

UNA PROVA SPERIMENTALE PER LE MATRICOLE DI MATEMATICA

Analisi dei risultati

di **Ottavio Giulio Rizzo**

Dipartimento di Matematica "F. Enriques"

Università degli Studi di Milano

Sunto: In questa nota studiamo le risposte date dagli studenti nella prova sperimentale presentata in questo numero da Sandra Mantovani, evidenziando in modo particolare la correlazione fra le varie risposte nei singoli test.

La metodologia

Abbiamo tabulato gli esiti delle risposte date dagli studenti alle singole domande, attribuendo il valore 0 alle risposte errate o mancanti, 1 a quelle corrette. Dopodiché, ci siamo interessati alle correlazioni fra le varie risposte, in modo da evidenziare possibili cause o spiegazioni alla mancata comprensione.

In particolare abbiamo usato due indicatori: la correlazione

$$\text{Corr}(X, Y) = \frac{\sum(X - \mu_X)(Y - \mu_Y)}{N\sigma_X\sigma_Y} = \frac{\sum XY - N\mu_X\mu_Y}{N\sigma_X\sigma_Y}$$

e uno che indicasse quanto spesso l'esercizio Y fosse stato sbagliato avendo sbagliato X :

$$C(X, Y) = \frac{\sum(1 - X)(1 - Y)}{\sum(1 - X)} = \frac{(1 - \mu_X)(1 - \mu_Y) + \sigma_X\sigma_Y\text{Corr}(X, Y)}{N(1 - \mu_X)}$$

dove X, Y sono due variabili campionarie, di media μ e deviazione standard σ , mentre N è il numero dei campioni.

Notiamo che Corr è compreso fra -1 e 1: valori negativi indicano una correlazione lineare inversa, positivi diretta. Invece C è compreso fra 0 ed 1: 0 indica che nessuno che ha sbagliato X ha poi sbagliato Y , 1 indica che tutti quelli che hanno sbagliato X hanno poi sbagliato Y . Notiamo infine che Corr è per costruzione simmetrica, mentre C non lo è (e questo è parte del suo interesse).

Le quattro domande

Le quattro domande 1, 2, 3 e 5 sono identiche per tutti gli studenti: tutte sono articolate in più punti, e ogni punto può essere corretto o sbagliato indipendentemente dagli altri. Riportiamo le domande, evidenziando in grassetto quelle corrette, insieme con la percentuale di studenti che ha risposto correttamente ad ogni punto.

Q.1 Quali tra i seguenti numeri NON è un quadrato perfetto?

Q1-A	92%	A) $(3^2 \times 5^2 \times 7^4 \times 121) \times (3^2 \times 7^2 \times 11^2)$.
Q1-B	79%	B) $(3^2 \times 8^2 \times 5^2 \times 7^4 \times 144) \times (8^2 \times 7^2 \times 12)$.
Q1-C	95%	
Q1-D	95%	C) $(3^2 \times 7 \times 11^4) \times (3^2 \times 5^2 \times 7 \times 121)$. D) $(3^2 \times 7^3 \times 11) \times (3^2 \times 5^2 \times 7 \times 11)$.
Q1-E	77%	
Q1-F	97%	E) $(3 \times 8^2 \times 5^2 \times 7^4 \times 144) \times (8^2 \times 7^2 \times 12)$. F) $8 \times 5 \times 2$.

Il valore estremamente alto di soluzioni corrette alle domande C, D, F mostra che il concetto di base di quadrato perfetto è stato compreso: le difficoltà si originano al momento di fare i calcoli.

Pochi gli studenti che hanno sbagliato a fattorizzare 121, ma c'è un'altissima correlazione fra gli errori in A e C (il 71% di chi ha sbagliato la C ha sbagliato anche la A, e il 50% di chi ha sbagliato la A ha sbagliato la C). Bassa la correlazione, invece, fra aver sbagliato B ed E.

Possiamo quindi concludere che la regola è stata compresa, e che chi ha sbagliato le (facili) domande A e C lo ha fatto perché non ha riconosciuto 121 come quadrato; mentre gli errori in B ed E sono più legati alla confusione fra la fattorizzazione di 144 e 12, piuttosto che al mancato riconoscimento di 144 come quadrato?

Gli studenti, in gran maggioranza, sanno così applicare la definizione di quadrato perfetto a casi semplici (F) oppure fattorizzare casi semplici (A, C); vengono mandati in confusione da quei casi (B,E) in cui compare sia un numero che il suo quadrato, indipendentemente dalla difficoltà di applicare la definizione. Una spiegazione alternativa è che nei casi B ed E, una percentuale non piccola di studenti si sia scordata della natura *negativa* della domanda.

Q.2

Q2-A	77%	A) Un numero primo non ha divisori propri.
Q2-B	81%	B) Ogni numero composto è il prodotto di almeno due numeri primi distinti.
Q2-C	2%	
Q2-D	96%	C) Un numero è composto se è il prodotto di numeri primi, tutti distinti tra loro.
Q2-E	99%	D) Ogni numero naturale è divisibile per 1 e per se stesso.
Q2-F	80%	
Q2-G	98%	E) Il prodotto di due numeri primi è un numero primo.
Q2-H	25%	F) Ogni numero composto è il prodotto di potenze di numeri primi, eventualmente con esponente 1.
Q2-I	80%	

G) Un numero composto è scomponibile nel prodotto di numeri primi, tutti distinti tra loro.

H) Un numero è primo se non ha divisori propri.

I) La somma di due numeri primi può essere un numero primo.

In questo esercizio va registrato il fallimento totale del punto C): solo tre studenti hanno risposto correttamente, e tutti e tre hanno svolto male l'esercizio, per cui è lecito supporre che la correttezza sia casuale.

Anche il punto H ha avuto pessimi esiti, e anche in questo caso non possiamo affermare che solo gli studenti migliori abbiano risposto correttamente: infatti, il numero medio di risposte corrette sui primi quattro quesiti è 20,6 in generale contro un 19,8 per chi ha risposto correttamente a questo punto. Il coefficiente di correlazione fra la risposta A e l'H è -0,69, che è il più alto (in valore assoluto) insieme a quello delle risposte Q3-B e Q3-C: ciò evidenzia l'incapacità di molti studenti di distinguere l'affermazione $A \Rightarrow B$ da $A \Leftarrow B$.

Molto alta la correlazione fra F ed A: in particolare i 2/3 di chi sbaglia la prima sbaglia la seconda. È difficile, però, da questi dati ipotizzarne la causa.

Q.3 Per verificare che 1987 è primo,

Q3-A	86%	A) è necessario verificare che 1987 non sia divisibile per tutti i numeri primi
Q3-B	89%	<1987.
Q3-C	92%	B) è necessario verificare che 1987 non sia divisibile per i numeri primi che sono
Q3-D	42%	<44, ma non è sufficiente.
Q3-E	82%	

C) non è sufficiente verificare che 1987 non sia divisibile per i numeri primi che sono < 44.

D) è sufficiente verificare che 1987 non sia divisibile per tutti i numeri primi <1987.

E) è sufficiente verificare che 1987 non sia divisibile per i numeri 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43.

La quasi totalità di chi ha sbagliato le risposte B e C ha sbagliato anche la E, come logica richiederebbe: gli stessi studenti hanno un punteggio medio inferiore alla media, malgrado ciò dimostrano di essere in grado di fare semplici inferenze logiche; o meglio: sono in grado di ricavare dalle stesse premesse risultati che non sono chiaramente in contrasto; infatti, solo i 2/3 di chi sbaglia B sbaglia anche C, come sarebbe richiesto dalle regole di inferenza logica.

Q.5 Con riferimento alle notazioni introdotte nel testo di Davenport, cosa accade se, applicando il metodo di Fermat a un numero primo N , otteniamo i numeri interi x e y ?

Q5-A	61%	A) $x + y = N$ e $x - y = 1$.	B) $x + y = 1$ e $x - y = N$.	C) $N = 2$.
Q5-B	80%	D) $N = x^2 - y^2$ con $x = 1$ o $y = 1$.		
Q5-C	95%	E) $N = x^2 - y^2 = (x - y)(x + y) = 9271$.		
Q5-D	83%			
Q5-E	77%			

Il 70% degli studenti dà una sola risposta, mentre il 25% ne dà due: nella gran parte dei casi, le due risposte sono la A e la B, il che evidenzia – nuovamente – una lettura troppo frettolosa del testo che porta a ritenere interscambiabili i ruoli di $x + y$ e di $x - y$.

Con l'eccezione di sopra, gli studenti tendono a scegliere una sola risposta, il che nuovamente dimostra una buona capacità di notare le contraddizioni più evidenti.

Troviamo anche interessante che solo il 4% degli studenti non abbia risposto ad alcuna domanda: chiaramente non erano ancora stanchi.

La quarta domanda

La quarta domanda è stata proposta in due versioni diverse, con lo scopo esplicito di verificare se le due diverse formulazioni danno risultati equivalenti. Notiamo che le risposte sono identiche, quindi misuriamo solo il diverso grado di comprensione. I due insiemi di studenti possono essere considerati equivalenti (la differenza nel punteggio medio delle altre quattro domande è dell'1%)

Q.4A Si consideri un numero naturale N e sia m un numero naturale tale che $m^2 > N$. Si consideri la sequenza di numeri

$$a_0 = m^2 - N, a_1 = (m + 1)^2 - N, a_2 = (m + 2)^2 - N, \dots$$

Q4A-A	95%
Q4A-B	97%
Q4A-C	97%
Q4A-D	92%
Q4A-E	60%
Q4A-F	63%

Con *differenza tra due termini consecutivi della sequenza* intendiamo il numero che si ottiene sottraendo a un termine della sequenza il termine che lo precede.

- A) La sequenza è crescente.
- B) La sequenza è decrescente.
- C) La differenza tra due termini consecutivi della sequenza è costante.
- D) La differenza tra due termini consecutivi della sequenza non è costante.
- E) La differenza tra due termini consecutivi della sequenza aumenta in modo costante.
- F) La differenza tra due termini consecutivi della sequenza aumenta, ma non in modo costante.

Q.4B Si consideri un numero naturale N e sia m un numero naturale tale che $m^2 > N$. Si considerino le due sequenze di numeri

$$1) \quad a_0 = m^2 - N, a_1 = (m + 1)^2 - N, a_2 = (m + 2)^2 - N, \dots$$

$$2) \quad \Delta_0 = a_1 - a_0, \Delta_1 = a_2 - a_1, \Delta_2 = a_3 - a_2, \dots$$

Q4A-A	90%
Q4A-B	97%
Q4A-C	91%
Q4A-D	67%
Q4A-E	74%
Q4A-F	86%

- A) La sequenza 1) è crescente.
- B) La sequenza 1) è decrescente.
- C) La differenza tra due termini consecutivi della sequenza 1) è costante.
- D) La differenza tra due termini consecutivi della sequenza 1) non è costante.
- E) La differenza tra due termini consecutivi della sequenza 2) è costante.
- F) La differenza tra due termini consecutivi della sequenza 2) non è costante.

Per entrambe le varianti, le coppie di risposte A/B, C/D e E/F non sono perfettamente correlate: vediamo però che questo è dovuto esclusivamente a casi in cui nessuna delle due risposte è stata data. Sarebbe stato interessante mescolare le risposte per rendere meno evidente le coppie antitetiche.

Come ci si potrebbe attendere, le coppie hanno una percentuale di risposte corrette decrescente. La correlazione fra l'aver sbagliato la prima coppia e la seconda coppia, sorprendentemente, non è alta — lo è rispetto alla terza — il che si può spiegare solo con l'incapacità di legare logicamente la prima coppia con la seconda.

Se confrontiamo ora la variante A con la variante B vediamo che, dal punto di vista statistico, sono differenti solo i risultati degli ultimi tre punti. Se da un lato è ben comprensibile che la formulazione B — molto più esplicita — dia esiti migliori sull'ultima coppia; dall'altro è sorprendente il crollo nel punto D.

Premettiamo che su 21 alunni che hanno sbagliato questa domanda, ben 17 non hanno risposto alla seconda coppia: gli errori sono dovuti, pertanto, alla mancata scelta. Rimane da capirne il motivo: le due coppie di domande sono funzionalmente identiche; e in entrambe le varianti occorre, come minimo, calcolare i primi valori a_i per rispondere. L'unica spiegazione ragionevole che riusciamo a supporre è che l'aver definito i Δ_i esplicitamente come $a_{i+1} - a_i$ abbia indotto non pochi studenti a riconoscere nella regolarità della definizione una sorta di *costanza*.

Se vogliamo attribuire a quest'ultimo fattore la differenza di risultato, dobbiamo dedurre che aver esplicitato maggiormente la domanda ha *peggiorato* la qualità delle risposte, demotivando gli studenti dal lavorare per interpretare la definizione degli a_i .

Il cloze

Nella correzione del cloze¹, un elemento di arbitrarietà viene introdotto dalla scelta di quali siano i termini accettabili. Nell'utilizzo standard dei test cloze per la valutazione della comprensione, l'unica risposta corretta è la parola presente nel brano originale²: questo sia per eliminare ogni arbitrarietà nel giudizio, sia — osiamo affermare *soprattutto* — per semplicità di correzione. Grazie, però, alla possibilità di correggere automaticamente i test cloze compilati online³ quest'ultimo problema decade. Rimane la soggettività nella scelta delle parole corrette: noi crediamo che sia un rischio che valga la pena correre per non considerare scorretta la frase: «se nessuno di questi due numeri è uguale a **N**»; quando la frase originale, ugualmente corretta, è: «se nessuno di questi due numeri è uguale a **1**».

Per comodità di argomentazione, riportiamo il test proposto con le parole presenti nel brano originale

¹ Vedi l'articolo di Paola Gario nel numero corrente per un'introduzione ai test cloze

² A. Gagatsis, *Come misurare la leggibilità dei testi matematici*, Pitagora editrice, Bologna 1999

³ P. Gario et al. <http://sites.google.com/site/valutarecompetenzenellescienze>

Molti matematici hanno inventato metodi che mirano a ridurre il numero di verifiche utili a decidere se un numero N sia o no primo. Il testo di⁽¹⁾ Davenport ne ha illustrati due⁽²⁾. Col primo di questi metodi⁽³⁾ si controlla se N è⁽⁴⁾ divisibile per un numero primo⁽⁵⁾ minore o uguale a \sqrt{N} ⁽⁶⁾: infatti, per verificare che N ⁽⁷⁾ è primo non è necessario⁽⁸⁾ controllare che esso non sia⁽⁹⁾ divisibile per numeri primi maggiori⁽¹⁰⁾ di \sqrt{N} . Col secondo⁽¹¹⁾, che è dovuto a Fermat⁽¹²⁾, si cerca invece di esprimere⁽¹³⁾ il numero N come differenza⁽¹⁴⁾ di due quadrati. Se N ⁽¹⁵⁾ = $a^2 - b$ ⁽¹⁶⁾², N è il prodotto⁽¹⁷⁾ dei due numeri $a - b$ e $a + b$ ⁽¹⁸⁾: se nessuno di questi due⁽¹⁹⁾ numeri è uguale a 1⁽²⁰⁾ possiamo concludere che N non⁽²¹⁾ è primo.

In questo test abbiamo visto che, a parte varianti irrilevanti come ‘>’ al posto di ‘maggiore’, gli unici casi dubbi sono il (20) dove N è una risposta matematicamente equivalente a 1 e, soprattutto, il (13) in cui le risposte date dagli studenti sono state *estremamente* varie. Abbiamo deciso di accettare come corrette, sia dal punto di vista grammaticale che di semantica matematica, i termini: *scomporre, rappresentare, vedere, scrivere*; abbiamo deciso di non accettare: *considerare, pensare, ridurre, trattare, definire, calcolare, determinare, dividere, fattorizzare, indicare, porre, trovare, verificare*.

Dividiamo i buchi in due categorie: quelli che verificano le competenze propriamente linguistiche — cioè quelli in cui la risposta è errata se risulta una frase sgrammaticata o senza senso logico — e quelli che verificano le competenze matematiche — siano queste competenze la padronanza del linguaggio tecnico (come la 13) che la capacità di applicare processi matematici per trovare la risposta corretta (come nel caso 16).

I buchi del primo tipo, e la relativa percentuale di risposte corrette, sono:

1	2	3	4	9	11	12	19
89%	91%	95%	95%	97%	81%	95%	88%

I buchi del secondo tipo sono:

5	6	7	8	10	13	14	15	16	17	18	20	21
86%	97%	95%	76%	73%	64%	79%	96%	96%	88%	96%	80%	80%

Fra i risultati scorretti, abbiamo tre studenti che non hanno risposto a *nessun* punto, ed uno che ha risposto solo ai primi punti: escludiamo questi outlier dalle considerazioni successive.

Notiamo pure che buona parte dei buchi non sono significativi, in quanto la quasi totalità delle risposte sono corrette.

Nei buchi di tipo linguistico gli errori sono dovuti al non aver compreso che *una sola parola* andava inserita (ad es., nel buco 1 molti hanno scritto «Aritmetica superiore», cioè il titolo del testo citato), il che di nuovo può essere attribuito ad una lettura frettolosa della consegna, oppure a frasi evidentemente sgrammaticate (ad es., nel buco 11, molti hanno scritto *metodo* che non è compatibile con la subordinata seguente: significativamente, il voto medio di maturità di chi ha commesso questi errori è abbondantemente inferiore alla media).

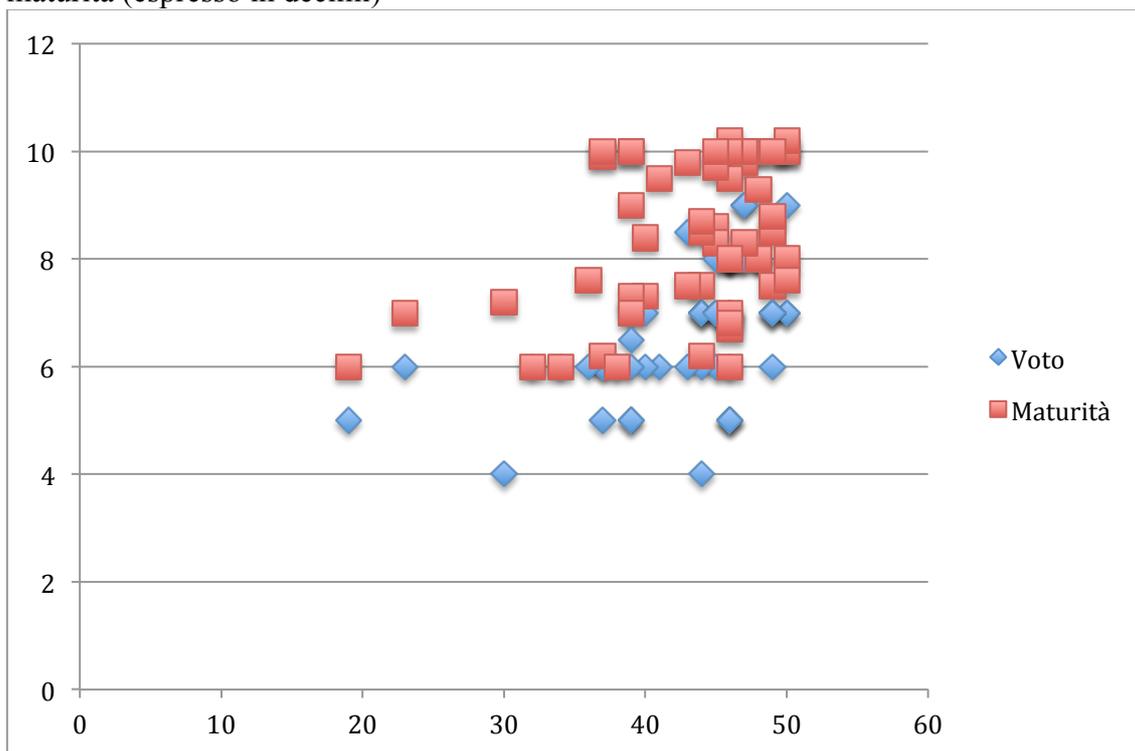
Per quanto riguarda i buchi di competenza matematica, è elevatissima la correlazione fra l'8 ed il 10: infatti molti hanno indicato, rispettivamente, *sufficiente* e *minori* come risposta, denotando una errata comprensione del testo (ma internamente coerente).

A conferma della nostra scelta di considerare il buco 13 come competenza matematica (o più specificatamente: di linguaggio matematico) c'è l'alta correlazione col buco 17: il 92% di chi ha sbagliato quest'ultimo ha sbagliato anche il precedente — si tratta, evidentemente, di studenti magari capaci di eseguire i calcoli, ma in difficoltà di fronte alla scelta del termine tecnicamente corretto (usano *quoziente* o *risultato* o *fattore* al posto di *prodotto*)

Per valutare meglio le capacità predittive del cloze, però, sarà necessario attendere la fine dell'anno accademico 2010/2011, in modo da confrontare i suoi esiti con l'andamento degli esami.

Confronto fra i voti

Confrontiamo infine i risultati grezzi del test, ottenuti contando il numero di risposte corrette, con il voto che gli studenti si sono auto attribuiti (in decimi) e il voto di maturità (espresso in decimi)



Sorprendentemente, o forse no?, la correlazione fra esito del test e voto di maturità non è molto buona: in particolare molti studenti con ottimi voti di maturità hanno avuto risultati mediocri nel test, mentre molti studenti con votazione intorno all'80 hanno avuto ottimi esiti. Molto migliore la correlazione fra autovalutazione e risultati reali.

Rinviamo ad una pubblicazione futura il confronto dell'esito del test con gli effettivi risultati nella carriera universitaria.