



# Valutazione della Qualità della Ricerca 2011-2014 (VQR 2011-2014)

Il confronto tra la distribuzione dei valori di IRAS1 degli atenei  
italiani nelle due VQR (2004-2010 e 2011-2014)

## APPENDICE A

In questa appendice si presentano i risultati preliminari di un confronto tra la varianza delle distribuzioni di IRAS1 (normalizzato per la dimensione degli atenei) nelle due VQR. Tenuto conto che IRAS1 è il principale indicatore utilizzato per misurare la qualità della produzione scientifica e per la distribuzione della componente premiale dell'FFO, l'analisi appare interessante per capire se, successivamente alla prima VQR, si è verificato un processo di convergenza tra gli atenei o al contrario, se c'è stata una divaricazione.

Gli esercizi qui effettuati mirano a rendere il più possibile omogeneo il confronto tra le distribuzioni di IRAS1 nelle due VQR, eliminando gli effetti di alcuni elementi spuri, relativi ai cambiamenti nelle modalità di assegnazione dei punteggi. Queste correzioni vengono applicate gradualmente, per maggiore trasparenza. In estrema sintesi, i risultati confermano la convergenza tra atenei rilevata sulla base dei dati non modificati ma tendono a ridimensionare l'intensità del fenomeno.

Va sottolineato che il confronto che qui si presenta è per sua natura imperfetto, per almeno due ordini di motivi. In primo luogo, le correzioni apportate, ovviamente, non possono rendere perfettamente confrontabili le due VQR, che differiscono per moltissimi aspetti, non ultimo il ricambio degli stessi valutatori; per tenere conto di questo elemento, ad esempio, avremmo dovuto chiedere ai GEV correnti di valutare anche i prodotti della VQR precedente (o un campione di essi). Molte delle differenze tra le due VQR non sembrano però avere, almeno a priori, un impatto univoco sul fenomeno qui esaminato; è il caso, ad esempio, della richiesta di conferire 3 prodotti su 7 anni di attività, nella prima VQR, a fronte di 2 prodotti su 4 anni nella seconda. In secondo luogo, l'impegno volto a rispettare le scadenze per la pubblicazione del Rapporto VQR ha limitato i tempi dell'analisi; nei prossimi mesi, quindi, si cercherà di affinare le metodologie utilizzate e di valutare se altri fattori di disomogeneità sono suscettibili di analisi.

Il confronto qui presentato rappresenta inoltre solo un punto di partenza, perché i risultati suggeriscono immediatamente ulteriori approfondimenti. Ad esempio, è importante identificare il ruolo avuto da due distinti ordini di fattori: b) le modifiche nella popolazione accademica; a) le modifiche nei comportamenti degli addetti coinvolti in entrambe le VQR, relativamente sia alla produzione scientifica e alle forme di pubblicazione sia alla selezione dei prodotti per il conferimento. Per questi fattori, inoltre, vanno chiarite le rispettive cause, quali ad esempio: tendenze già in atto nei comportamenti accademici; modifiche nelle scelte di reclutamento da parte degli atenei; ecc.

Gli esercizi vogliono correggere disomogeneità nella scala di valutazione adottata dai GEV nelle due VQR, in base ai rispettivi bandi. La Tabella 1, costruita sulla base delle indicazioni dei bandi, mette a confronto le due scale di valutazione adottate e, nella terza colonna, mostra le

scelte qui effettuate per rendere omogenei i due esercizi. I decili della tabella fanno riferimento alla distribuzione della produzione scientifica internazionale dell'area in cui si colloca il prodotto valutato. Ad esempio, sulla base del bando VQR 2011-2014, il giudizio *Elevato* (0,7) andava attribuito a un prodotto che “si colloca nel segmento 10%-30% della distribuzione della produzione scientifica internazionale dell'area cui appartiene”.

Caratteristiche del prodotto	VQR 2004-2010	VQR 2011-2014	Esercizio controfattuale	
Scorretti (plagio, ecc.)	Plagio: -2	Non valutabile 0	Limitato: 0 (esercizio 1)	
Sottomesso ma errato	Non valutabile: -1	Non valutabile 0		
Non sottomesso	Mancante: -0.5	Non valutabile 0		
Decile 1	Limitato: 0	Limitato: 0		
Decile 2				
Decile 3				
Decile 4		Accettabile: 0.1		
Decile 5				
Decile 6	Accettabile: 0.5	Discreto: 0.4		Discreto: 0.4 (esercizio 2)
Decile 7	Buono: 0.8	Elevato: 0.7		Elevato: 0.7 (esercizio 2)
Decile 8				
Decile 9	Eccellente: 1			
Decile 10			Eccellente: 1	

Tabella 1 – Scala di valutazione adottata nelle due VQR e proposta di omogeneizzazione.

La prima VQR penalizzava la mancata o l'errata presentazione di prodotti con punteggi negativi, mentre la seconda attribuisce valore nullo a queste situazioni<sup>1</sup>. **Nel primo esercizio**, tutti i prodotti con punteggio negativo nella prima VQR vengono quindi convertiti in valori nulli. Questo produce un accorciarsi della coda bassa nella prima VQR con conseguente innalzamento della media (ma non della mediana) e una riduzione della dispersione tra atenei (otteniamo quindi la distribuzione *VQR 2004-10 HI* della Figura 1). Inoltre, per migliorare ulteriormente la comparabilità delle due distribuzioni, anche i giudizi di “accettabile” con voto 0.1 nella nuova VQR vengono convertiti in valori nulli, producendo un effetto analogo a quello delle

<sup>1</sup> La modifica del punteggio ha mirato a ridurre gli incentivi a comportamenti opportunistici, quali la produzione di lavori puramente per sottomettere formalmente un prodotto.

trasformazioni che subisce la prima VQR (otteniamo quindi la distribuzione *VQR 2011-14* della Figura 1)<sup>2,3</sup>.

Le due scale di valutazione differiscono anche nella parte centrale e sommitale della distribuzione. In particolare, la definizione di prodotti *Eccellenti* faceva riferimento ai primi due decili della distribuzione della produzione internazionale nella prima VQR, e al primo decile nella seconda. **Nel secondo esercizio**, partendo dai risultati del primo, cerchiamo quindi di correggere queste ulteriori disomogeneità. I risultati di questa ulteriore elaborazione corrispondono alla distribuzione *VQR 2004-10 H2* della Figura 1. Essi vanno però considerati con estrema cautela, perché vengono divise a metà due categorie della prima VQR senza disporre di graduatorie interne ad esse. In questo caso abbiamo infatti proceduto randomizzando la valutazione delle categorie *Eccellente* (voto 1) e *Buono* (voto 0.8) e attribuendo le metà casualmente più basse nella categoria inferiore vedi terza colonna della Tabella 1). Va inoltre rilevato come il peso della categoria “eccellente” sul totale dei prodotti valutati si sia ridotto solo del 12%, mentre ci si attenderebbe un calo dell’ordine del 50% nelle ipotesi di una sostanziale stabilità della performance della ricerca italiana (come rappresentata dai prodotti conferiti) nel confronto internazionale e di un numero approssimativamente simile dei prodotti italiani nei due decili.

La Figura 1 presenta le distribuzioni di IRAS1 (normalizzato per la dimensione degli atenei e con media posta pari a zero) per la VQR 2011-2014 e per la VQR 2004-2010 sulla base delle diverse ipotesi sopra indicate. Ai fattori spuri qui esaminati è attribuibile circa la metà della riduzione nella deviazione standard della distribuzione tra la prima e la seconda VQR (in base sia ai risultati del primo esercizio sia a quelli del secondo). Anche correggendo per disomogeneità nelle scale di valutazione rimarrebbe quindi una convergenza tra atenei, che risulta significativa sulla base di un test F di Fisher sul rapporto delle due varianze. Prendendo come riferimento il primo esercizio (in base al quale l’intensità della convergenza risulta minore), le due varianze, rispettivamente pari a 0,05 e 0,03, forniscono un valore di F pari a 1,51 che risulta significativo con un p-value inferiore al 3%.

---

<sup>2</sup> In questo esercizio e in tutte le seguenti analisi si fa riferimento ai 91 atenei presenti in entrambe le VQR.

<sup>3</sup> Per semplicità, visto che le differenze per effetto di questa correzione sono minime, non si riporta la distribuzione non modificata della VQR 2011-14.

La convergenza tra atenei è inoltre confermata anche quando si ponderi le deviazioni dalla media dei singoli atenei sulla base delle loro dimensioni (*deviazione standard ponderata* riportata in Figura 1).

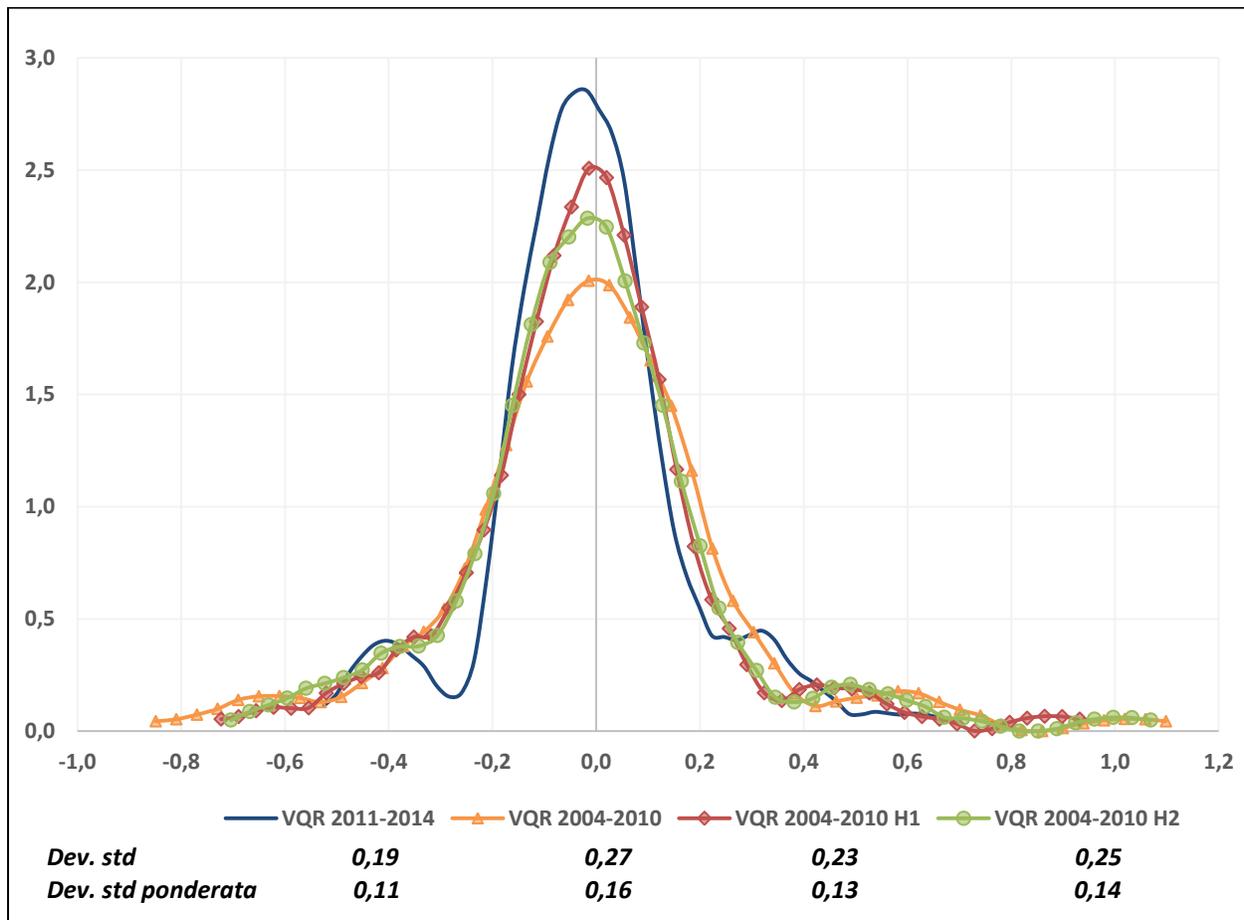


Figura 1 – Distribuzione dei valori di IRAS1 (normalizzato per la dimensione dell’ateneo e ponendo la media pari a zero) per gli atenei italiani.

Indicando con I1 e I2 i valori di IRAS1 (normalizzato per la dimensione degli atenei e con media posta pari a zero) rispettivamente nella prima e nella seconda VQR, è stata inoltre stimata una funzione di regressione lineare avente come variabile dipendente la differenza tra I2 e I1 e come variabile indipendente I1. In questo caso, la stima del coefficiente di regressione relativo a I1 fornisce una misura di convergenza (valori negativi) divergenza (valori positivi) degli atenei nel passaggio tra la prima e la seconda VQR.

Come riportato in Tabella 2, la stima del coefficiente di regressione risulta negativa e altamente significativa per entrambi gli esercizi. Questo indica che all'aumentare di I1 la differenza tra le due VQR diminuisce, ossia che le università con un valore inferiore alla media di IRAS1 nella prima VQR hanno aumentato la propria valutazione più di quelle che presentavano un valore maggiore (Figura 2).

	Parametri	Coefficienti	Errore standard	Stat t	Valore di significatività
Esercizio 1	Intercetta	-0,002	0,011	-0,175	0,861
	I1	-0,326	0,049	-6,694	0,000
Esercizio 2	Intercetta	-0,005	0,011	-0,424	0,673
	I1	-0,383	0,045	-8,448	0,000

Tabella 2 – Statistiche sulla regressione lineare stimata (variabile dipendente I2-I1)

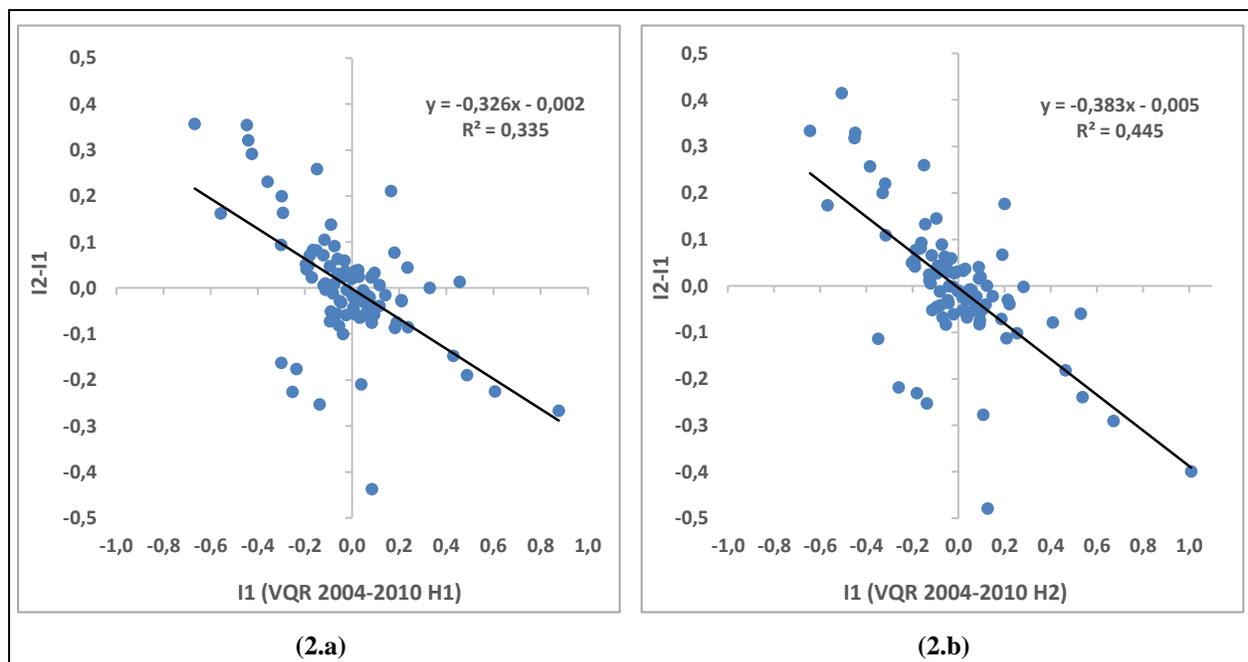


Figura 2 – Regressione lineare avente per variabile dipendente la differenza tra I2 e I1 e per variabile indipendente I1 (2.a: I1= VQR 2004-2010 H1; 2.b: I1= VQR 2004-2010 H2)

Questo avvicinamento degli atenei tra la prima e la seconda VQR non avviene per tutte le aree di valutazione. L'analisi della distribuzione per atenei del voto medio standardizzato (Tabella 3)

mostra che per 3 delle sedici aree si registra invece un aumento della deviazione standard (quest'ultima rimane costante in una quarta area), prendendo come riferimento il secondo esercizio; il numero delle aree in cui aumenta la dispersione degli atenei sale a 4 (e in tre aree la deviazione standard rimane costante) prendendo come riferimento il primo esercizio.

Area	VQR 2011-2014	VQR 2004-2010	VQR 2004-2010 H1	VQR 2004-2010 H2
1	0,30	0,45	0,33	0,36
2	0,17	0,17	0,14	0,15
3	0,21	0,19	0,18	0,20
4	0,27	0,33	0,32	0,33
5	0,24	0,38	0,36	0,38
6	0,31	0,47	0,39	0,39
7	0,34	0,40	0,37	0,38
8a	0,32	0,38	0,29	0,33
8b	0,25	0,31	0,27	0,27
9	0,24	0,33	0,28	0,31
10	0,19	0,28	0,19	0,22
11a	0,22	0,24	0,19	0,19
11b	0,36	0,39	0,36	0,39
12	0,21	0,30	0,24	0,23
13	0,36	0,59	0,51	0,54
14	0,34	0,39	0,34	0,34

**Tabella 3 – Deviazione standard del voto medio standardizzato (R) degli atenei per ciascuna area di valutazione**