

Le opinioni degli studenti sulla didattica in università: sono attendibili le differenze nei risultati aggregati di Facoltà?

Andrea Scagni

*Dipartimento di Statistica e Matematica applicata alle scienze umane
Università degli Studi di Torino*

Riassunto: Negli ultimi anni l'attività di indagine sulle opinioni degli studenti sulla didattica nell'Università italiana si è andata profondamente trasformando, sia in relazione alle metodologie adottate che alle finalità implicite nel lavoro. Scopo dell'intervento è analizzare alcuni dei problemi di interpretazione e valenza dei dati raccolti, con riferimento al lavoro del Nucleo di Valutazione dell'Università di Torino. In particolare l'attenzione è mirata alle cause e i significati delle differenze tra i risultati aggregati delle facoltà di ambito disciplinare umanistico e scientifico.

L'analisi acquista grande importanza quando la crescente diffusione e standardizzazione dei metodi di indagine diviene incentivo ad un utilizzo dei risultati a fini di valutazione di intere strutture come le Facoltà o i Corsi di Laurea.

Il lavoro mostra, anche in base ad un'analisi delle componenti di devianza sui principali parametri di valutazione, le diverse chiavi di lettura di tali differenze. Si evidenzia la necessità di compendiare più elementi e approfondire la valutazione per comprenderne pienamente il significato.

Parole chiave: Qualità della didattica; Università; Indici di dispersione

1. Introduzione

Negli ultimi anni l'attività di indagine sulle opinioni degli studenti sulla didattica nell'Università italiana si è andata profondamente trasformando, sia in relazione alle metodologie adottate che alle finalità implicite nel lavoro. L'adozione di un questionario base comune ai diversi atenei, come quello del Doc. 9/02 del Comitato Nazionale per la Valutazione del Sistema Universitario (già oggi utilizzato da più del 40% degli Atenei italiani), è il presupposto di un processo che amplia la valenza di tali indagini, trasformandole potenzialmente in un ulteriore strumento di *benchmarking* quantitativo di cui, tuttavia, devono ancora essere valutate con attenzione e cautela l'attendibilità e la rilevanza.

Un approccio che le interpreta come un esercizio di analisi della *customer satisfaction* dello studente/cliente va infatti incontro a numerose e valide obiezioni, legate tra l'altro a:

- la molteplicità di soggetti individuali e collettivi che fruiscono dei “prodotti” del sistema universitario;
- il ruolo dello studente non fruitore passivo di un servizio ma partecipante attivo al processo formativo;
- l'accentuata eterogeneità delle aspettative individuali dei laureandi (l'interesse per un “lasciapassare” per svolgere una certa professione; l'aspirazione ad una crescita personale e culturale; l'attrazione per il mondo della ricerca e della scienza; il consolidamento di posizioni lavorative già acquisite; ecc.) e le inevitabili ricadute sul modo di valutare la didattica.

A tali elementi strutturali di complessità si aggiungono diversi problemi organizzativi e metodologici relativi alla conduzione dell'indagine, tra cui:

- la natura *microsystemica* dei dati e la valenza di loro analisi aggregate;
- il problema della copertura potenziale e reale;
- la rilevanza degli aspetti organizzativi e dell'efficienza dei sistemi informativi dell'Ateneo;
- le cause e i significati delle differenze tra i risultati aggregati delle facoltà di ambito disciplinare umanistico e scientifico;
- l'opportunità e le prospettive di diversificazione e verifica incrociata dei risultati mediante attività di valutazione di tipo diverso.

Da un approccio consapevole e condiviso a tali temi dipende in modo cruciale il corretto avvio della riflessione sull'utilizzo esplicito dei risultati delle rilevazioni da parte degli Atenei, ed ancor più del Ministero. Tale aspetto, già delicato in sé, specialmente nel contesto dell'Università italiana, dove la capacità docimologica ha tradizionalmente un ruolo marginale, presuppone infatti un riconoscimento chiaro del significato dei dati raccolti, prima di tutto in termini di validità statistica e capacità di rappresentare il fenomeno di interesse senza marcate distorsioni.

È d'altronde evidente che, sia per giustificare l'investimento di risorse necessario per le indagini, sia per evitare che esse si riducano, agli occhi di studenti e docenti, ad un mero adempimento amministrativo, risulta ormai urgente definire e rendere trasparenti procedure di *follow-up* coerenti e credibili.

Scopo di questo intervento è in particolare analizzare alcuni aspetti di interpretazione e valenza delle differenze riscontrabili nei risultati aggregati per facoltà, anche in riferimento ai risultati della prima Indagine coordinata a livello di Ateneo condotta dal N.d.V. dell'Ateneo dell'Università di Torino per l'a.a. 2002-2003. A tale fine vengono proposte e discusse alcune tecniche di lettura dei dati basate sulla scomposizione della dispersione relativa ai principali parametri di giudizio della qualità didattica.

2. La popolazione e la copertura: definizione e controllo

Una delle considerazioni critiche più comuni e rilevanti che vengono mosse all'indagine, in particolare alla luce dei risultati che ne derivano, è legata all'**effettiva copertura della popolazione**: gli studenti coinvolti dall'indagine non sono tutti quelli che dovrebbero parteciparvi, in alcuni casi in modo macroscopico, e in particolare tale discrepanza si potrebbe manifestare in modo diverso tra le facoltà, inficiando specificamente la possibilità di confronto tra risultati distorti in modo diseguale.

La definizione stessa della **popolazione oggetto di indagine** è, come già accennato, problematica: l'insieme degli studenti frequentanti è generalmente diverso in ogni lezione; nella maggior parte dei casi, esso presenta dimensioni decrescenti con l'avvicinarsi della fine dei corsi; infine è influenzato anche da elementi transienti (concomitanza con appelli d'esame di altre discipline, sovrapporsi temporale delle lezioni ad altre attività didattiche pratico-sperimentali, ecc.) il cui pieno controllo richiede un impegno logistico locale estremamente accurato.

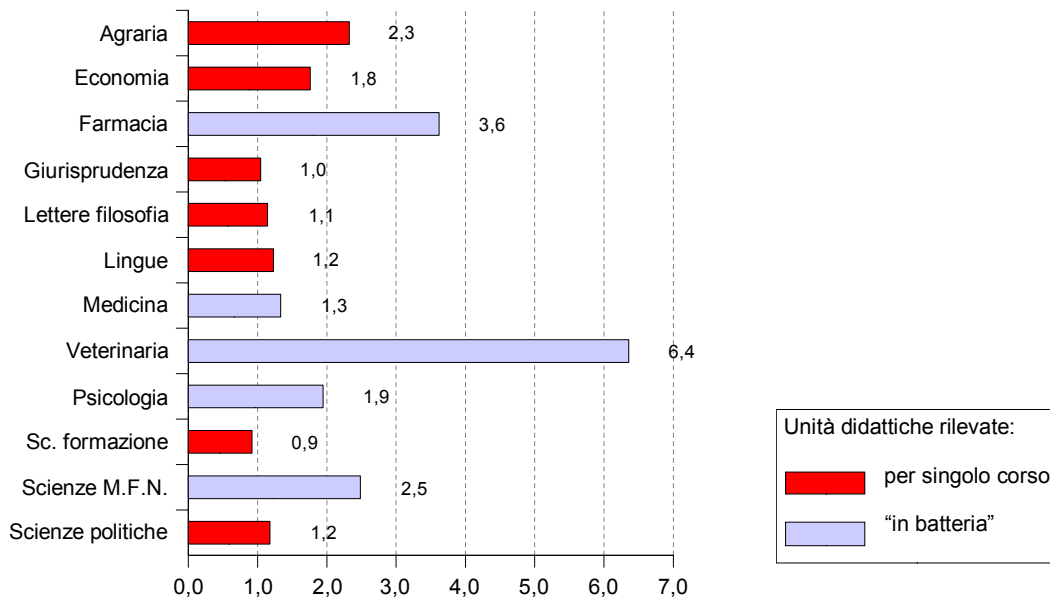
Inoltre è altrettanto difficile identificare in modo attendibile la **popolazione potenziale** per ciascun insegnamento, comprendere cioè quanti studenti *dovrebbero* frequentarne il corso durante un certo anno accademico. Per tenere sotto controllo tale fattore sarebbe necessario disporre del numero di studenti che ogni a.a. inseriscono nel proprio Piano degli Studi ciascun insegnamento; tuttavia, di norma, la codifica amministrativa delle unità didattiche non coincide con quella proposta dalle facoltà per le indagini sulle opinioni degli studenti, rendendo problematico l'abbinamento delle informazioni.

In tale situazione, l'estrema eterogeneità nel frazionamento delle unità didattiche nelle Facoltà rende impossibili valutazioni comparative della copertura in termini di rapporti tra questionari potenziali e effettivamente compilati.

La Figura 1 mostra, a titolo d'esempio, l'apparente copertura per ciascuna Facoltà dell'Ateneo torinese in termini di "unità didattiche" valutate mediamente da ciascuno studente iscritto. In tale computo ogni Facoltà ha definito **in modo autonomo** l'identificazione delle unità elementari soggette a distinta valutazione. Così ad esempio il valore pari a 6,4 di Medicina Veterinaria riflette la nota polverizzazione delle attività didattiche presso tale facoltà, con "corsi integrati" spesso costituiti dall'accorpamento di molti brevi moduli di 10-15 ore ciascuno. Anche l'analisi di dettaglio tra i Corsi di Studio di una stessa facoltà evidenzia quasi sempre forte eterogeneità.

La Figura 1 evidenzia inoltre un'ulteriore particolarità dell'indagine torinese; per meglio adattarsi alle differenti modalità di organizzazione della didattica, sono previsti due metodi alternativi di rilevazione dei dati, recependo le istanze delle Facoltà che negli anni precedenti li avevano sperimentati con successo:

- per singolo corso (in scuro nella Figura 1): un questionario distinto viene compilato dallo studente per ciascun insegnamento o corso integrato frequentato, identificato come unità didattica da un unico momento di verifica del profitto (un esame unico con un voto finale unico, anche se eventualmente condotto in sequenza da più colla-

Figura 1. *N. di corsi o moduli didattici valutati per iscritto, disaggregazione per le Facoltà dell'Università di Torino, indagine '02-03*

boratori alla docenza). Il singolo questionario prevede comunque la possibilità di esprimere con identiche modalità un giudizio su più docenti, collaboratori, esercitatori ecc. fino ad un massimo di quattro soggetti, mentre sugli aspetti più generali del corso il giudizio resta unico. I questionari "per singolo corso" ben si adattano ai contesti in cui lo studente si organizza in modo prevalentemente autonomo l'attività di frequenza e studio.

- per tutti i moduli didattici di un semestre "in batteria" (in chiaro nella Figura 1): adatto ai Corsi di Studio caratterizzati da una programmazione didattica rigidamente strutturata, prevede un momento unificato di rilevazione, di durata maggiore, in cui lo studente è chiamato a fornire indicazioni per tutti i moduli didattici frequentati durante il semestre in esame. Il questionario (di aspetto diverso da quello della rilevazione per singolo corso) contiene un certo numero di domande con possibilità di fornire risposte "in batteria" distinte per ciascun modulo, mentre le domande più generali sono unificate. In tal modo è possibile limitare sia l'impegno richiesto allo studente per la compilazione che il disturbo all'attività didattica. Tuttavia, nelle Facoltà in cui la strutturazione della didattica non è rigida e i corsi non sono frequentati da coorti omogenee di studenti, è quasi inevitabile la somministrazione separata corso per corso.

Da un punto di vista metodologico la rilevazione "in batteria" si caratterizza inevitabilmente per il **maggiore sforzo di memoria** richiesto ai rispondenti, che devono nello stesso momento effettuare valutazioni su un notevole numero di fenomeni, alcuni dei quali riferiti a moduli didattici che potrebbero essersi addirittura già

conclusi da qualche tempo. Per contro, il momento unitario di rilevazione permette al rispondente di valutare anche *in termini effettivamente comparativi* le valutazioni fornite, dato che le risposte sui vari moduli fisicamente affiancate sullo stesso questionario rendono immediato il confronto tra i giudizi espressi per i medesimi. Con la tradizionale rilevazione per singolo corso il rispondente difficilmente può ricordare nel dettaglio le valutazioni espresse giorni prima su altri insegnamenti e molto più forte è quindi il rischio di incoerenza trasversale nei giudizi dovuta a fattori contingenti e momentanei.

La scelta tra tali alternative di rilevazione ha comunque anch'essa influenzato il livello di frazionamento delle unità didattiche, generalmente più spinto presso le facoltà dove si è adottato l'approccio "in batteria".

In conclusione, pur ricordando come la norma di legge richieda ai N.d.V. la raccolta di informazioni sul grado di soddisfazione per la didattica dei **sol** studenti **frequentanti**, è opportuna un'analisi accurata dei fattori macro e micro che determinano la dimensione della quota di studenti frequentanti: estremizzando, è evidente che un corso valutato in modo eccellente, ma solo da una ristretta quota degli studenti che ne sosterranno poi l'esame non rappresenta necessariamente un caso di successo, per lo meno in relazione ad un efficiente impiego delle risorse umane e strutturali. Inoltre, come discusso nel successivo paragrafo, dietro a situazioni di questo genere si possono nascondere problemi di distorsione anche rilevanti.

3. La copertura e gli effetti distorsivi sui dati

Il problema generale di copertura è in realtà legato non solo alla rappresentatività dei dati raccolti rispetto all'intera popolazione studentesca, ma anche al rischio di distorsione dei risultati che le mancate risposte possono indurre.

L'esigenza di rilevare l'opinione dello studente verso la fine dei corsi, in modo che questa sia espressa in modo consapevole, influisce ulteriormente sulla copertura, dato che nella maggior parte dei casi la classe presenta dimensioni decrescenti con l'avvicinarsi della fine dei corsi. In tal modo, alla mancata rilevazione delle opinioni di coloro che non hanno mai frequentato, si aggiunge quella di chi ha rinunciato alla frequenza dopo una o più lezioni.

I motivi di tale processo di *autoselezione* sono vari, e possono essere classificati come segue in relazione ai problemi posti all'analisi dei dati:

- a) casi mancanti casualmente (*missing at random*), relativi a studenti che hanno deciso di non frequentare per fattori del tutto estranei ai temi che l'indagine studia;
- b) casi mancanti non casualmente, relativi a studenti che non frequentano per motivi legati direttamente o indirettamente agli aspetti valutati nell'indagine.

Nel primo caso non si evidenziano effetti distorsivi sui risultati, ma solo una minore rappresentatività della popolazione. Nel secondo caso, invece, vi sono mag-

giori rischi di produrre un quadro non completamente attendibile della situazione. Il fattore più evidente in questo senso è l'*insoddisfazione* e/o lo *scarso interesse* che le prime lezioni possono avere suscitato nello studente, inducendolo a non frequentare più, oppure a rarefare la propria presenza con elevata probabilità di essere assente al momento della somministrazione del questionario¹.

Altro fattore rilevante è anche la percezione di *utilità* (o addirittura *necessità*) della frequenza in funzione del superamento dell'esame. Ad esempio, un corso ritenuto generalmente poco interessante può risultare molto frequentato fino alla fine perché si ritiene sia comunque difficile superare l'esame con una preparazione sui soli testi. La combinazione di tali fattori può comportare una distorsione nei risultati apparenti dell'indagine in termini comparativi; si pensi alla contrapposizione tra:

1. un corso di studi che prevede corsi di lezione quasi tutti ritenuti "necessari" per l'esame, e che quindi "costringe" alla frequenza fino alla fine anche gli studenti insoddisfatti dalla qualità dei corsi;
2. un corso di studi dove molti corsi di lezione risultano superabili anche con una preparazione sui soli libri di testo, e in cui quindi frequentano quasi sempre solo gli studenti che ritengono interessanti e utili in sé le lezioni.

Con ogni probabilità l'indagine produrrà risultati *più soddisfacenti* per il corso di studi 2; tuttavia, se fosse possibile coinvolgere per tale c.d.s. tutti gli studenti che hanno partecipato alle prime lezioni i risultati potrebbero diventare equivalenti o addirittura inferiori a quelli emersi per il c.d.s. 1².

In conclusione può essere elevato il rischio di correlazione tra le scelte di frequenza e il livello di soddisfazione sui diversi aspetti legati all'erogazione della didattica: dove la frequenza è minore, e quindi anche la copertura dell'indagine è modesta, gli studenti partecipanti all'indagine saranno più facilmente quelli convinti delle qualità positive del docente e del suo corso.

4. Il significato delle differenze tra risultati aggregati: l'esempio del contrasto tra ambito umanistico e scientifico

L'effetto descritto nel paragrafo precedente potrebbe spiegare, in proporzione anche rilevante, le valutazioni aggregate meno positive delle Facoltà di ambito scientifico rispetto a quelle di ambito umanistico, emerse anche a Torino nell'indagine 2002-03.

Per quantificare in modo sintetico le valutazioni, è stata applicata una codifica

¹ In alcuni casi il diffondersi di opinioni negative tra gli studenti può portare a rinunciare alla frequenza sin dall'inizio, rendendo mancanti non casualmente anche alcuni degli studenti che non hanno frequentato del tutto il corso.

² Si tratta di un'ipotesi di difficile verifica empirica, dato che gli studenti che hanno smesso di frequentare dopo poche lezioni (spesso una o due solamente) potrebbero non essere in grado di valutare il corso in modo comparabile a chi l'ha frequentato interamente.

numerica arbitraria, con interi da 1 a 4, alla scala di risposte prevista dal CNVSU³.

Con tale codifica "naturale" la media dei giudizi può essere espressa in funzione delle sole frequenze relative⁴. Infatti, denominata X la variabile discreta così ottenuta, $k = 4$ il numero di modalità possibili, n_i le frequenze assolute con cui ogni modalità è stata osservata nella popolazione di interesse di dimensione N , e definite:

$$f_i = \frac{n_i}{N}; \quad F_i = \sum_{j=1}^i f_j$$

si ha:

$$\begin{aligned} \bar{x} &= \sum_{i=1}^k x_i f_i = \sum_{i=1}^k i(F_i - F_{i-1}) = \sum_{i=1}^k i F_i - \sum_{i=2}^k i F_{i-1} = \\ &= \sum_{i=1}^k i F_i - \sum_{i=1}^{k-1} (i+1) F_i = k F_k + \sum_{i=1}^{k-1} i F_i - \sum_{i=1}^{k-1} (i+1) F_i = k - \sum_{i=1}^{k-1} F_i \end{aligned}$$

La Figura 2 mostra, in base a tali valori medi calcolati per due tra le principali domande di valutazione, la collocazione delle Facoltà dell'Ateneo torinese.

Un altro elemento di interesse è il confronto tra le collocazioni delle facoltà in termini di valori medi e quelle in termini di variabilità interna dei giudizi. Per valutare tale aspetto può essere utilizzato un indice di dispersione per caratteri qualitativi ordinali, come quello di diversità (Leti, 1983, pp. 290-297):

$$D = 2 \sum_{i=1}^{k-1} F_i (1 - F_i).$$

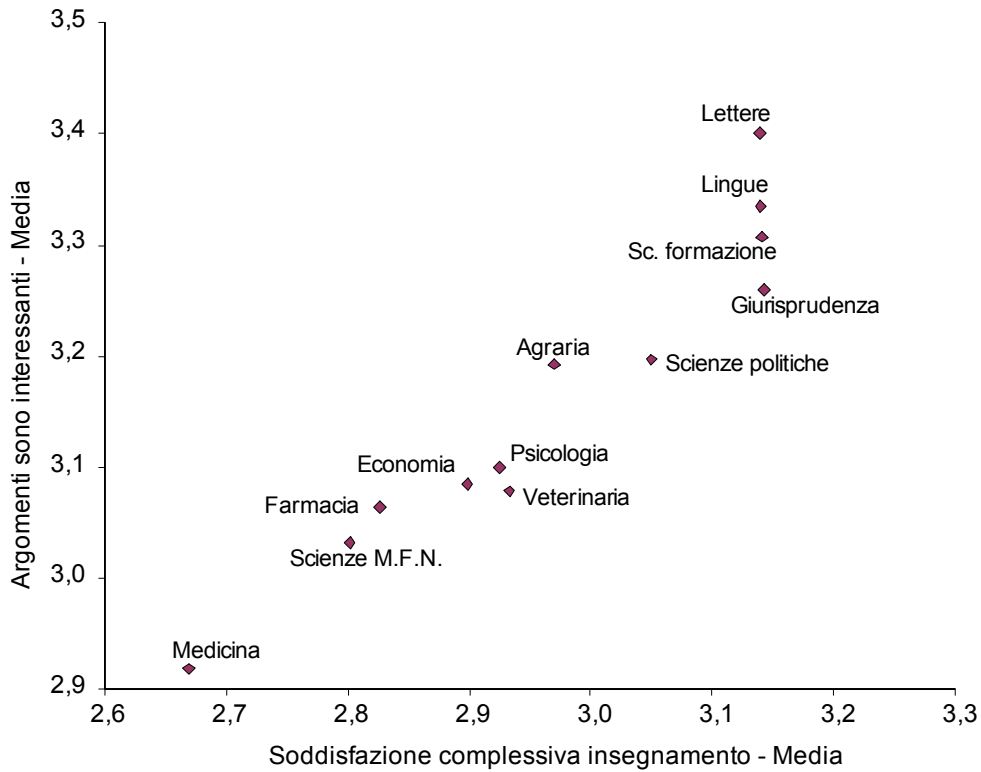
In alternativa può essere utilizzato lo scarto quadratico medio della variabile ottenuta dalla codifica sopra descritta. In tal caso, infatti, i due indici hanno un comportamento molto simile, presentando lo stesso campo di variazione da 0 a $(k-1)/2$ (asintoticamente per N dispari nel caso di D). Si ha, infatti:

$$\sigma^2 = \sum_{i=1}^k (x_i - \bar{x})^2 f_i = \sum_{i=1}^k (x_i)^2 f_i - \bar{x}^2;$$

³ *Decisamente no* = 1; *Più no che sì* = 2; *Più sì che no* = 3; *Decisamente sì* = 4. L'utilizzo della scala di risposta proposta dal CNVSU nell'indagine è derivato dalla scelta dell'Ateneo di aderire in modo completo alle indicazioni del DOC 9/02, anche al fine di contribuire al processo di armonizzazione a livello nazionale delle procedure relative alla rilevazione delle opinioni degli studenti. In tal senso non è stata effettuata un'analisi specifica della validità della rappresentazione delle opinioni rilevate mediante tale scala. In realtà, il gruppo di lavoro composto dal N.d.V. e dai delegati di tutte le facoltà aveva deciso, poche settimane prima della pubblicazione del DOC 9/02, di utilizzare una scala decimale esplicitamente numerica per tutte le domande, in base a considerazioni di natura metodologica.

⁴ La scelta di usare la semplice trasformazione in ranghi delle modalità di risposta, invece di tentare più complessi approcci di approssimazione di una quantificazione ottimale, è dovuta a diversi motivi: la semplicità delle elaborazioni seguenti che essa rende possibile; la coerenza con i risultati delle elaborazioni di base che vengono prodotte in via immediata dalla società che effettua la lettura dei dati a livello di singolo insegnamento, di corso di laurea e Facoltà; le finalità non di quantificazione assoluta ma di comparazione relativa tra aggregati che rendono meno urgente disporre di una traduzione quantitativa intrinsecamente attendibile.

Figura 2. *Media dei giudizi - disaggregazione per le Facoltà dell'Università di Torino, indagine 02-03*



Il massimo di σ^2 , nel caso presente, si ha quando

$$f_i = \frac{1}{2}; i=1, k \quad f_i = 0; i=2, 3, \dots, k-1 \Rightarrow$$

$$\Rightarrow \sigma_{max}^2 = \frac{k^2+1}{2} - \left(\frac{k+1}{2}\right)^2 = \frac{k^2-2k+1}{4} = \left(\frac{k-1}{2}\right)^2 \Rightarrow \sigma_{max} = \frac{k-1}{2}$$

Nel caso in esame ha quindi un significato preciso il concetto di *scarto quadratico medio relativo*, ottenuto dividendo il valore di definizione per il suo massimo:

$$\sigma_r = \frac{\sigma}{\sigma_{max}} = \frac{\sigma}{\frac{k-1}{2}}$$

Inoltre, su una variabile X come quella utilizzata in questa sede, l'indice di diversità D e lo scarto quadratico medio, anch'esso esprimibile in funzione delle sole F_i :

$$\begin{aligned}
\sigma &= \left[k^2 - \sum_{i=1}^{k-1} (2i+1) F_i - \left(k - \sum_{i=1}^{k-1} F_i \right)^2 \right]^{1/2} = \\
&= \left[k^2 - \left(2 \sum_{i=1}^{k-1} i F_i \right) - \left(\sum_{i=1}^{k-1} F_i \right) - k^2 - \left(\sum_{i=1}^{k-1} F_i \right)^2 + 2k \sum_{i=1}^{k-1} F_i \right]^{1/2} = \\
&= \left[\sum_{i=1}^{k-1} F_i \left(2k-1 - \sum_{i=1}^{k-1} F_i \right) - 2 \sum_{i=1}^{k-1} i F_i \right]^{1/2} = \\
&= \left[\sum_{i=1}^{k-1} F_i \left(2k-1 - \sum_{i=1}^{k-1} F_i \right) - 2 \sum_{i=1}^{k-1} \sum_{j=i}^{k-1} F_j \right]^{1/2}
\end{aligned}$$

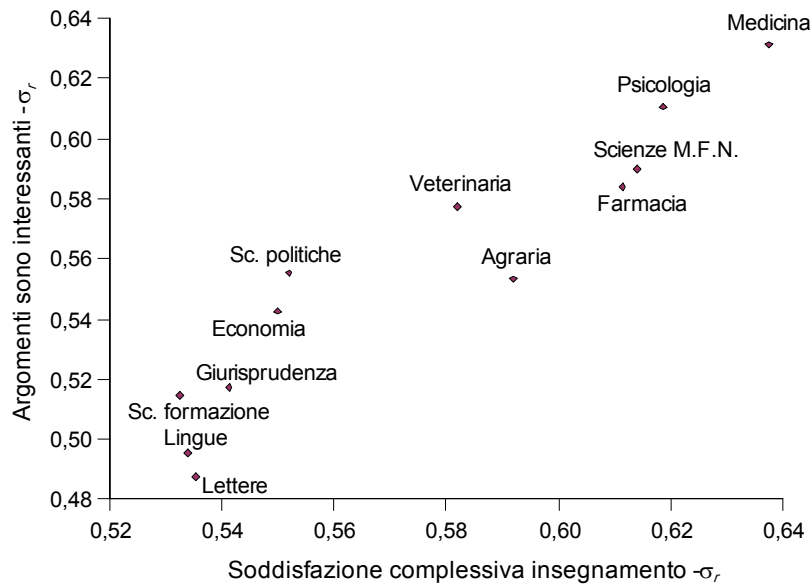
pur restando funzioni distinte, presentano un comportamento simile su tutto il campo di variazione. Le Figure in Appendice mostrano i valori di σ e D ottenuti in una simulazione casuale di 130 distribuzioni multinomiali, ordinate in base al valore di σ , e in una sequenza di distribuzioni ad hoc con crescente dispersione con $N = 50$; in entrambi i casi l'indice D risulta leggermente più sensibile, con valori prima inferiori a σ vicino a 0 e successivamente superiori.

Un'altra similarità rilevante dei due indici è la possibilità di scomposizione tra gruppi della dispersione, nel modo usuale per la varianza e come evidenziato da Grilli e Rampichini (2004) per l'indice di diversità. Infatti, distinta la popolazione di interesse in h sottogruppi di dimensione N_j ($j=1,2,\dots,h$), e indicate con F_{ij} le frequenze cumulate relative fino alla modalità i nel gruppo j , si ha:

$$\begin{aligned}
D &= 2 \sum_{j=1}^h \frac{N_j}{N} \left[\sum_{i=1}^{k-1} F_{i|j} (1 - F_{i|j}) + \sum_{i=1}^{k-1} F_{i|j} (F_{i|j} - F_i) \right] = \\
&= 2 \sum_{j=1}^h \frac{N_j}{N} \left[\sum_{i=1}^{k-1} D_j + \sum_{i=1}^{k-1} (F_{i|j} - F_i)^2 \right]
\end{aligned}$$

Nell'espressione il primo addendo rappresenta una media ponderata degli indici di dispersione "locali" D_j , mentre il secondo può essere visto come una misura della dissomiglianza media tra la distribuzione sui singoli gruppi e quella sulla popolazione complessiva.

La Figura 3 presenta, in modo simile alla Figura 2, i valori di σ_r per le 12 facoltà torinesi rispetto ai giudizi di tutti gli studenti per le domande indicate. Il confronto tra i due grafici mostra che i due gruppi di Facoltà (rispettivamente scientifiche e umanistiche) che si evidenziano nel primo caso (Figura 2) si trovano in posizione ribaltata nel secondo. In altre parole, dove la valutazione media è più bassa vi è però maggiore varietà di giudizio; dove la media è invece più elevata, le opinioni sono più concordi. Ciò potrebbe essere dovuto proprio al fatto che nelle Facoltà dove la frequenza è più

Figura 3. Variabilità dei giudizi - disaggregazione per le Facoltà dell'Università di Torino, indagine '02-03

generalizzata partecipano maggiormente all'indagine, valutando negativamente i corsi, anche studenti che in altri contesti avrebbero rinunciato alla presenza in aula.

Naturalmente si tratta di una interpretazione tutt'altro che univoca; altri effetti potrebbero avere un ruolo significativo. Ad esempio, un esame delle componenti di dispersione distinte⁵, all'interno di singole Facoltà o loro raggruppamenti, tra variabilità del giudizio medio tra i diversi insegnamenti (*inter-corso*) e variabilità dei giudizi dei singoli studenti per uno stesso insegnamento (*intra-corso*) evidenzia, passando dalle Facoltà umanistico-letterarie a quelle di ambito scientifico, un incremento molto più sensibile della variabilità inter-corso rispetto a quella intra-corso. In altri termini, presso le Facoltà scientifiche si evidenzia non tanto un'eterogeneità di giudizi maggiore tra gli studenti che valutano uno stesso insegnamento, quanto una maggiore presenza di insegnamenti "critici" in cui la valutazione di tutti i frequentanti converge su giudizi vicini al limite inferiore della scala.

Le Tabelle 1 e 2 e la Figura 4 dettagliano tale contrasto con riferimento sia alla varianza che all'indice D di Leti. Le indicazioni fornite dalla scomposizione dei due indici sono pressoché identiche: si registra solo l'inversione nell'ordine delle facoltà di Economia e Sc. Politiche, comunque contigue. In modo analogo ai valori assunti, anche le percentuali di dispersione tra i corsi sono leggermente più discriminate dall'indice D , per cui la percentuale di variabilità inter-corso più elevata (Medicina, 27,5) è più del doppio della percentuale minore (Economia, 13,4).

⁵ Riferite alla domanda sulla *Soddisfazione complessiva per come è stato svolto l'insegnamento*.

Tabella 1. Variabilità totale, inter-corso e intra-corso dei giudizi di Soddisfazione complessiva sul corso per le Facoltà dell'Università di Torino, indagine 2002-03

	σ^2 totale	σ^2 inter-corso	σ^2 intra-corso	% inter-corso	% intra-corso
Medicina	0,925	0,346	0,579	37,4	62,6
Veterinaria	0,777	0,268	0,509	34,5	65,5
Farmacia	0,850	0,275	0,575	32,3	67,7
Scienze M.F.N.	0,866	0,269	0,597	31,1	68,9
Psicologia	0,874	0,267	0,607	30,6	69,4
Agraria	0,821	0,203	0,618	24,8	75,2
Lettere	0,658	0,162	0,496	24,6	75,4
Lingue	0,652	0,148	0,504	22,7	77,3
Sc. formazione	0,649	0,134	0,515	20,7	79,3
Giurisprudenza	0,669	0,131	0,538	19,6	80,4
Economia	0,689	0,135	0,554	19,5	80,5
Sc. politiche	0,701	0,132	0,570	18,8	81,2
Fac. Umanistiche	0,725	0,170	0,555	23,5	76,5
Fac. Scientifiche	0,856	0,277	0,579	32,4	67,6

Figura 4. Composizione % della variabilità totale tra inter-corso e intra-corso dei giudizi di soddisfazione complessiva sul corso per le Facoltà dell'Università di Torino, indagine 2002-03 - Varianza e Indice di diversità D

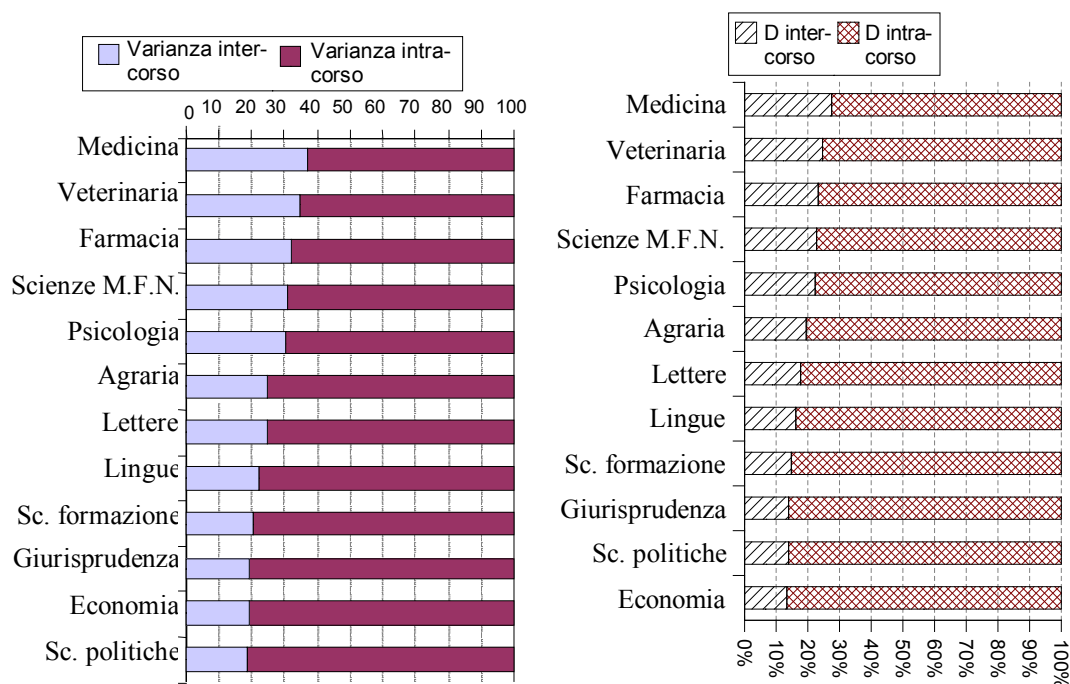


Tabella 2. *Indice di diversità D totale, inter-corso e intra-corso dei giudizi di Soddisfazione complessiva sul corso per le Facoltà dell'Università di Torino, indagine 2002-03*

	<i>D</i> totale	<i>D</i> inter-corso	<i>D</i> intra-corso	% inter-corso	% intra-corso
Medicina	1,037	0,285	0,752	27,5	72,5
Veterinaria	0,925	0,228	0,697	24,6	75,4
Farmacia	0,986	0,229	0,757	23,3	76,7
Scienze M.F.N.	0,994	0,226	0,768	22,8	77,2
Psicologia	1,000	0,223	0,776	22,3	77,7
Agraria	0,945	0,184	0,761	19,5	80,5
Lettere	0,831	0,147	0,683	17,7	82,3
Lingue	0,828	0,134	0,694	16,2	83,8
Sc. formazione	0,827	0,122	0,705	14,8	85,2
Giurisprudenza	0,839	0,117	0,722	13,9	86,1
Sc. politiche	0,865	0,120	0,745	13,9	86,1
Economia	0,861	0,115	0,746	13,4	86,6

Un riscontro della suddetta ipotesi è ricavabile dal computo dell'incidenza degli insegnamenti che ottengono un giudizio medio inferiore a 2 (cioè più critico della moderata insoddisfazione, e con modesta eterogeneità dei pareri dato lo schiacciamento verso il limite inferiore della scala), con un 5,5% in area scientifica e un 1,7% in area umanistica.

D'altro canto la relazione tra ambito disciplinare e livello della varianza intra-corso potrebbe essere *in parte spuria*: in generale tra i corsi con classi di maggiori dimensioni essa tende ad assumere peso maggiore, e questo avviene sia in ambito umanistico che scientifico: il suo peso percentuale passa dal 72% al 91% per l'umanistico e dal 67% all'89% per lo scientifico, distinguendo tra le classi con meno di 50 frequentanti e le restanti. Tuttavia, essendo la dimensione media delle classi maggiore in ambito umanistico (45 studenti contro 32), la differenza della varianza intra-corso tra i due ambiti disciplinari potrebbe essere in buona parte il riflesso della maggiore dimensione delle classi in ambito umanistico. A sua volta, questa appare difficile da interpretare, essendo sicuramente influenzato dalla prevalenza di impiego del questionario per singolo corso nelle facoltà umanistiche e del questionario “in batteria” tra quelle scientifiche.

5. Conclusioni

Come evidenziato nelle riflessioni precedenti, numerosi fattori di carattere logistico, metodologico, interpretativo e psicologico rendono incerta l'attendibilità dei dati ottenuti dalle indagini sulle opinioni degli studenti sulla qualità della didattica in termini aggregati per “tastare il polso” di un corso di studio o di una intera facoltà. L'esperienza di prima applicazione del Doc. 9/02 del CNVSU ha fornito al riguardo

significativi elementi di valutazione e ha contribuito ad incrementare la consapevolezza della complessità di tali fattori. In questo senso due sono le principali direzioni in cui si ritiene sia opportuno muoversi:

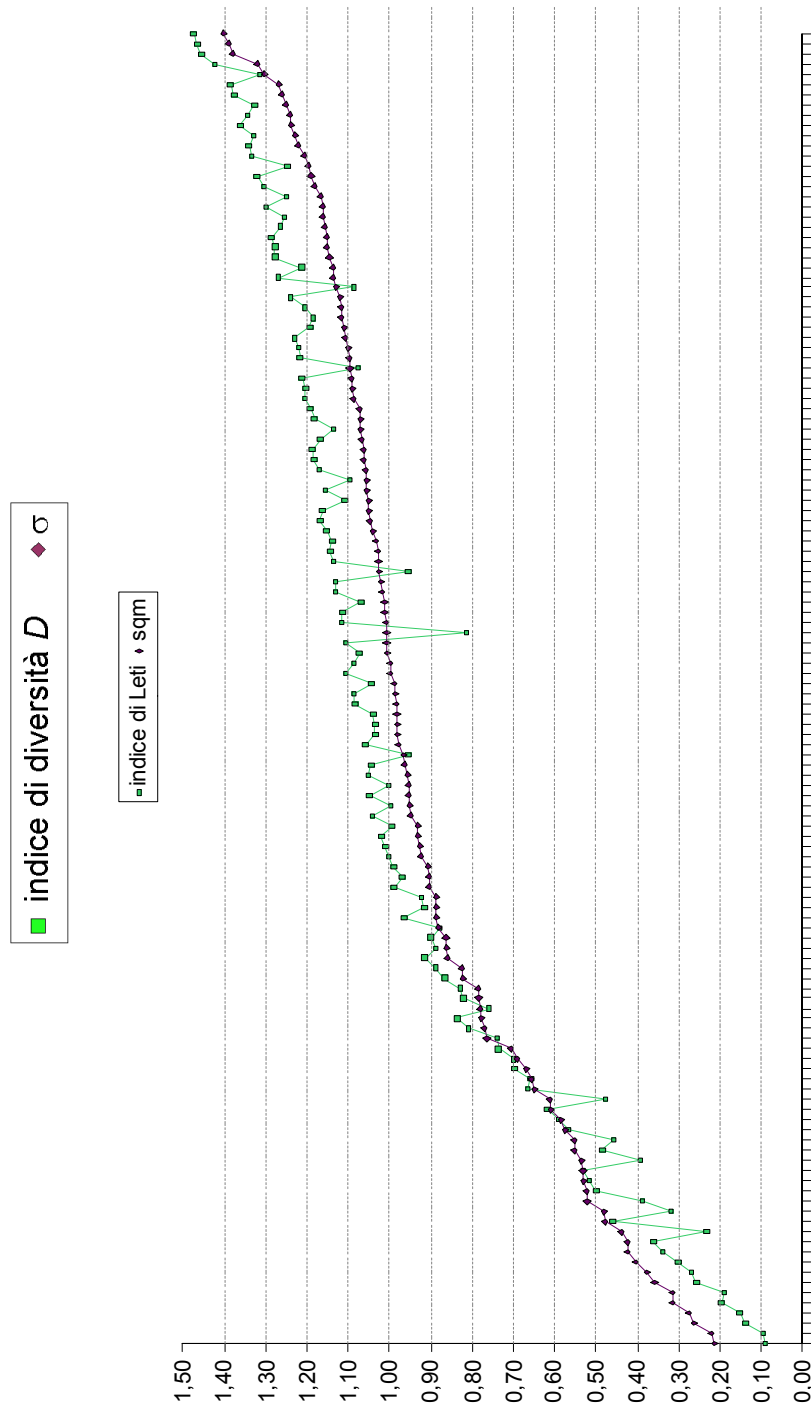
- sperimentazione e verifica di metodologie di indagine esplicitamente rivolte alla valutazione degli aspetti aggregati, anche al fine di confrontarne i risultati con quelle delle attuali indagini sulla qualità della didattica, verificando empiricamente le eventuali distorsioni che queste ultime comportano;
- impulso verso il rafforzamento e razionalizzazione dei sistemi informativi sull'offerta didattica, come base di dati univoca ed attendibile su cui basare il lavoro di indagine in modo efficiente e consapevole.

E' comunque opportuno ricordare che i risultati descritti, come molti altri dati ricavati dall'indagine, potrebbero essere più utilmente inquadrati in un contesto nazionale in cui il ruolo del CNVSU è preminente. Tuttavia ad oggi sono solo parzialmente noti gli indirizzi di lavoro e analisi che il CNVSU sta perseguendo o intende concretizzare nel prossimo futuro al riguardo. Su questi aspetti è sicuramente auspicabile una maggiore trasparenza sull'attività del Comitato Nazionale, nonché una maggiore diffusione da parte del Comitato di informazioni relative allo status del lavoro di valutazione della didattica per il complesso del sistema universitario italiano.

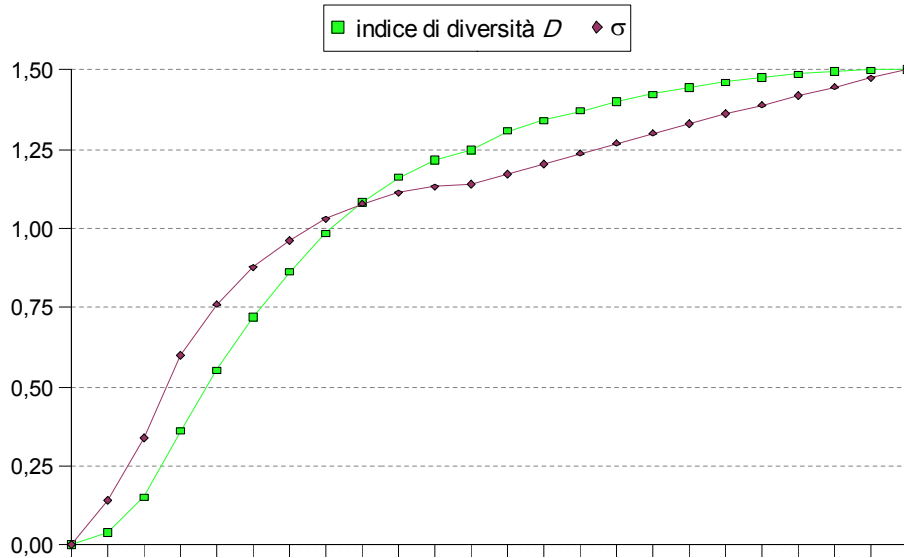
Riferimenti bibliografici

- Comitato Nazionale Valutazione del Sistema Universitario (CNVSU) (2002), *Proposta di un insieme minimo di domande per la valutazione della didattica da parte degli studenti frequentanti*, DOC 09/02, Rapporto finale del gruppo di lavoro approvato dal Comitato nella seduta del 26 luglio 2002.
- GRILLI L. e RAMPICHINI C. (2002) Scomposizione della dispersione per variabili ordinali, *Statistica*, anno LXII, 1: 111-116.
- VIOLANI C. (2001), Rilevazione delle opinioni degli studenti sulla didattica nell'Università di Roma "La Sapienza", *CAPTOR 2000: Qualità della didattica e sistemi computer-assisted*, Cleup, Padova: 11-26.

Appendice A - Valori di D e σ su 130 distribuzioni multinomiali a 4 modalità simulate mediante estrazione casuale dei valori delle frequenze, ordinate in base a σ



Appendice B - Valori di D e σ su 130 distribuzioni multinomiali a 4 modalità simulate mediante estrazione casuale dei valori delle frequenze



x_i	1	2	3	4	D di Leti	s.q.m.
n_i	50	0	0	0	0,0000	0,0000
	49	1	0	0	0,0392	0,1400
	47	2	1	0	0,1520	0,3370
	44	3	2	1	0,3632	0,6000
	41	4	3	2	0,5520	0,7600
	38	5	4	3	0,7184	0,8754
	35	6	5	4	0,8624	0,9625
	32	7	6	5	0,9840	1,0284
	29	8	7	6	1,0832	1,0770
	26	9	8	7	1,1600	1,1107
	23	10	9	8	1,2144	1,1307
	20	11	10	9	1,2464	1,1377
	14	11	11	14	1,3064	1,1705
	15	10	10	15	1,3400	1,2042
	16	9	9	16	1,3704	1,2369
	17	8	8	17	1,3976	1,2689
	18	7	7	18	1,4216	1,3000
	19	6	6	19	1,4424	1,3304
	20	5	5	20	1,4600	1,3601
	21	4	4	21	1,4744	1,3892
	22	3	3	22	1,4856	1,4177
	23	2	2	23	1,4936	1,4457
	24	1	1	24	1,4984	1,4731
	25	0	0	25	1,5000	1,5000

Student's opinions of university teaching: are differences among aggregate results reliable?

Summary: *In recent years, student's opinion surveys on teaching quality in Italian Universities have changed considerably in methods and scope. This paper considers the validity and meaning of the data from such surveys, with reference to the work conducted by the "Nucleo di valutazione" (Evaluation Committee) of Turin University. Two sets of problems are of particular interest:*

- *the individual context of the data vs. their relevance in aggregate analysis;*
- *causes and meanings of the differences between aggregate results for nature sciences Faculties and human sciences Faculties.*

The first item has great relevance due to the diffusion and standardization of survey techniques, which encourages the assessment of whole Faculties and Courses based on the collected data.

Regarding the second item, some of the possible factors behind such differences are assessed and compared, also considering appropriate dispersion decomposition techniques. It is shown how many effects must be considered to really understand the nature of this phenomenon.

Keywords. *Teaching quality; University; Dispersion indexes*