigente Tecnologo presso l'INVALSI, dove è la responsabile del Servizio Statistico che gestisce l'acquisizione, l'analisi e la restituzione dei dati, riguardanti le Rilevazioni nazionali e internazionali sugli apprendimenti, alle singole istituzioni scolastiche, agli *stakeholders* e alla comunità scientifica. Dal 2016 è anche Responsabile dell'Ufficio Statistico dell'INVALSI nell'ambito del Sistema statistico nazionale (SISTAN).

Roberto Ricci è dirigente di Ricerca presso l'INVALSI dove, dal 2008, è il responsabile delle Rilevazioni nazionali e il coordinatore delle aree di ricerca. È membro del Governing Board del PISA, del TALIS, dello Strategic Development Group del PISA e della General Assembly della IEA. Si è occupato del passaggio alla somministrazione informatizzata (CBT) delle prove INVALSI, anche in una prospettiva adattiva a livello individuale. È autore di molti articoli, sia divulgativi sia tecnico-metodologici, sugli aspetti legati alla misurazione standardizzata su larga scala.