

Il VQR ed il vino a due stadi

2012-03-03 12:45:44 By Alberto Baccini



ovvero cosa può accadere se i giudici di una gara di vini devono astenersi dal berlo

Ci sarà modo di commentare [ancora](#) in dettaglio i criteri di valutazione appena pubblicati. Gli elementi di originalità della metodologia ANVUR di valutazione sono tali che il VQR si configura come una sperimentazione di inediti (e poco fondati) strumenti bibliometrici, condotta sull'intero sistema della ricerca italiano.

In un paese come il nostro in cui la cultura della valutazione non è poi così diffusa, si è deciso di inventarsi un modello originale di valutazione della ricerca. In cosa consiste l'originalità? Si è scelto di valutare la produzione migliore (3 prodotti/6 prodotti) di ogni ricercatore, come nel REF britannico. Si è dunque scelta la via più raffinata: la valutazione del singolo prodotto di ricerca. La qualità della ricerca è come quella del vino: si riconosce solo dopo averlo bevuto. Fuori di metafora: l'unico modo per valutare il singolo prodotto di ricerca è leggercelo, e dare un giudizio. Che sarà idiosincratico, personale, soggettivo etc. etc. Ma è un giudizio informato di un pari. Il sistema che ha fatto progredire la scienza dai tempi della rivoluzione scientifica. Nel REF britannico i giudici assaggiano il vino: tutti i prodotti sono sottoposti a peer review. Purtroppo il "capitale sociale" dell'accademia italiana è così basso che si tende a non fidarsi della peer review; e soprattutto l'ex-ministro Gelmini ed i suoi consiglieri non si fidavano dei baroni. Per cui l'ex-ministro e l'ANVUR da lei nominata hanno preferito ricorrere all'oggettività della bibliometria. L'originalità consiste quindi nell'aver disegnato un concorso enologico dove si pretende di valutare il vino senza che i giudici possano berlo.

Si può fare un concorso di vino senza che i giudici possano berlo? Non lo so. Per la ricerca si può fare

e si fa: moltissimi esercizi di valutazione sono condotti in giro per il mondo senza sottoporre tutti i prodotti a peer review, utilizzando gli strumenti della bibliometria. Ci vogliono però accortezza e competenze tecniche. Si parte quindi dall'ipotesi che contare le citazioni di un prodotto di ricerca sia una buona approssimazione del giudizio che si avrebbe se quel prodotto lo si fosse letto davvero. Approssimativo, ma accettabile e accettato. Un po' come dire: ipotizziamo che il numero di volte che un vino è citato nelle riviste enologiche sia una misura della sua qualità. Da questa ipotesi sono partiti verosimilmente anche l'ex-ministro e l'ANVUR.

Nella letteratura internazionale c'è diffusa consapevolezza che gli strumenti bibliometrici siano particolarmente poco adatti alle valutazioni puntuali, quelle su cui, come abbiamo visto, è basato il VQR. La speranza dei bibliometrici è gli indicatori siano statisticamente affidabili, ma nessun bibliometrico si sognerebbe di dare un giudizio su un singolo prodotto di ricerca, basandosi esclusivamente sul numero di citazioni che ha ricevuto o sulla rivista su cui è stato pubblicato. In analisi statistiche di grandi insiemi di prodotti di ricerca si spera che le distorsioni si annullino ed i risultati siano in media affidabili. La bibliometria è cioè poco adatta a fare il mestiere che l'ex-ministro e l'ANVUR vogliono farle fare: fornire un giudizio affidabile sul singolo prodotto di ricerca. L'esempio forse più noto di valutazione basata esclusivamente sulla "qualità delle riviste" fu condotto in Australia nel 2010. Ad ogni prodotto di ricerca fu attribuito il valore di "qualità" del contenitore in cui era apparso. Il sistema, adesso abbandonato, aveva sicuramente alcuni pregi: assoluta trasparenza e controllabilità; facilità estrema di implementazione, costi ridotti. La realizzazione del sogno: datemi il ranking delle riviste che solleverò il mondo! A differenza di quanto è previsto nel VQR, nessuno in Australia si sognò di comunicare ai singoli ricercatori il giudizio sui loro prodotti: perché era già noto ex ante! E soprattutto perché poco affidabile in relazione ai singoli prodotti.



Come si risolve il problema? Se si abbandona l'idea di una valutazione singolarmente affidabile e ci si concentra sulla ragionevolezza dei risultati aggregati, il problema potrebbe essere risolto con una tecnica usuale e diffusissima negli esercizi di valutazione, di cui ho già detto nella pars construens di un [post precedente](#). Essa consiste nell'attribuire ad ogni prodotto un valore citazionale normalizzato rispetto al campo disciplinare.

Nella VQR italiana questa soluzione standard è impedita dal 50-10-20-20. Cosa significa? L'ex-ministro e l'ANVUR hanno pensato bene che i prodotti di ricerca siano stratificati secondo questa formula: 50% di prodotti di qualità limitata; 10% di prodotti accettabili; 20% di prodotti buoni; 20% di prodotti eccellenti. Se si hanno a disposizione le citazioni di tutti i prodotti di ricerca di una certa disciplina, tutti i prodotti che hanno un numero di citazioni inferiori al numero mediano possono essere classificati come limitati; quelli che hanno un numero di citazioni superiore al valore mediano, ma inferiore al valore assunto dalla distribuzione nel 60° percentile, sono accettabili e così via. Il problema nasce da questo: le distribuzioni delle citazioni sono a code pesanti; per cui i valori soglia di una distribuzione 50-10-20-20 tendono ad essere terribilmente vicini. Per capirci: è del tutto possibile

che in moltissime discipline siano sufficienti 1-2 citazioni in più per essere classificati in una delle categorie. Potrebbe accadere per esempio che in una disciplina di Web Of Science un prodotto con 0 citazioni sia classificato come di qualità limitata; con 1 citazione diventi accettabile; con 2 buono, e con 4 eccellente. O ancora che non sia possibile distinguere la soglia tra limitato e accettabile. Risultati che confliggono con il buon senso. O meglio risultati che mostrano la mutua inconsistenza del ranking in quattro classi del singolo prodotto di ricerca e della tecnologia bibliometrica al momento disponibile.

Soluzione possibile: buttare via la formula 50-10-20-20, correggendo il bando. Ma se una formula è magica non si può toccare. L'ANVUR aveva originariamente pensato ad una inconsistente procedura a due stadi, di cui abbiamo [già scritto](#). Nei criteri finali le soluzioni bibliometriche a due stadi sono diventate due. Una per i prodotti pubblicati nel periodo 2004-2007; una per quelli prodotti nel periodo 2008-2010. Questi due sistemi sono alla base delle "matrici di corrispondenza tra classi iniziali della rivista (colonne) e citazioni (righe)".

Per i prodotti pubblicati nel periodo 2004-2007 prima si assaggia il vino, sia mai, scusate, si contano le citazioni al vino sulle riviste enologiche, e poi si corregge per la regione di provenienza. Per quelli 2008-2010 si dà il giudizio sulla regione di provenienza, poi si contano le citazioni sulle riviste enologiche. E' ben noto, almeno dal [seminal paper](#) di P.O. [Seglen](#), che la correlazione tra impact factor della rivista e numero di citazioni ricevute da ogni paper che vi è pubblicato è molto bassa. Per capirsi, pubblicare un articolo su una rivista con elevato impact factor non garantisce che si riceverà un numero elevato di citazioni. Anche per ogni singolo autore, il grado di correlazione tra citazioni agli articoli e impact factor della rivista è variabilissimo, come si vede nella Figura 1. Ogni grafico rappresenta un ricercatore ed i punti le sue pubblicazioni, posizionate in base al numero di citazioni ricevute e all'IF della rivista dove ogni articolo è pubblicato. Come si può vedere ci sono articoli con un numero elevato di citazioni pubblicati su riviste con basso IF e viceversa.

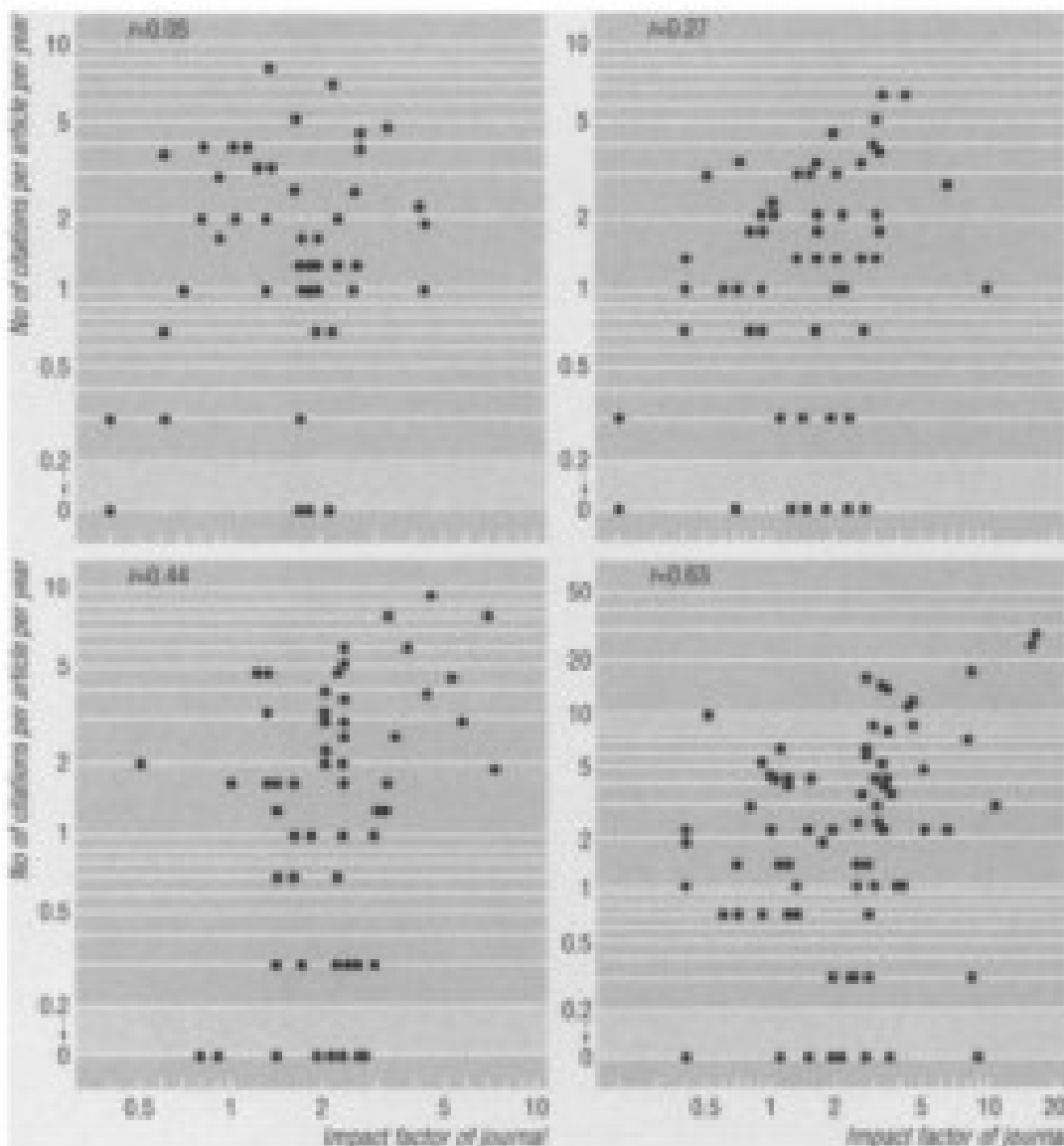


Fig 3 Correlation between article citation rate and journal impact for four authors¹¹

Figura 1. Correlazione tra IF di una rivista e citazioni degli articoli per quattro ricercatori

L'assenza di correlazione rende impossibile costruire un ordinamento consistente degli articoli che tenga conto contemporaneamente delle due dimensioni considerate. Questo significa che dietro la retorica scientifica della costruzione della matrice e della raffinatezza dei criteri ci sono in realtà scelte arbitrarie dei valutatori. Per questo sarebbe più onesto chiamare quelle matrici i "quadrati magici bibliometrici" dell'ANVUR. Sono i giudici del concorso che hanno deciso che se un Chianti è di qualità limitata vale quanto un vino abruzzese eccellente, e tutti e due vanno in B, o C o D o all'assaggio (peer review).

A mia conoscenza, ma forse nella sterminata letteratura bibliometrica mi sono perso qualcosa, non esistono procedure a due stadi così disegnate. Si ribatterà: nella scienza c'è sempre spazio per l'innovazione. Ma le idee scientifiche innovative dovrebbero essere validate e diffuse secondo la prassi usuale, ovvero la pubblicazione su riviste scientifiche internazionali peer reviewed. L'ANVUR invece sta sperimentando la sua originale bibliometria direttamente sull'intero sistema italiano della



ricerca.

POST SCRIPTUM. La metafora del vino è ispirata da un articolo pacato e competente sui rating delle università scritto dal prof. Andrea Bonaccorsi, vice responsabile del VQR, sul Sole24ore di domenica scorsa. Vi si legge:

Non è vero che la realtà sociale è misurabile esattamente allo stesso modo di quella fisica (e quindi le pretese di obiettività vanno senz'altro messe in discussione). ... Tra visibilità mediatica, basata su un'illusoria certezza dei numeri, e pretesa di incommensurabilità esiste dunque una terza alternativa, meno roboante, ma per fortuna più robusta.

Sarei curioso di sapere se il prof. Bonaccorsi ritiene i criteri ANVUR per il VQR un esempio di questa terza alternativa.

Copyright :

All this contents are published under [Creative Commons Attribution-NonCommercial-ShareAlike 2.5 Generic License](http://creativecommons.org/licenses/by-nc-sa/2.5/).

for reproduced, please specify from this website [ROARS](http://www.roars.it/) AND give the URL.

Article link : <http://wp.me/p1WBc2-1li>