

## L'indicatore standardizzato di performance dipartimentale (ISPD)

*Risposta al Comunicato CUN del 18 luglio 2017*

*Approvato dal Consiglio Direttivo dell'ANVUR del 9 ottobre 2017*

Nel corso dell'estate è stato diffuso dal CUN un comunicato (Prot. 20400 del 18/7/2017) che solleva alcune perplessità sull'utilizzo dell'indicatore standardizzato di performance dipartimentale (ISPD). Questa nota tecnica intende chiarire le principali differenze fra indicatori standardizzati e indicatori semplicemente normalizzati e rispondere per quanto possibile ai dubbi sollevati dal CUN.

È innanzitutto opportuno rammentare come la valutazione della qualità dei prodotti della ricerca misuri la reazione, nel senso del grado di approvazione, della comunità scientifica di riferimento a un risultato della ricerca. Va dunque tenuto in conto che diverse comunità differiscono per standard e metodologie di valutazione, così come diverse sono le tipologie più comuni di prodotti (articoli, progetti, monografie, ecc.), le modalità operative della ricerca (individuale, in piccoli o grandi gruppi, teorica o sperimentale), e quelle relative alle pubblicazioni (riviste nazionali o internazionali, monografie, ecc.). Inoltre – e questo riguarda specificamente gli esercizi VQR ad oggi completati – il numero di prodotti richiesti per addetto è stato fissato uguale per tutte le discipline ed è quindi diversamente rappresentativo della produzione complessiva dei ricercatori nei diversi settori scientifico-disciplinari.

Queste forti disomogeneità di contesto rendono di fatto impossibile il confronto trasversale fra le valutazioni dei prodotti di settori scientifico-disciplinari diversi.

Un esempio può essere fatto considerando due settori di aree attigue (Fisica e Matematica) e che condividono la stessa tipologia di prodotti (prevalentemente articoli scientifici): osserviamo che i prodotti del settore FIS/01 (Fisica Sperimentale) hanno un punteggio medio di 0,81, mentre quelli del settore MAT/01 (Logica Matematica) hanno il punteggio medio 0,6. Trarne la conseguenza che la ricerca italiana nel campo della Fisica Sperimentale sia di qualità superiore a quella nella Logica Matematica sarebbe del tutto ingiustificato. Nell'esempio specifico pesa inoltre il fatto che gli articoli consegnati nel settore FIS/01 hanno mediamente 611 autori, contro 2,1 di MAT/01, e che ciascun prodotto è stato presentato alla VQR mediamente da 6 addetti diversi, contro 1,2 di MAT/01. Un così diverso grado di proprietà dei prodotti rende, già da solo, di fatto non confrontabili le valutazioni fra i due settori. Ancora più eclatante è il confronto interno all'area Fisica fra i settori FIS/01 e FIS/02 (Fisica Teorica): in quest'ultimo il voto medio è più basso 0,75, e registriamo mediamente 13 autori per articolo, ognuno dei quali è stato presentato 1,2 volte.

Non è dunque appropriato utilizzare i punteggi VQR come strumento per stabilire, tramite confronto diretto, se un gruppo di ricercatori di un'area o di un settore ha prodotto una ricerca migliore di un altro.

Nel tentativo di raggiungere questo scopo si possono percorrere due strade, entrambe guidate dall'idea di sostituire il confronto diretto dei voti ottenuti in settori diversi con il posizionamento del gruppo entro il medesimo settore nella scala nazionale.

La prima via consiste nel considerare, tramite una somma o una media, i *voti normalizzati per settore* (o per area), ottenuti dividendo il voto del singolo prodotto per il voto medio del settore scientifico-disciplinare (o dell'area). In questo maniera sono costruiti i noti indicatori qualitativi IRAS1 e IRAS2 e quelli puramente qualitativi R e X. I primi sono la somma dei voti

normalizzati, i secondi ne costituiscono invece il valore medio dell'aggregato in considerazione. In entrambi i casi, il prodotto con valutazione eccellente di un fisico sperimentale risulta valere meno di quello di un logico ( $1/0,81=1,18$  il primo contro  $1/0,6=1,67$  il secondo), riflettendo la maggiore difficoltà di ottenere tale giudizio nel settore della logica rispetto a quello della fisica sperimentale.

Anche tale metodologia non è scevra da distorsioni e, mentre gli indicatori normalizzati di tipo somma sono troppo sensibili alla numerosità dell'aggregato, gli indicatori di tipo media risentono invece del difetto opposto, favorendo gli aggregati meno numerosi. Risultano poi invariabilmente penalizzati i settori con medie più elevate rispetto a quelli con medie meno elevate.

La seconda strada è un raffinamento della normalizzazione e conduce all'elaborazione di indicatori standardizzati. I *voti standardizzati* possono essere ottenuti tramite sottrazione della media e divisione per la varianza del settore scientifico-disciplinare. Per esempio, il voto standardizzato di un prodotto eccellente di un fisico sperimentale vale ora 0,57, contro 1,06 nel caso di un logico matematico. Di nuovo, la differenza rispecchia la maggiore difficoltà, in sede di VQR, di ottenere un voto eccellente nel settore della Logica Matematica rispetto alla Fisica Sperimentale. Da notare che il valore medio dei voti standardizzati è zero, mentre nella versione normalizzata era uno. Osserviamo inoltre che, rispetto alla semplice normalizzazione, la standardizzazione ha moderato la differenza relativa rispetto al valore medio.

La procedura di standardizzazione però non risolve completamente i problemi già evidenziati nel caso del voto normalizzato. Per questo motivo non è corretto confrontare dipartimenti o altri aggregati disomogenei utilizzando il voto standardizzato.

ANVUR ha quindi adottato una diversa grandezza, l'ISPD, che consente di stimare il posizionamento percentuale di quell'aggregato<sup>1</sup> nella scala di tutti i suoi simili (i cosiddetti "dipartimenti virtuali", vale a dire tutti i dipartimenti con la stessa composizione in termini di SSD teoricamente costruibili combinando tutti i docenti italiani ad essi appartenenti).<sup>2</sup> Viene così evitato il confronto tra dipartimenti non omogenei.

Il CUN paventa che nel passaggio fra voto standardizzato e ISPD il valore venga distorto da effetti non lineari e che "due dipartimenti con voti standardizzati molto vicini tra loro possono invece avere valori di ISPD estremamente diversi". Ciò non è però possibile: infatti vale una disuguaglianza per cui il valore assoluto della differenza fra i valori di ISPD (ovviamente non espresso in percento) di due dipartimenti diversi è sempre inferiore a quello della differenza fra i corrispondenti voti standardizzati.

Nello stesso documento del CUN si afferma inoltre che "è possibile infatti costruire esempi teorici di due dipartimenti della medesima area disciplinare per i quali il voto standardizzato di uno è del 7% più grande di quello dell'altro e, per di più, i due voti standardizzati non differiscono significativamente né tra loro né con il voto standardizzato medio nazionale dell'area (sulla base dei test statistici di significatività più diffusi), mentre invece l'ISPD del primo dipartimento è 14 volte più grande di quello del secondo". Ora, il voto standardizzato può assumere ogni valore positivo o negativo, compreso lo zero che è il suo

---

<sup>1</sup> Negli usi pratici dell'indicatore standardizzato, è sempre più opportuno utilizzare il valore della funzione di ripartizione a posto del valore del voto standardizzato, nonostante i due producano la medesima graduatoria. Il voto standardizzato non costituisce un parametro affidabile su cui basare una ripartizione lineare delle risorse. Valori assoluti del voto standardizzato superiori a 2 vanno considerati come anomalie statistiche: infatti al voto standardizzato 1,8 corrisponde il posizionamento nel percentile 96% nella scala dei dipartimenti virtuali, mentre un voto standardizzato doppio (3,6) si situa nella posizione nel percentile 99%.

<sup>2</sup> Per consentire il confronto diretto tra settori diversi, la correzione della standardizzazione richiederebbe che le distribuzioni delle votazioni dei dipartimenti virtuali fossero rigorosamente normali, proprietà vera solamente nel limite asintotico.

valore medio (per effetto della stessa standardizzazione). Se due voti standardizzati differiscono di poco dal valore medio standardizzato (che è, come già osservato, zero), differiranno di poco anche fra loro (in valore assoluto) e tale prossimità verrà conservata dall'ISPD relativo in virtù della succitata disuguaglianza: si troveranno entrambi verso il cinquantesimo percentile.

L'affermazione del CUN risulta dunque di difficile interpretazione, anche perché l'esempio teorico in questione non viene riportato nel documento. Vi è forse la possibilità che l'esempio si riferisca a confronti fra il voto NON standardizzato, ISPD e il voto medio nazionale di area NON standardizzato (giacché quello standardizzato, come detto, è nullo). Tali confronti fra numeri in scale diverse andrebbero però sempre evitati.

In conclusione, la definizione dell'indicatore standardizzato si fonda su un modello matematico solido (il modello dei dipartimenti virtuali), utilizza nozioni relativamente elementari della teoria della probabilità, ed è calcolabile con facilità nonostante rifletta un'elaborazione un po' più complicata di quella necessaria dell'indicatore semplicemente normalizzato. È d'altra parte uno strumento molto più potente perché permette di confrontare efficacemente la qualità della ricerca di aggregati (dipartimenti o altre sottostrutture accademiche) di grandezze e composizioni diverse anche all'interno dello stesso Ateneo, pur evitando inappropriati confronti diretti fra le valutazioni dei prodotti di aree disciplinarmente eterogenee e quindi di natura diversa.

ANVUR è lieta che il CUN riconosca che l'ISPD sia un indicatore adatto a formare la graduatoria dei primi 350 dipartimenti italiani (alla fine della Raccomandazione si legge infatti che "... l'indicatore scelto dall'ANVUR è coerente con l'obiettivo di formulare...") e da parte sua condivide pienamente la preoccupazione sottesa alla Raccomandazione circa il rischio di possibili usi distorti degli indicatori. A più riprese ANVUR ha raccomandato di non utilizzare gli indicatori messi a punto per specifici obiettivi al di fuori del loro corretto ambito di applicazione.<sup>3</sup> ANVUR ribadisce che tali usi distorti sono da evitare, anche a livello di Ateneo, e segnala il rischio che essi possano portare all'assunzione di decisioni non sostenute da reali elementi valutativi.

---

<sup>3</sup> A questo fine, l'ANVUR ricorda che l'indicatore standardizzato è stato introdotto nel 2014, al termine di una collaborazione tra l'ANVUR e la Commissione Ricerca della CRUI, e come nel documento finale siano espressamente menzionati alcuni possibili esempi di applicazione da parte degli Atenei [ <http://www.anvur.org/attachments/article/609/Dipartimento%20virtuale%20associato%20e%20voto%20standardizzato%20FINALE.pdf>].