

## Un elemento di valutazione della qualità della didattica universitaria

Gianluca Sbardella\*, Francesco Sebastianelli\*, Carlo Mariani\*,  
Vincenzo Nesi\*, Andrea Pelissetto\*

**Abstract:** For many years, Italian Universities have been committed to collecting “assessment questionnaires” of the courses from their students. It is not fully clear how to manage these questionnaires, though. In the present work, we propose a method to single out significant information from the wide amount of issues raised by the students. With this work, we aim at improving the services offered by Universities. Human resources dedicated to the evaluation of teaching are definitely scarce, if not absent. Thus, we believe that teaching evaluation committees must focus on the most critical issues, rather than dispersing information.

In this work, we show how an almost automatic method allows one to single out a very limited fraction (about 7% in the case of our Faculty of Natural Sciences at Sapienza University in Rome) of really critical courses. Our political objective is to give a clear and objective answer to the students’ request of improving those courses. With this approach we do not consider the students’ opinions as absolute. We rather aim at respecting their implicit request of verification of their opinions.

Our approach also suggests some improvements in the formulation of the questions raised in the questionnaires, so as to reduce ambiguity and obtain a more accurate statistical treatment. In the example presented here, a very high appreciation of our courses emerges if we focus on the average values. However, a punctual control of opinions for a fraction of courses is indispensable. Technically, we use the so called trimming method to single out the problematic courses. This method, used in a variety of fields, helps to control for carelessness of some students who may not take the questionnaires seriously. In our case, for each course and for each question, we neglect 50% of the questionnaires, namely those corresponding to 25% of the most positive and 25% of the most negative. Still, our analysis leads to very stable results, if we focus on the analysis of the most critical courses, which is the main objective of this work.

---

\* Facoltà di Scienze Matematiche, Fisiche e Naturali, Sapienza Università di Roma [gianluca.sbardella@uniroma1.it](mailto:gianluca.sbardella@uniroma1.it)

\* Facoltà di Scienze Matematiche, Fisiche e Naturali, Sapienza Università di Roma  
[francesco.sebastianelli@uniroma1.it](mailto:francesco.sebastianelli@uniroma1.it)

\* Facoltà di Scienze Matematiche, Fisiche e Naturali, Sapienza Università di Roma [carlo.mariani@uniroma1.it](mailto:carlo.mariani@uniroma1.it)

\* Facoltà di Scienze Matematiche, Fisiche e Naturali, Sapienza Università di Roma [presidenzasmfn@uniroma1.it](mailto:presidenzasmfn@uniroma1.it),  
[nesi@mat.uniroma1.it](mailto:nesi@mat.uniroma1.it)

\* Facoltà di Scienze Matematiche, Fisiche e Naturali, Sapienza Università di Roma [andrea.pelissetto@roma1.infn.it](mailto:andrea.pelissetto@roma1.infn.it)

**Keywords:** Valutazione, Valutazione didattica universitaria, Impatto sociale della didattica.

## 1. Obiettivi

In molti sistemi universitari, e scolastici in generale, vi è una diffusa e crescente domanda di miglioramento della qualità dell'offerta didattica. Questa esigenza è molto sentita da tutta la società. Riteniamo che sia doveroso tener conto di tale richiesta anche nelle valutazioni del sistema universitario. Negli ultimi anni l'Università italiana si è concentrata molto sulla valutazione della qualità della ricerca. Il Ministero dell'Istruzione, Università e Ricerca (MIUR), poi l'Agenzia Nazionale di Valutazione del Sistema Universitario e della Ricerca (ANVUR)<sup>1</sup> e infine, come conseguenza, tutta la comunità universitaria hanno dedicato una quantità di energie molto grande a questo tema. La Valutazione della Qualità della Ricerca (VQR) è stata oggetto di grandi dibattiti fra sostenitori e detrattori. Tuttavia bisogna riconoscere che il mondo accademico è arrivato largamente impreparato alla richiesta legittima, molto naturale, che proviene dalla società di rendere conto del modo come vengono utilizzate le risorse pubbliche. Con il tempo la comunità scientifica ha reagito mettendo in luce una serie di debolezze dell'approccio ministeriale ed in generale di un approccio legato soltanto a valutazioni numeriche. Bisogna anche riconoscere che una parte della comunità accademica, inizialmente, ha rifiutato il concetto stesso di valutazione della qualità della ricerca scientifica. A nostro parere tale atteggiamento è antistorico. La richiesta della società di una maggiore capacità di influenzare i processi che attengono alle distribuzioni delle risorse, anche in ambito scientifico, è stata messa in luce da molti studiosi ed osservatori (si veda ad esempio Greco e Silvestrini, 2010; 2009).

L'obiettivo del presente lavoro è quello di contribuire ad elaborare un metodo utile a rendere conto delle attività di una grande università pubblica in un campo alquanto inesplorato quale quello della valutazione della didattica. In un qualche modo si vorrebbe stimolare la comunità accademica ad elaborare un metro valutativo condiviso che possa fungere da antidoto ed argine a proposte poco meditate da parte degli organi di governo del Paese che, sulla formazione delle nuove classi dirigenti, oggettivamente hanno operato negli anni con grande frammentarietà, in un quadro di eccesso di produzione di leggi, norme e regolamenti.

In analogia con quanto accade con la valutazione della ricerca, noi siamo fermi sostenitori del fatto che valutazioni numeriche possano essere utili, quando i numeri sono abbastanza grandi per valutare comparativamente grandi popolazioni di studiosi. Ma possono essere anche utili per valutare individualmente quella che, volendo scherzosamente creare un neologismo, potremmo definire la "pessimenza". Valutare l'eccellenza è una attività che non può prescindere da una valutazione di merito molto attenta ed individuale. Viceversa individuare la coda più problematica è più facile, presenta maggiori caratteri di oggettività ed è anche più utile.

Noi riteniamo auspicabile che, nell'ambito della valutazione della didattica, la comunità scientifica preceda tutti gli apparati di ispirazione ministeriale proponendo un modello basato sull'esperienza di chi lavora all'Università, che abbia come obiettivo il miglioramento dei servizi agli studenti e alle

---

<sup>1</sup> <http://www.anvur.org/index.php?lang=it>

studentesse e che, in definitiva, si ponga come strumento per valorizzare il ruolo dell'Università pubblica come istituzione preposta alla crescita culturale del Paese.

Ricordiamo che sempre più frequentemente si pensa di legare, direttamente o indirettamente, l'erogazione di maggiori risorse, o, per essere più obiettivi, una minore diminuzione dell'erogazione delle risorse, alla valutazione delle attività di insegnamento.

In questo lavoro presentiamo l'analisi di uno studio, durato due anni, teso a ricercare un metodo, con forti elementi di oggettività, per selezionare insegnamenti che risultino essere responsabili principali di risultati insoddisfacenti della valutazione, con il fine di migliorare nel tempo la qualità della didattica.

Per la natura dell'istituzione che ha condotto lo studio (una Facoltà universitaria), in questi due anni ci si è concentrati su moltissimi aspetti legati all'offerta formativa, che partono dal rafforzamento dell'orientamento, alla verifica delle scelte degli studenti, ai tanti gradini della scala che conduce ad una buona offerta formativa. Tuttavia, in questa nota, ci concentriamo su un solo aspetto che riteniamo di grande importanza in tutto il percorso che porta al miglioramento della qualità della didattica: la valutazione che gli studenti e le studentesse fanno dei singoli insegnamenti a loro proposti, attraverso le cosiddette schede OPIS (OPInioni degli Studenti).

La rilevazione delle opinioni degli studenti su tutti gli insegnamenti dei Corsi di Studio o Corsi di Laurea (CdS, nel seguito per brevità) universitari è prevista su base nazionale. Il questionario proposto ha un certo numero di domande con quattro gradazioni di risposta (dal gradimento pieno al gradimento nullo), con domande stabilite a livello nazionale dal MIUR e comuni per tutti gli Atenei. Pertanto, si presta molto bene ad essere utilizzato per un confronto fra realtà omogenee (CdS presenti in tutta Italia, quali ad esempio il CdS in Chimica o in Scienze Biologiche, ecc.).

## 2. Aspetti metodologici e ambito di intervento

Il nostro approccio è radicalmente diverso da quelli in uso all'Università di Roma "La Sapienza" (Sapienza, nel seguito) ed anche nella nostra Facoltà di Scienze Matematiche, Fisiche e Naturali (Scienze MFN, nel seguito). Il presente studio vuole evidenziare i meriti di un nuovo approccio che, per amore di sintesi, chiameremo "ricerca delle code"; più precisamente, la ricerca della coda *meno virtuosa*. Esso differisce sostanzialmente da quelli in uso nei vari organismi valutativi che si concentrano sul valutare valori medi. Ad esempio, se vi fossero 10 insegnamenti, 9 dei quali eccellenti (dieci decimi di valutazione) ed uno pessimo (0 decimi di valutazione), sarebbe del tutto evidente che la media non potrà mai essere migliorata se non attraverso il miglioramento dell'unico insegnamento che viene valutato negativamente dall'utenza. Ogni tentativo di migliorare ulteriormente gli altri 9 insegnamenti è destinato a fallire e pertanto a rimanere un esercizio del tutto inutile. Per quanto possa sembrare paradossale, tale metodo, fatte le opportune proporzioni, è un esempio abbastanza calzante per spiegare la situazione generale.

Proponiamo inoltre di adottare un approccio pragmatico. La valutazione ci viene richiesta dal Ministero e noi, con l'analisi quantitativa delle schede OPIS, rispondiamo cercando tendenzialmente di far migliorare il giudizio che gli studenti e le studentesse danno degli insegnamenti.

Ci sono molti argomenti, quasi sempre pretestuosi, per non considerare rilevanti le valutazioni degli studenti. Quasi sempre questi si condensano in due obiezioni principali.

(i) La prima obiezione è che le schede OPIS sono inaffidabili se trattate numericamente. Dimosteremo che, al contrario, qualsiasi sia il metodo utilizzato per la valutazione delle schede, viene essenzialmente sempre individuata la stessa coda per quanto riguarda gli insegnamenti più problematici. Essa consiste, almeno per la realtà da noi analizzata in questo lavoro, in pochi insegnamenti con evidenti aspetti critici, qualunque siano i dettagli con i quali si valutano le schede.

(ii) La seconda obiezione è che un peso eccessivo dato alle schede OPIS potrebbe instaurare una spirale “buonista”, favorendo docenti troppo generosi nelle valutazioni degli studenti. Anche questo argomento, che pure ha ragioni per essere formulato, risulta indebolito dalla stabilità dei risultati in funzione del trattamento numerico (descritto più avanti).

D'altronde non si può nemmeno congetturare che il docente più bravo sia quello che opera una selezione maggiore. Il livello di selezione di un CdS deve dipendere da una decisione collegiale e può essere (e forse dovrebbe essere) programmato. Non può invece dipendere dal fissare in modo autoreferenziale una soglia diversa da quella stabilita globalmente per il numero di laureati che si ritiene debbano formare l'ossatura culturale e produttiva del Paese.

Su questa seconda obiezione torneremo nella sezione dedicata alle conclusioni con un risultato di notevole interesse e che sembra smentire l'idea che lo scarso gradimento degli insegnamenti sia altamente correlato allo scarso interesse per gli stessi. Questo argomento sarà ripreso nella Sezione 7 nei commenti alla Tabella 5.

L'analisi quantitativa delle opinioni degli studenti quindi si concentra sui seguenti punti:

- 1) definire per ogni insegnamento un punteggio da zero ad uno (più avanti specificheremo i tanti modi per farlo);
- 2) definire l'ambito culturalmente considerato adeguato per confronti omogenei (docenti dello stesso CdS);
- 3) verificare che su domande quali: “Quanto è soddisfatto/a complessivamente dell'insegnamento?”, la *media* della valutazioni su tutti gli insegnamenti dello stesso CdS fornisce pochissima informazione;
- 4) verificare che, al contrario, all'interno di un CdS, la percentuale di insegnamenti considerati “critici” secondo un semplice parametro, è pressoché costante e fornisce un numero esiguo di insegnamenti i quali ricevono valutazioni talmente al di sotto delle *media* da non essere in alcun modo compensabili dal miglioramento degli insegnamenti per i quali il gradimento è già alto.

La conclusione di questa analisi è che un sensibile miglioramento della valutazione *media* di un CdS è realmente possibile soltanto impegnandosi a riportare *la coda meno virtuosa della distribuzione* verso la *media* del CdS. Quindi un miglioramento significativo è solo possibile se si affrontano con coraggio, equilibrio e competenza le ragioni per le quali quei pochi insegnamenti si discostano in maniera così radicale, in peggio, dalla *media*.

Pertanto la strategia generale consiste nell'introdurre un metodo, totalmente esplicito e pubblico, con il quale segnalare alle comunità che hanno la responsabilità di prendere decisioni in merito (quali i Consigli di CdS, i Consigli di Dipartimento, le Scuole o le Facoltà), la lista degli insegnamenti che le schede OPIS segnalano come critici. Sarà poi cura degli organi preposti individuare le azioni necessarie per correggere le sofferenze individuate.

Le ragioni di scarso gradimento possono essere molteplici (aule troppo piccole o inadeguate, malattia del docente, insegnamenti troppo anticipati rispetto alle conoscenze di base, ed altro ancora). Quindi è cruciale che le decisioni siano prese da un organismo *collegiale*. Spesso il singolo docente non ha gli strumenti per migliorare il proprio insegnamento (si pensi alle aule, di competenza dei Dipartimenti, o anche al posizionamento dell'insegnamento al primo semestre

invece che al terzo, da ripensare nell'ambito del CdS, oppure ad un programma troppo denso per le richieste che provengono dagli insegnamenti successivi).

Crediamo che fornire un metodo affidabile, stabile e replicabile per quantificare il gradimento dell'offerta didattica dei singoli insegnamenti da parte degli studenti, sia un passo in avanti nel miglioramento dell'offerta didattica in ogni CdS, in ogni Facoltà o Scuola ed ogni Ateneo, che rappresenta uno degli obiettivi principali, insieme alla valutazione della qualità della ricerca, di una corretta posizione dell'Università nei confronti della società per la formazione delle nuove generazioni.

Nel seguito discuteremo i risultati di un'analisi condotta sui dati raccolti per tutti i CdS della Facoltà di Scienze MFN della Sapienza. Per motivi di sintesi e di efficacia, nel presente lavoro ci limitiamo a mostrare i dati relativi ai CdS triennali dell'A.A. 2013-2014, sulla base dei questionari OPIS compilati dagli studenti in forma elettronica durante l'anno.

### 3. Popolazione di riferimento

Ai fini dell'analisi considereremo i questionari OPIS relativi all'A.A. 2013-2014 per i CdS triennali della Facoltà di Scienze MFN della Sapienza. Di seguito, la sigla con cui verranno identificati:

- (1) CdS\_01
- (2) CdS\_02
- (3) CdS\_03
- (4) CdS\_04
- (5) CdS\_05
- (6) CdS\_06
- (7) CdS\_07
- (8) CdS\_08
- (9) CdS\_09
- (10) CdS\_10

Utilizzeremo solo i questionari degli studenti *frequentanti*, rilevati dall'Ateneo per via telematica (OPIS On-Line<sup>2</sup>) collegata con il sistema gestionale delle carriere INFOSTUD<sup>3</sup>, per un totale di 14156 questionari. Tale numero appare significativo in assoluto in quanto il numero di iscritti totale è di poco inferiore a 8000 e, anche se è difficile calcolare il numero di coloro che effettivamente frequentano i corsi di studio, è ragionevole pensare che siano intorno ai cinque o seimila. In questo senso la popolazione presa in considerazione appare molto significativa.

---

2 [http://w3.uniroma1.it/cersites/files/\\_Linee-Guida-Studenti-cv.pdf](http://w3.uniroma1.it/cersites/files/_Linee-Guida-Studenti-cv.pdf)

3 <https://stud.infostud.uniroma1.it/Sest/Log/Corpo.html>

	N insegnamenti	N insegnamenti > 5	N schede	N schede > 5
CdS_01	19	17	272	268
CdS_02	19	19	527	527
CdS_03	33	33	2281	2281
CdS_04	18	18	697	697
CdS_05	64	62	3132	3130
CdS_06	49	45	2031	2020
CdS_07	20	19	626	624
CdS_08	64	55	2077	2058
CdS_09	24	23	1127	1122
CdS_10	31	31	1386	1386

**Tabella 1.** Per ciascun CdS riportiamo il numero di insegnamenti (N insegnamenti), il numero di insegnamenti (N insegnamenti > 5) per cui il numero di schede è maggiore di 5 (ossia degli insegnamenti che verranno considerati nell'analisi), il numero di schede totali (N schede), ed il numero di schede totali (N schede > 5) relative agli insegnamenti analizzati.

Inoltre il numero è significativo in quanto rappresenta circa il 75% del totale dei questionari che vengono compilati e, giova ricordare, la compilazione è obbligatoria per chiunque voglia sostenere un esame di profitto.

Inoltre, sono stati considerati solo gli insegnamenti per i quali sono stati raccolte almeno 6 schede. In Tabella 1 per ciascun CdS riportiamo il numero di insegnamenti (N insegnamenti), il numero di insegnamenti (N insegnamenti > 5) per cui il numero di schede è maggiore di 5 (ossia degli insegnamenti che verranno considerati), il numero di schede totali (N schede), ed il numero di schede totali (N schede > 5) relative agli insegnamenti analizzati.

Come si vede dalla tabella, gli insegnamenti con al più 5 schede sono una piccola minoranza, assolutamente inessenziale per l'analisi dei vari CdS triennali.

#### 4. Metodologia

L'analisi della distribuzione delle risposte è stata condotta per identificare quei (pochi) insegnamenti i quali hanno indici di gradimento molto basso, proprio perché su quei pochi insegnamenti occorre agire per portarli verso i valori medi della Facoltà. Vedremo che tutti i metodi utilizzati individuano quasi sempre gli stessi insegnamenti.

I questionari OPIS somministrati nell'A.A. 2013-2014 contengono 12 domande che permettono di determinare il gradimento da parte degli studenti di vari aspetti, di natura organizzativa, didattica e culturale, relativa ai vari CdS. In questo documento ci focalizzeremo, a titolo di esempio, su tre domande del questionario:

**Domanda 5.** *Gli orari di svolgimento di lezioni, esercitazioni e altre eventuali attività didattiche sono rispettati?*

**Domanda 11.** *È interessato/a agli argomenti trattati nell'insegnamento?*

**Domanda 12.** *Sono complessivamente soddisfatto/a di come è stato svolto questo insegnamento?*

La domanda 5 è sostanzialmente una domanda di carattere deontologico, la quale permette di verificare se il docente rispetti i propri doveri professionali, comportandosi in modo corretto nei confronti degli studenti. Da un punto di vista metodologico, la domanda 5 è particolarmente interessante dato che la risposta è obiettiva: tutti i questionari dovrebbero riportare la stessa risposta. Differenze eventuali danno un'indicazione della affidabilità delle singole risposte ottenute dai questionari.

La domanda 11 è utile per comprendere l'interesse degli studenti verso gli argomenti trattati, mentre la domanda 12 permette di rilevare il gradimento complessivo dell'offerta didattica, nei suoi aspetti sia organizzativi, sia culturali. È inoltre particolarmente importante confrontare i risultati delle due domande, in quanto tale confronto permette di distinguere la soddisfazione sugli aspetti organizzativi e sulla qualità della didattica, *due cose sulle quali si può e si deve agire*, dall'interesse per la materia oggetto del corso.

Al fine di effettuare un'analisi quantitativa è necessario associare a ciascuna domanda ed a ciascun insegnamento un valore numerico  $r$ , che chiameremo *grado di soddisfazione*, che ne caratterizzi il gradimento. La definizione di tale indicatore non è univoca. Nel seguito presenteremo i risultati ottenuti con tre procedure differenti, facendo vedere che, al fine di discriminare gli *insegnamenti problematici*, tali procedure sono tutte tra loro sostanzialmente equivalenti.

Prima di definire i vari metodi, è necessario definire le quantità più rilevanti. In sostanza, per ogni domanda, esistono quattro gradazioni di risposta, da quella molto positiva, a quella molto negativa. Consideriamo le risposte ad una domanda specifica per un insegnamento fissato. Ovviamente quel che conta sono le percentuali di queste quattro categorie di risposte. Per una facile associazione mentale useremo le seguenti notazioni:

$$(4.1) \quad \begin{aligned} f_{++} & \text{ è la frazione di studenti e studentesse decisamente soddisfatti/e,} \\ f_{+} & \text{ è la frazione di studenti e studentesse mediamente soddisfatti/e,} \\ f & \text{ è la frazione di studenti e studentesse mediamente insoddisfatti/e,} \\ f_{-} & \text{ è la frazione di studenti e studentesse decisamente insoddisfatti/e.} \end{aligned}$$

In questo modo si ha, ovviamente, che la somma dei quattro numeri vale uno ed ognuno dei quattro numeri è compreso fra zero e uno.

Dal punto di vista matematico, per definire le frazioni che compaiono nelle formule (4.1) dobbiamo definire i seguenti numeri interi:

$$(4.2) \quad \begin{aligned} N_{++} & \text{ è il numero delle risposte "decisamente sì",} \\ N_{+} & \text{ è il numero delle risposte "più sì che no",} \\ N & \text{ è il numero delle risposte "più no che sì",} \\ N_{-} & \text{ è il numero delle risposte "decisamente no".} \end{aligned}$$

Il numero totale di schede, denotato con  $N_{tot}$ , è dato dalla somma dei quattro diversi contributi:

$$N_{tot} = N_{++} + N_{+} + N + N_{-}.$$

Pertanto la formulazione matematica delle quantità definite in (4.1) è

$$f_{++} = \frac{N_{++}}{N_{tot}}, \quad f_{+} = \frac{N_{+}}{N_{tot}}, \quad f_{-} = \frac{N_{-}}{N_{tot}}, \quad f_{--} = \frac{N_{--}}{N_{tot}}.$$

Siamo a questo punto in grado di definire tre metodi per il calcolo del grado di soddisfazione  $r$ .

**Metodo 1 (M1).**

Il metodo più semplice e diretto per definire il grado di soddisfazione consiste nel porre

$$(4.3) \quad r = \frac{N_{++} + N_{+}}{N_{tot}} = f_{++} + f_{+}.$$

In questo caso il grado di soddisfazione è definito come la percentuale di studenti che ha risposto positivamente alla domanda, ossia la percentuale dei “decisamente sì” e dei “più sì che no”. Questo metodo è quello usualmente utilizzato a livello di Ateneo.

**Metodo 2 (M2).**

Il metodo 1 non distingue tra i decisamente soddisfatti e coloro che sono solo abbastanza soddisfatti. Analogamente, i decisamente insoddisfatti sono considerati equivalenti a coloro che sono solo parzialmente insoddisfatti. Per tenere conto di tali differenze si può dare un peso diverso alle varie risposte, definendo

$$(4.4) \quad r = \frac{a^{++}N_{++} + a^{+}N_{+} + a^{-}N_{-} + a^{--}N_{--}}{N_{tot}} = a^{++}f_{++} + a^{+}f_{+} + a^{-}f_{-} + a^{--}f_{--},$$

dove

$$1 \geq a^{++} \geq a^{+} \geq a^{-} \geq a^{--} \geq 0$$

sono delle costanti che ora definiremo. Il metodo 1 corrisponde a prendere  $a^{++}=a^{+}=1$ ,  $a^{-}=a^{--}=0$ . Per tenere conto del diverso grado di soddisfazione utilizzeremo

$$(4.5) \quad a^{++} = 1, \quad a^{+} = 0.8, \quad a^{-} = 0.2, \quad a^{--} = 0.$$

Si noti la scelta simmetrica per i valori, ossia  $a^{--}=1-a^{++}$  e  $a^{-}=1-a^{+}$ , che fa sì che possiamo considerare il valore  $r=1/2$  come il discriminante tra un corso globalmente non gradito ed un corso di cui invece gli studenti sono sostanzialmente soddisfatti. La scelta fatta contiene un certo grado di arbitrarietà. Un'altra possibilità consiste nel prendere pesi equispaziati, ossia  $a^{++} = 1$ ,  $a^{+} = 2/3$ ,  $a^{-} = 1/3$ ,  $a^{--} = 0$ . Come discuteremo nel seguito, la scelta dei pesi non influenza significativamente il risultato finale.

**Osservazione.**

La critica più frequente fatta agli indicatori (4.3) e (4.4) è quella di essere sensibili (e tale sensibilità aumenta al diminuire del numero di schede) a risposte anomale, date da studenti poco attenti o poco

interessati alla rilevazione. Per esempio, la risposta alla domanda 5 dovrebbe essere univoca: tutti gli studenti dovrebbero dare la stessa risposta, dato che la puntualità del docente è un dato oggettivo. Invece, si rileva spesso che qualche studente dia una risposta molto diversa da quella data dalla maggioranza (per esempio, capita che la maggioranza risponda “decisamente sì”, e qualche raro studente risponda “più no che sì”); tali risposte, chiaramente non corrette, potrebbero falsare i risultati nel caso di poche schede. Inoltre la loro presenza viene spesso utilizzata per giustificare l'affermazione che i risultati ottenuti dall'analisi dei questionari OPIS sarebbero poco attendibili. Al fine di tener conto di tali obiezioni, introduciamo ora un terzo metodo di analisi.

### Metodo 3 (M3).

Il terzo metodo, che chiameremo metodo *cenurato* (*trimming*), elimina dall'analisi le *code*, ossia quelle risposte che sono molto dissimili dalla risposta media. M3 è un metodo molto noto ed usato in letteratura<sup>4</sup>. Il metodo M3 con la scelta di eliminare il 25% superiore ed il 25% inferiore fornisce la cosiddetta media interquartile (in inglese talvolta indicata con *midmean*). Una discussione più precisa da un punto di vista statistico può essere trovata in (Tukey, 1977). Medie *winsorized* sono anche discusse in (Tukey, 1962; Hasings et al., 1947). Curiosamente viene adoperato in applicazioni molto diverse, dalla determinazione degli indici monetari in campo economico e finanziario, alla determinazione del punteggio di prove quali tuffi o ginnastica nelle competizioni olimpiche<sup>5</sup>.

### Presentazione qualitativa del metodo.

Prima di introdurre le definizioni formali illustriamo la sostanza della procedura con un esempio. L'idea è molto semplice e, di fatto, tende a produrre un *campione modificato* rispetto a quello originale tentando di eliminare tutte quelle risposte ai questionari che possono apparire, su base oggettiva, poco affidabili. Supponiamo ad esempio di voler misurare la lunghezza di un foglio di carta, formato A4, con “grande” precisione, ad esempio con la precisione di un millimetro. Teoricamente la lunghezza di tale foglio deve misurare 297 millimetri. Ovviamente il singolo foglio potrebbe deviare da tale misura. Si procede a fare 10 misurazioni. Esse si concentreranno su valori molto vicini a 297 mm. Misure del tipo 397 mm oppure 197 mm possono essere presenti solo a causa di *grossolani* errori. Vanno “censurate”. La misura più accreditata, quindi, verrà ricercata facendo una media dei valori ristretta a quelli considerati affidabili.

Per calare questa idea nel contesto della valutazione delle schede OPIS e apprezzare l'utilità del metodo censurato, faremo un esempio costruito ad arte, considerando due insegnamenti, entrambi con 10 schede di valutazione. Supponiamo che, alla domanda 5, le risposte relative al primo insegnamento, che chiameremo A, siano le seguenti (si veda la Tabella 2): “decisamente sì” (8 volte), “più no che sì” (una volta) e “decisamente no” (una volta). Le risposte relative al secondo insegnamento, che chiameremo B, siano: “decisamente sì” (2 volte) e “più sì che no” (8 volte). È evidente che il docente dell'insegnamento A è più puntuale di quello del secondo. Infatti, data

---

<sup>4</sup> Le definizioni matematiche dei metodi utilizzati possono essere trovate in [http://en.wikipedia.org/wiki/Trimmed\\_estimator](http://en.wikipedia.org/wiki/Trimmed_estimator)

<sup>5</sup> Per una discussione informale, si veda l'articolo *Removing Judges' Bias Is Olympic-Size Challenge* di Carl Bialik, pubblicato sul *Wall Street Journal* del 27 luglio 2012, <http://www.wsj.com/news/articles/SB10000872396390443477104577551253521597214>; le regole per la definizione dei tassi monetari sono spiegate, per esempio, in [https://www.theice.com/publicdocs/services/ISDAFIX\\_EUR.pdf](https://www.theice.com/publicdocs/services/ISDAFIX_EUR.pdf)

l'oggettività della condizione “essere puntuali”, le due risposte anomale sia per l'insegnamento A che per l'insegnamento B sono poco credibili. Di fatto, sono erranee. Vediamo ora, in questo esempio, cosa accade al grado di soddisfazione ottenuto nei vari metodi. Maggiori dettagli vengono forniti alla fine della presente sezione.

Risposta		Insegnamento A	Insegnamento B
decisamente sì	++	8	2
più sì che no	+	0	8
più no che sì	–	1	0
decisamente no	--	1	0

**Tabella 2.** Risposte ipotetiche alla domanda 5 per i due insegnamenti A e B.

Il metodo 1 dà  $r=0.8$  per l'insegnamento A e  $r=1.0$  per l'insegnamento B, di fatto capovolgendo l'evidenza. La ragione di questo apparente paradosso è che, adottando il Metodo 1, le risposte anomale pesano troppo significativamente e danno un *ranking* opposto a quello che dovrebbe essere corretto. Il metodo 2 fornisce  $r=0.82$  per l'insegnamento A e  $r=0.84$  per l'insegnamento B. Ora i due insegnamenti ottengono, sostanzialmente la stessa valutazione ed in entrambi i casi i docenti vengono valutati come “abbastanza puntuali”. Qui la considerazione è che, mentre la valutazione ottenuta con il metodo 2 è sostanzialmente corretta per l'insegnamento B, essa continua ingiustamente a penalizzare il docente dell'insegnamento A.

Il metodo 3, basato sulla distribuzione censurata, fornisce, come vedremo,  $r=1.00$  per l'insegnamento A e  $r=0.8$  per l'insegnamento B. Le risposte inattendibili sono state scartate ed il primo insegnamento ha ottenuto il risultato migliore. In poche parole, il metodo 3 è quello più equo ed affidabile.

### Presentazione quantitativa ma informale del metodo 3.

Per descrivere quantitativamente il metodo 3 è necessario innanzitutto stabilire una soglia di *inaffidabilità*; in altre parole, quale frazione di dati si vuole scartare. Anticipiamo che, proprio per garantire una grandissima stabilità dei risultati e preservarli dalla possibile obiezione di essere soggetti a grandi fluttuazioni, vorremo fissare questa soglia in modo che sia “grande”. Nelle applicazioni, scarteremo il 50% dei dati! Più precisamente il 25% più favorevole e il 25% più sfavorevole. I dati rimasti sono quelli che *riteniamo decisamente affidabili* e solo ad essi viene applicata la strategia valutativa complessiva. Si noti che la scelta di scartare il 25% delle code è una scelta arbitraria. Si sarebbe potuto scartare una frazione diversa (che chiameremo  $q$  nella trattazione matematica) individuata soggettivamente come la soglia *al di sotto della quale è lecito dubitare dell'affidabilità delle risposte*. La nostra scelta  $q=1/4$ , che corrisponde a scartare la metà delle risposte fornite, ha lo scopo di utilizzare distribuzioni molto diverse rispetto a quelle considerate nei metodi precedenti. Quindi un'eventuale coincidenza dei risultati delle tre analisi rappresenterebbe una conferma della validità delle conclusioni raggiunte.

Una volta effettuata l'operazione di “pulizia” dei dati, si torna ad utilizzare la media pesata dei valori esibiti dal campione ripulito. Nel nostro caso quindi l'indicatore che fornisce la valutazione dell'insegnamento è

$$r = a^{++} p_{++} + a^+ p_+ + a^- p_- + a^{--} p_{--},$$

dove i valori  $p_{++}$ ,  $p_+$ ,  $p_-$ ,  $p_{--}$  con  $p_{++}+p_++p_-+p_{--}=1$ , rappresentano le frazioni di schede, valutate nel nuovo campione ripulito, corrispondenti alle quattro possibili valutazioni. Per semplicità, ci limiteremo a considerare il caso

$$a^{++} = 1, a^+ = 0.8, a^- = 0.2, a^{--} = 0.$$

Per i dettagli matematici lettori e lettrici possono consultare l'appendice A.

Possiamo, a questo punto, comprendere come sia stato ottenuto il grado di soddisfazione per i due insegnamenti ipotetici A e B, discussi nella sezione precedente (vedi Tabella2). In entrambi i casi le code corrispondono a 2 insegnamenti<sup>6</sup>. Per l'insegnamento A, nel campione ripulito vengono tolte le due risposte anomale e due risposte “decisamente sì”. Nel campione ripulito rimangono quindi solo risposte “decisamente sì”, per cui  $r=1$ . Nel caso dell'insegnamento B vengono eliminate le due risposte “decisamente sì” assieme a due risposte “più sì che no”. Nel campione ripulito rimangono quindi solo risposte “più sì che no”, per cui  $r=0.8$ .

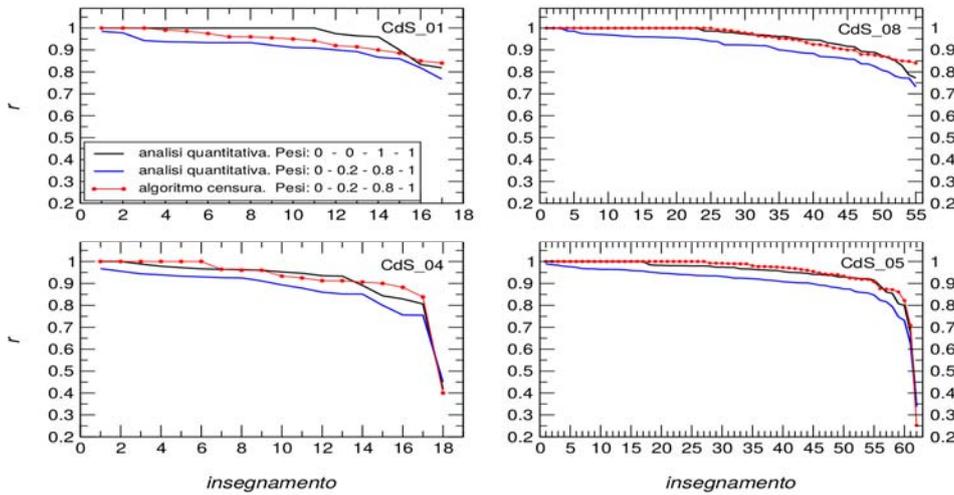
## 5. Confronto dei vari metodi

Al fine di confrontare i tre metodi introdotti e di verificarne l'affidabilità, considereremo 4 diversi CdS, due ad alta numerosità, CdS\_08 e CdS\_05, e due con un numero di studenti molto inferiore, CdS\_01 e CdS\_04. Confronteremo i risultati per le domande 5 e 12 che sono metodologicamente molto diverse. Per la prima ci si aspetta una distribuzione delle risposte molto stretta, essendo la risposta oggettiva.

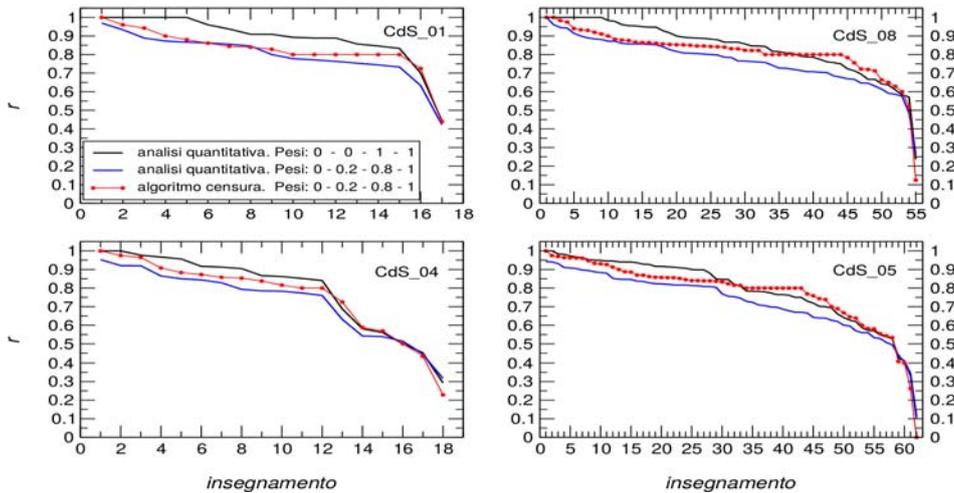
---

<sup>6</sup> Propriamente, il 25% di 10 è pari a 2.5. Ignoreremo questa differenza, inessenziale in questo caso. La trattazione esatta può essere fatta seguendo la discussione riportata nell'Appendice A.

OPIS 2013-14. Domanda 5. Gli orari di svolgimento di lezioni, esercitazioni e altre eventuali attività didattiche sono rispettati?



OPIS 2013-14. Domanda 12. Sono complessivamente soddisfatto di come è stato svolto questo insegnamento?



**Figura 1.** Gradi di soddisfazione ottenuti con i tre diversi metodi come funzione del *ranking* (metodo M1, nero; metodo M2, blu; metodo M3, rosso).

La risposta alla seconda domanda invece è molto più soggettiva e quindi ci si aspetta una distribuzione di risposte più ampia.

In Figura 1 riportiamo la distribuzione delle risposte per le due domande relativamente ai 4 CdS considerati. In ciascun grafico vengono riportati i risultati ottenuti con i tre metodi, come funzione del *ranking*. In pratica, per ogni insegnamento con  $N_{tot} > 5$  viene calcolato il grado di soddisfazione  $r$ . Tali valori vengono quindi ordinati in modo decrescente e riportati in figura. Come ci si attende, il grado di soddisfazione ottenuto con M1 è sempre superiore a quello ottenuto con M2. Le differenze non sono grandi (inferiori a 0.1 tipicamente, ovvero inferiori al 10%) e rispecchiano il fatto che, per molti insegnamenti, i “decisamente soddisfatti” ed i “parzialmente soddisfatti” hanno percentuali

simili. Un confronto più quantitativo può essere fatto utilizzando i dati riportati nelle Tabelle 3 e 4. Anche in questo si vede che sia le mediane, sia la posizione del quartile inferiore si spostano verso il basso quando si passa da M1 ad M2.

	Mediana			quartile inferiore			N insegnamenti con $r < 0,5$				N insegnamenti
	M1	M2	M3	M1	M2	M3	M1	M2	M3	M123	
CdS_01	1.00	0.92	0.96	0.96	0.89	0.91	0	0	0	0	17
CdS_04	0.96	0.90	0.95	0.89	0.85	0.91	1	1	1	1	18
CdS_05	0.97	0.93	0.99	0.94	0.89	0.94	1	1	1	1	62
CdS_08	0.98	0.92	0.99	0.93	0.87	0.92	0	0	0	0	55

**Tabella 3.** Domanda 5. Per ciascun CdS riportiamo la mediana della distribuzione dei valori di  $r$ , la posizione del quartile inferiore, ossia il valore  $r_q$  tale che la percentuale di insegnamenti con  $r < r_q$  è pari al 25%, il numero di insegnamenti con grado di soddisfazione inferiore a 0.5 ed infine il numero totale di insegnamenti considerati. M1, M2 e M3 si riferiscono al metodo utilizzato. La colonna M123 dà il numero di insegnamenti che hanno  $r < 0.5$  secondo tutti e tre i metodi.

La differenza nel valore dell'estimatore  $r$  tra i due metodi non è *a priori* l'unica cosa rilevante. Allo scopo di individuare gli insegnamenti sui quali è necessario agire, ciò che conta è il *ranking*, ossia il gradimento di un corso rispetto al gradimento medio di tutti gli insegnamenti di quel CdS. Ossia si vuole sapere se il grado di soddisfazione di un corso è, per esempio superiore/inferiore alla mediana, oppure se appartiene al primo quartile inferiore. A questo fine i due metodi risultano essere sostanzialmente equivalenti. Consideriamo, per esempio, la domanda 12 ed il CdS\_05. Se consideriamo i 15 insegnamenti che appartengono al quartile superiore secondo il metodo M1, 12 di essi appartengono allo stesso quartile quando viene utilizzato il metodo M2. Se consideriamo il quartile inferiore, i metodi M1 ed M2 identificano gli stessi 15 insegnamenti. Se consideriamo invece il CdS\_01, dei 4 insegnamenti che il metodo M1 identifica come appartenenti al quartile inferiore, 3 appartengono allo stesso quartile anche secondo il metodo M2. Più interessante è il confronto con i risultati del metodo M3, che ci permette di valutare l'effetto delle *code*. In quasi tutti i casi si vede che  $r_{M2} < r_{M3} < r_{M1}$ . Questo dato indica che l'effetto delle *code* è veramente marginale. Anche per il *ranking*, M3 risulta sostanzialmente equivalente a M1 ed M2. Per esempio, consideriamo di nuovo la domanda 12 ed il CdS\_05.

	Mediana			quartile inferiore			N insegnamenti con $r < 0.5$				N insegnamenti
	M1	M2	M3	M1	M2	M3	M1	M2	M3	M123	
CdS_01	0.91	0.80	0.83	0.86	0.75	0.80	1	1	1	1	17
CdS_04	0.86	0.78	0.83	0.58	0.54	0.59	2	2	2	2	18
CdS_05	0.83	0.75	0.82	0.70	0.64	0.74	4	5	4	4	62
CdS_08	0.87	0.78	0.82	0.76	0.70	0.80	1	1	1	1	55

**Tabella 4.** Domanda 12. Per ciascun CdS riportiamo la mediana della distribuzione dei valori di  $r$ , la posizione del quartile inferiore, ossia il valore  $r_q$  tale che la percentuale di insegnamenti con  $r < r_q$  è pari al 25%, il numero di insegnamenti con grado di soddisfazione inferiore a 0.5 ed infine il numero totale di insegnamenti considerati. M1, M2 e M3 si riferiscono al metodo utilizzato. La colonna M123 dà il numero di insegnamenti che hanno  $r < 0.5$  secondo tutti e tre i metodi.

Se consideriamo i 15 insegnamenti che appartengono al quartile inferiore secondo il metodo M2, 14 di essi appartengono allo stesso quartile secondo il metodo M3. Siamo quindi abbastanza certi dell'affidabilità dei risultati.

Come detto nell'introduzione, uno degli scopi dell'analisi è quello di individuare i pochi insegnamenti che godono di uno scarso gradimento. La scelta più naturale è porre il discrimine a  $r=1/2$  e quindi investigare per quali insegnamenti si abbia  $r < 1/2$ . Si noti che la posizione del quartile inferiore corrisponde a valori di  $r$  significativamente più alti (unica eccezione è il CdS\_04, ma in questo caso il numero di insegnamenti è piccolo ed un quartile corrisponde a 4 insegnamenti). Quindi gli insegnamenti che soddisfano tale disequaglianza sono realmente anomali rispetto alla media. Abbiamo quindi identificato tali insegnamenti per ciascun CdS e ciascuna domanda. Come si vede dalle Tabelle 3 e 4 i tre metodi forniscono quasi sempre lo stesso risultato: il non gradimento di un corso è un dato oggettivo, indipendente dal metodo di analisi utilizzato.

Alla luce di questa discussione, nel seguito riporteremo le distribuzioni ottenute con il metodo censurato per tutti i CdS e per le due domande ed individueremo gli insegnamenti problematici come quelli per cui  $r < 1/2$  utilizzando *tutti i tre metodi*.

## 6. Risultati

Nell'Appendice B abbiamo riportato i gradi di soddisfazione per le domande 5 e 12 dei diversi insegnamenti come funzione del *ranking* per tutti i CdS triennali della Facoltà di Scienze MFN alla Sapienza. Riportiamo solo i risultati ottenuti con il metodo M3; come già discusso, i risultati M1 e M2 sono analoghi. Per la domanda 5 il grado di soddisfazione è molto alto: per la stragrande maggioranza degli insegnamenti si ha  $r > 0.9$ , confermando la sostanziale puntualità dei docenti. L'identificazione degli insegnamenti "problematici" dai grafici è pressoché immediata: un rapido sguardo mostra che vi sono 4 corsi per cui  $r < 0.5$ , sui quali è certamente necessario agire. Per quel che riguarda la domanda 12, l'indicatore  $r$  mostra una maggiore variazione, ma anche qui le *code* possono essere facilmente discriminate. Se identifichiamo come insegnamenti problematici quelli per cui  $r < 0.5$  per tutti e tre i metodi, troviamo 21 insegnamenti, un numero piccolo rispetto al numero dei insegnamenti analizzati (322 insegnamenti). Questo risultato è chiaramente positivo e conferma che gli studenti sono estremamente selettivi: non vi sono lamentele generalizzate, ma lamentele ben specifiche relative solo a pochi insegnamenti, per risolvere le quali devono chiaramente agire gli organi preposti (Consigli di CdS, Dipartimento, Facoltà), individuandone le cause e trovando le soluzioni.

## 7. Discussione e conclusioni

1) L'obiettivo più interessante di una valutazione degli insegnamenti consiste nel migliorare i servizi a studentesse e studenti. Si deve quindi concentrare l'attenzione sul superiore interesse generale ed

accantonare, in quanto socialmente meno rilevanti, le considerazioni che riguardano i pochi individui che certamente non avranno piacere ad essere associati ad insegnamenti problematici.

2) In strutture di raccordo sufficientemente ampie, quali Scuole e Facoltà in ambito universitario, ma anche comprensori scolastici in ambito di scuola secondaria, sembra essenziale trovare un metodo semi automatico per fare una prima cernita degli insegnamenti considerati critici dal punto di vista dell'utenza. Senza attribuire valore sacrale a tali suggerimenti, bisogna affermare il dovere di rispondere con puntualità alle critiche analizzandole con serietà e con un deciso grado di *terzietà*.

**Osservazione.** *Spesso le critiche si riveleranno estremamente ben fondate. Tale affermazione è suffragata dall'esperienza: quasi sempre le valutazioni negative da parte di studentesse e studenti confermano impressioni molto diffuse nella comunità con la differenza, abissale, che non si tratta più, appunto di "impressioni", soggettive per definizione, ma di indicazioni e suggerimenti provenienti da un campione di utenza numeroso e ben selezionato. Viceversa, nell'ambito degli insegnamenti molto graditi si riscontrano, a volte, della gradite sorprese. Ma questo è un altro discorso in quanto gli insegnamenti più graditi, secondo l'esperienza messa in moto nella Facoltà di Scienze MFN della Sapienza, vanno selezionati con un metodo diverso e molto più raffinato.*

3) L'approccio presentato in questo lavoro fornisce indicazioni fortissime che esiste un metodo molto robusto per determinare una percentuale ristretta (sempre minore del 10% nei casi studiati), di insegnamenti che necessitano nettamente di maggior attenzione degli altri. Il nostro auspicio è che la presente proposta possa essere conosciuta, eventualmente raffinata, ed applicata in contesti culturali diversi. Ma siamo convinti che il cuore del metodo per la determinazione degli insegnamenti problematici possa essere esportato in molti ambiti, a cominciare da altre realtà particolarmente affini a quella studiata in questo lavoro.

4) La nostra analisi dimostra l'utilità di effettuare alcune modifiche nei questionari. Più precisamente sarebbe utilissimo introdurre nuove domande, o modificare leggermente alcune di quelle presenti, al fine di permettere una ripulitura ancora più efficiente del campione iniziale. Un buon esempio potrebbe essere la domanda 5. Si tratta di una valutazione oggettiva richiesta a chi è presente in aula. Sarebbe meglio formulare almeno una domanda dove il grado di oggettività sia altissimo.

Ad esempio:

“la percentuale delle assenze a lezione del/della docente è:”

(7.1)                    0, compresa fra 0 e 10%, compresa fra 10 e 20%, maggiore 20%

Questo permetterebbe di escludere più facilmente dal campione le risposte chiaramente non affidabili.

5) La Tabella 5 mostra un dato particolarmente interessante. Rozzamente parlando, dimostra che la teoria secondo cui gli insegnamenti poco graditi dagli studenti e dalle studentesse siano quasi sempre quelli “difficili” oppure quelli meno centrati sui supposti interessi principali delle studentesse e degli studenti (ad esempio l'apprendimento della Matematica in un CdS di Biologia oppure, viceversa, l'apprendimento di nozioni basilari di Biologia in un CdS di Matematica), non è verificata dai dati statistici a nostra disposizione.

$r$	N insegnamenti
$0.9 \leq r$	1
$0.8 \leq r < 0.9$	12
$0.7 \leq r < 0.8$	1
$0.6 \leq r < 0.7$	3
$0.5 \leq r < 0.6$	3
$r \leq 0.5$	1

**Tabella 5.** Distribuzione del grado di soddisfazione  $r$  per la domanda 11, che quantifica l'interesse per i contenuti degli insegnamenti, relativamente ai 21 insegnamenti problematici individuati nella Facoltà di Scienze MFN, lauree triennali.

Accade “spesso” che si risponda che “si è molto interessati” ai contenuti dell'insegnamento ma si dia un parere “molto negativo” del modo come l'insegnamento è impartito. Sembra di poter dire che ci sia correlazione più stretta fra il/la docente e il gradimento, piuttosto che fra il contenuto dell'insegnamento e la formazione degli studenti e delle studentesse, anche se per dimostrare questo fatto occorrerà uno studio più prolungato nel tempo per monitorare l'attività di docenti su corsi di studio molto diversi fra loro.

## 8. Appendice A: definizione del metodo 3

In questa appendice vogliamo dare una definizione matematicamente rigorosa del metodo 3. Per rendere la trattazione matematica più semplice, è utile modificare le notazioni introdotte nel testo, indicando con  $f_1$  la frazione  $f_{++}$ , con  $f_2$  la frazione  $f_+$ , con  $f_3$  la frazione  $f$  e con  $f_4$  la frazione  $f_-$ .

Sia ora  $q$  la soglia di affidabilità, associata al  $q$ -ile inferiore. Ovviamente deve valere  $0 \leq q < 1/2$ . Si può allora definire un unico intero, denotato con  $k_{min}$ , che per costruzione soddisfa  $1 \leq k_{min} \leq 4$ , associato al  $q$ -ile inferiore. Esso è definito implicitamente dalla relazione

$$\sum_{i < k_{min}} f_i < q \leq \sum_{i \leq k_{min}} f_i.$$

Analogamente si può definire un unico intero, denotato con  $k_{max}$ , che per costruzione soddisfa  $k_{min} \leq k_{max} \leq 4$ , associato al  $q$ -ile superiore. Esso è definito implicitamente dalla relazione

$$\sum_{i > k_{max}} f_i < q \leq \sum_{i \geq k_{max}} f_i.$$

Definiamo ora una nuova distribuzione (*censurata o trimmed*)  $p_i$ . In questo passaggio stiamo, di fatto, riducendo il campione di una frazione sufficiente a garantirne l'affidabilità. Come sopra, utilizzeremo qui una notazione leggermente diversa rispetto al testo:  $p_i$  in luogo di  $p_{++}$ , eccetera.

Si possono verificare due casi. Se  $k_{min} < k_{max}$ , la nuova distribuzione è definita da

$$\begin{aligned}
 p_i &= 0, \quad i < k_{min}, \\
 p_i &= \frac{1}{1-2q} \left[ \sum_{i=k_{min}} f_i - q \right], \quad i = k_{min}, \\
 p_i &= \frac{1}{1-2q} f_i, \quad k_{min} < i < k_{max}, \\
 p_i &= \frac{1}{1-2q} \left[ \sum_{i \geq k_{max}} f_i - q \right], \quad i = k_{max}, \\
 p_i &= 0, \quad i > k_{max}.
 \end{aligned}
 \tag{8.1}$$

Se invece  $k_{min} = k_{max}$ , la nuova distribuzione è definita da

$$\begin{aligned}
 p_i &= 0, \quad i < k_{min}, \\
 p_i &= 1, \quad i = k_{min}, \\
 p_i &= 0, \quad i > k_{max}.
 \end{aligned}
 \tag{8.2}$$

Le espressioni per le quantità  $p_i$  non sono molto trasparenti. I lettori e le lettrici con una formazione scientifica avanzata possono trovare utile visualizzare le formule precedenti con l'aiuto di una analogia nel continuo. Se si considera il caso di una densità di probabilità  $f(x)$  con  $x \in \mathcal{R}$ , allora  $k_{min}$  e  $k_{max}$  sono definiti da

$$\int_{-\infty}^{k_{min}} dx f(x) = \int_{k_{max}}^{\infty} dx f(x) = q$$

e

$$\begin{aligned}
 p(x) &= 0, \quad x < k_{min}, \\
 p(x) &= \frac{f(x)}{1-2q}, \quad k_{min} < x < k_{max}, \\
 p(x) &= 0, \quad x > k_{max}.
 \end{aligned}$$

In altre parole, la nuova distribuzione  $p(x)$  è ottenuta eliminando le code su entrambi i lati della distribuzione  $f(x)$  e normalizzando. La percentuale dei dati eliminati dipende dalla scelta di  $q$ .

**Osservazione.**

*Nella nostra analisi abbiamo scelto*

$$q = 1/4,$$

*eliminando la metà delle risposte fornite, in modo da avere distribuzioni decisamente diverse da quelle considerate nei due casi precedenti. In altre parole abbiamo assunto un atteggiamento prudentiale, al possibile costo di una perdita di informazione, pur di poter garantire l'affidabilità delle valutazioni finali.*

Definita la distribuzione  $p_i$ , l'indicatore  $r$  è stato ottenuto, come nei casi precedenti, usando la notazione

$$a^{++} = a_1, \quad a^+ = a_2, \quad a^- = a_3, \quad a^{--} = a_4,$$

e ponendo

$$r = a_1 p_1 + a_2 p_2 + a_3 p_3 + a_4 p_4.$$

Anche per la distribuzione censurata potremmo effettuare le due diverse scelte per i parametri  $a_i$ , che definiscono i metodi 1 e 2. Per semplicità ci limiteremo a considerare il caso

$$a_1 = 1, \quad a_2 = 0.8, \quad a_3 = 0.2, \quad a_4 = 0.$$

#### **Osservazione.**

*Orvviamente la scelta fatta per la distribuzione censurata non è la sola possibile. Un'altra possibilità, fra le più frequentemente utilizzate, è l'utilizzo delle distribuzioni dette winsorised. Per esempio avremmo potuto definire*

$$(8.3) \quad \begin{aligned} p_i &= 0, & i < k_{\min}, \\ p_i &= \sum_{i \leq k_{\min}} f_i, & i = k_{\min}, \\ p_i &= f_i, & k_{\min} < i < k_{\max}, \\ p_i &= \sum_{i \geq k_{\max}} f_i, & i = k_{\max}, \\ p_i &= 0, & i > k_{\max}. \end{aligned}$$

*Non ci aspettiamo risultati diversi rispetto al caso censurato considerato prima.*

## **9. Appendice B: distribuzioni M3 per i CdS triennali della Facoltà**

In questa Appendice riportiamo le distribuzioni per le domande 5 e 12 ottenute utilizzando il metodo M3. Riportiamo i risultati per tutti i CdS della Facoltà di Scienze MFN.

Un elemento di valutazione della qualità della didattica universitaria

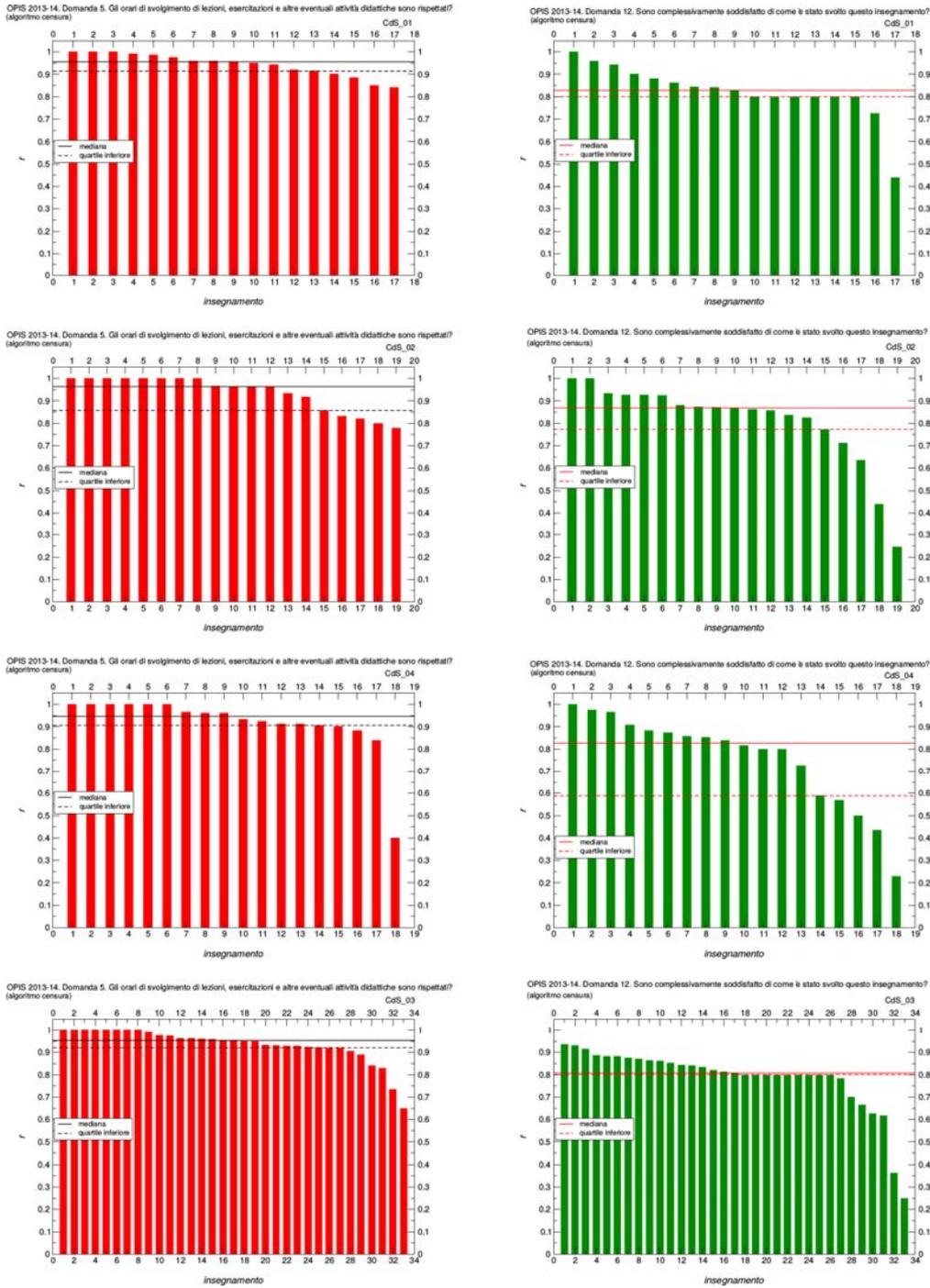


Figura 2. Gradi di soddisfazione come funzione del *ranking* ottenuti con il metodo M3. Pannelli di sinistra: domanda 5 (rosso); pannelli di destra: domanda 12 (verde). (I)

Un elemento di valutazione della qualità della didattica universitaria

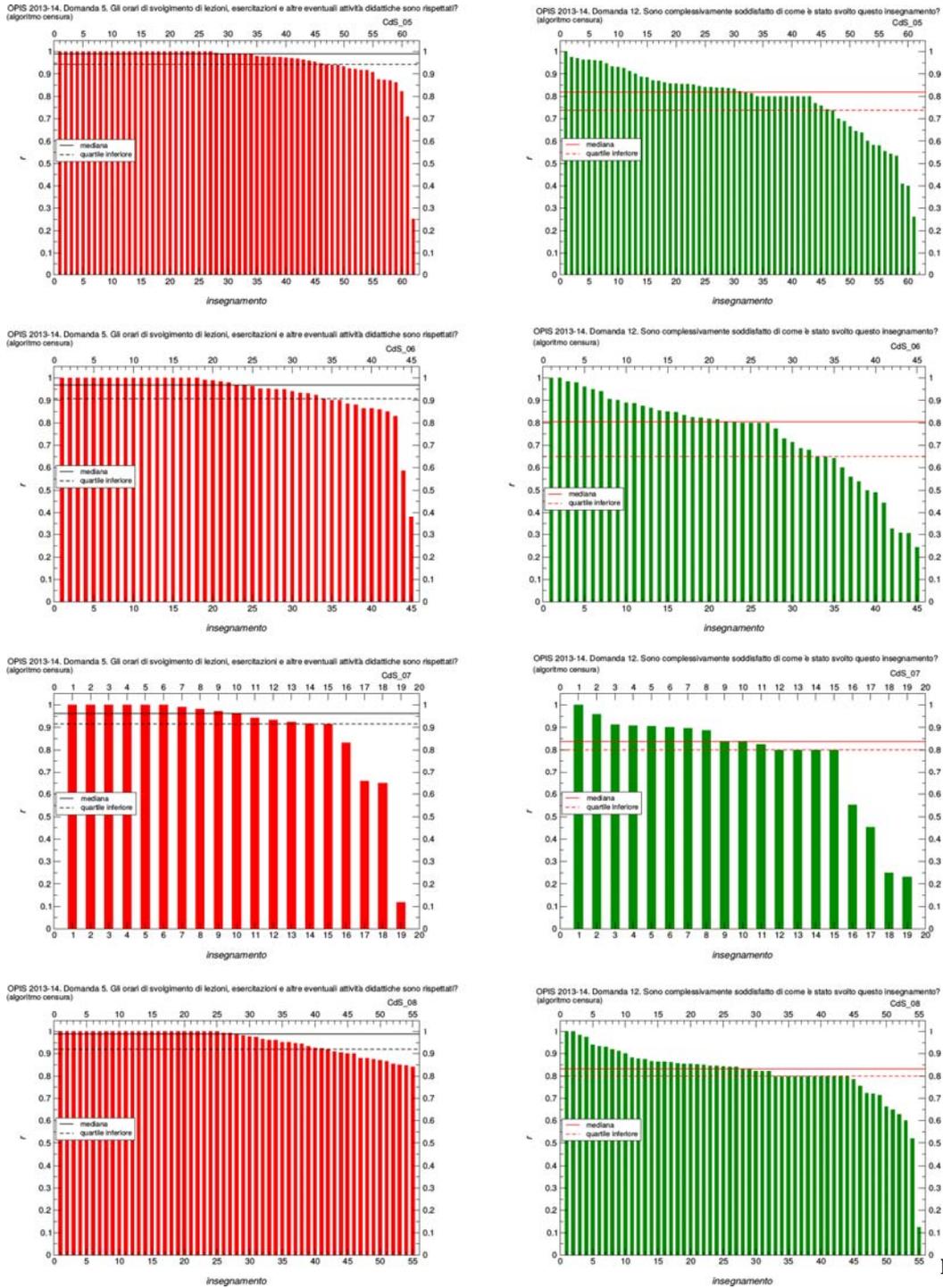


Fig  
ura

Figura 3. Gradi di soddisfazione come funzione del *ranking* ottenuti con il metodo M3. Pannelli di sinistra: domanda 5 (rosso); pannelli di destra: domanda 12 (verde). (II)

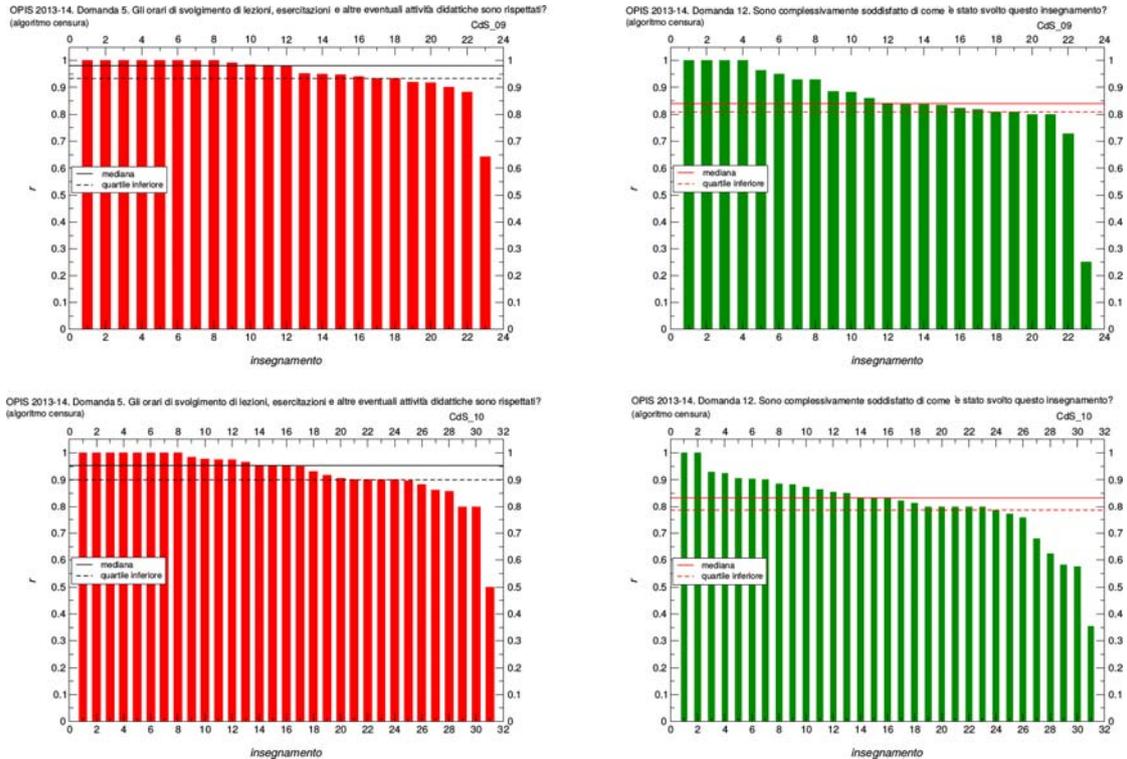


Figura 4. Gradi di soddisfazione come funzione del *ranking* ottenuti con il metodo M3. Pannelli di sinistra: domanda 5 (rosso); pannelli di destra: domanda 12 (verde). (III)

## References

- Greco Pietro, Silvestrini Vittorio (2009) “La risorsa Infinita. Per una società democratica della conoscenza”. Editori Riuniti
- Greco Pietro, Silvestrini Vittorio (2010) “The never-ending resource. For a democratic society of knowledge”. CUEN, Naples.
- Hastings Cecil, Mosteller Frederick, Tukey John W., Winsor Charles P. (1947) “Low moments for small samples: a comparative study of order statistics” *The Annals of Mathematical Statistics*, 18: 413-426
- Tukey John W. (1962) “The future of data analysis” *The Annals of Mathematical Statistics*, 33:1-67
- Tukey John W. (1977), “Exploratory Data Analysis” Pearson