Il confronto basato sul Dipartimento Virtuale Associato e sul "Voto standardizzato"

(G. Poggi - 24/2/2014)

Premessa:

Questo documento illustra un metodo di valutazione dei Dipartimenti universitari basato sulla qualità dei prodotti di ricerca presentati nell'esercizio VQR. Il metodo consente di definire indicatori della qualità dei Dipartimenti basati sull'analisi dei prodotti di ricerca aggregati per settore scientifico-disciplinare (SSD nel seguito); procedure analoghe possono essere applicate anche aggregando i prodotti per Area CUN.

Gli indicatori proposti consentono un utilizzo completo dell'informazione raccolta nell'esercizio VQR; in questo senso l'approccio proposto va visto come complementare e aggiuntivo alle analisi delle valutazioni dei prodotti di ricerca prodotte fino ad ora dall'ANVUR, che — come previsto dal bando ministeriale — sono confinate alle Aree CUN.

Il documento è diviso in quattro parti.

La prima parte ha lo scopo di descrivere il senso della proposta e di motivare l'efficacia degli indicatori proposti, anche riguardo alla loro capacità di affinare l'analisi della "qualità" della produzione scientifica dipartimentale basata sui risultati dell'esercizio VQR. Una illustrazione di tali concetti, accompagnata da esempi di calcolo, si può trovare anche nelle presentazioni effettuate il giorno 9/1/2014 presso la CRUI.

La seconda parte illustra due indicatori ispirati a tale metodologia. Essi forniscono valori estremamente simili, del tutto coerenti fra loro come risultati. Il primo è interpretabile come estensione diretta di quanto già prodotto dall'ANVUR; il secondo, detto *Voto standardizzato*, è più soddisfacente dal punto di vista statistico. Stante la loro sostanziale equivalenza e la maggiore solidità teorica del secondo, si suggerisce agli Atenei di utilizzare quest'ultimo, se lo riterranno opportuno nella loro autonomia, per operare un confronto della qualità della ricerca dei dipartimenti, coerente con l'esercizio VQR.

Nella terza parte si commentano le tabelle fornite dall'ANVUR che contengono l'indicatore di Voto standardizzato; si fornisce anche qualche esempio di possibile uso dell'indicatore per l'erogazione di risorse.

La quarta parte, più tecnica, è dedicata a coloro che vogliono approfondire e verificare la solidità dell'approccio seguito.

Queste tabelle rappresentano la prima di una serie di elaborazioni dei dati ANVUR-VQR che verranno a breve pubblicate, frutto di una collaborazione tra l'ANVUR e la Commissione Ricerca della CRUI.

1. Il Dipartimento Virtuale Associato e l'indicatore di Voto standardizzato

L'esercizio VQR, e in particolare la valutazione dei prodotti della ricerca, si presta ad essere utilizzato da parte degli Atenei per distribuire ai Dipartimenti risorse non dipendenti dai costi della ricerca quali, ad esempio, i punti organico. Ricordiamo che il loro utilizzo richiede l'individuazione dei Settori Concorsuali (SC nel seguito) su cui bandire le procedure, che a loro volta consentono la definizione di profili unicamente basati sui SSD del SC. Poiché l'esercizio VQR ha prodotto, in accordo con una esplicita richiesta della CRUI all'ANVUR, valutazioni accorpate anche a livello di SSD, pare opportuno partire proprio da queste ultime[1] per costruire indicatori utilizzabili sia per lo scopo anzidetto sia, più in generale, per permettere agli atenei interessati di confrontare fra loro i propri dipartimenti secondo la qualità di ricerca, così come misurata dai risultati VQR.

I problemi principali che si pongono nella costruzione di indicatori di dipartimento sono sostanzialmente dovuti al fatto che ogni dipartimento è una realtà a sé stante, composta spesso di un numero variabile di membri appartenenti a diversi SSD. Confrontare secondo la qualità della ricerca un Dipartimento di Lettere e Filosofia con uno di Matematica oppure con uno di Medicina è sostanzialmente impresa complessa, da molti addirittura considerata impossibile, proprio per la eterogeneità delle composizioni e delle aree scientifico-disciplinari coinvolte e quindi della presunta incomparabilità delle valutazioni adottate dai vari GEV nell'esercizio VQR.

Per dare un'idea di come sia possibile superare questo problema, consideriamo un caso limite: confrontare valutazioni assegnate dalla VQR a membri[2] di SSD molto diversi. Il confronto è senz'altro non banale, ma forse è sbrigativo sostenere che sia impossibile. Riteniamo infatti che sia identificabile una soluzione sufficientemente generale e soddisfacente al problema. Il metodo che si propone si ispira a quello cui fanno ricorso le Università più prestigiose — tipicamente negli USA — quando chiedono informazioni su un nostro studente che ha fatto domanda per essere ammesso ad uno dei loro corsi: non ci chiedono il voto che gli abbiamo assegnato in un nostro insegnamento, ma piuttosto in quale percentile (*top* %) della distribuzione dei nostri studenti esso si colloca. Se dichiariamo, dopo aver confrontato il voto che abbiamo assegnato allo studente con la distribuzione completa dei voti dell'insegnamento, che egli si colloca nel top 5%, vuol dire che la probabilità di trovare uno studente migliore di quello (ovviamente secondo il nostro metro di giudizio) è bassa, pari appunto solo al 5%.

L'idea è di procedere in modo simile per confrontare le valutazioni in due SSD completamente differenti: confrontare non il valore assoluto delle due valutazioni VQR ma la loro posizione nella distribuzione nazionale delle valutazioni nei rispettivi SSD. A seconda del percentile della distribuzione delle votazioni dei propri rispettivi SSD nel quale esse si collocano possiamo decidere quale delle due valutazioni è migliore o peggiore: *la valutazione migliore è quella per la quale è minore il numero di votazioni più alte assegnate nello stesso SSD*. Usando questa tecnica potremmo ad esempio scoprire che una valutazione pari a 0.866 in un certo SSD è un voto migliore di una valutazione di 0.933 in un altro SSD

---

[1] Le considerazioni svolte in questo documento, basate sulle valutazioni di SSD, possono, *mutatis mutandis,* basarsi anche sulle valutazioni di Area. Se un Ateneo fosse interessato all'utilizzo delle valutazioni di Area nell'ambito dell'approccio proposto, troverà (come descritto brevemente nella terza parte di questo documento) nelle tabelle l'indicatore corrispondente.

[2] Conviene ribadire che la VQR non dev'essere mai usata per valutare i singoli; questo caso limite è introdotto solo per chiarire i ragionamenti successivi.

Manifestamente questa metodologia di confronto richiede che le valutazioni VQR nei rispettivi SSD siano state compiute con equità e ragionevolezza (così come si assume che noi abbiamo operato con i nostri studenti), ma non impone, per essere valida, alcun comune metro di giudizio fra i due GEV che hanno operato sui due SSD nazionali, né richiede che tali GEV abbiano prodotto valutazioni compatibili in termini di valore medio di voti o di ampiezza di escursione degli stessi. Per essere applicato, questo criterio richiede esclusivamente di conoscere le distribuzioni dei voti assegnati dalla VQR ai vari membri dei due SSD considerati.

Accettato questo principio per il confronto fra membri di SSD differenti, il problema è ora quello di vedere se è possibile estendere il metodo al caso, indubbiamente più complesso, del confronto fra dipartimenti non mono-settoriali (o più generalmente e forse più elegantemente, fra qualunque aggregazione di membri di vari SSD — per esempio Atenei, collegi di Dottorato e così via). La difficoltà del problema, manifestamente, dipende dal fatto che dato un certo dipartimento è impossibile valutare il percentile in cui si colloca: per farlo occorrerebbe disporre di un numero molto ampio di dipartimenti composti in termini di SSD come il nostro e che tutti fossero stati valutati nella VQR. Ovviamente ciò non si verifica mai e ogni dipartimento costituisce un caso a sé stante.

Limitandoci ai dipartimenti (ma il ragionamento si può applicare a qualunque insieme di membri), è comodo introdurre, per ogni dipartimento dell'ateneo, un Dipartimento Virtuale o "Dipartimento Specchio" ad esso Associato (DVA nel seguito): questo è un dipartimento ipotetico (ovvero inesistente ma perfettamente definibile in termini operativi) composto da una distribuzione di membri nei vari SSD identica a quella del nostro dipartimento reale (DR nel seguito). Ovviamente, mentre la VQR ha assegnato un voto a tutte le pubblicazioni presentate e quindi un voto medio a tutti i membri del nostro dipartimento reale, la VQR non ha assegnato alcun voto ai membri (anch'essi virtuali) del DVA. Possiamo però farlo noi stessi; anzi possiamo — concettualmente e anche operativamente è possibile — calcolare tutte le combinazioni possibili per i voti che i membri del DVA avrebbero potuto ottenere, estraendoli a caso dall'insieme delle valutazioni nazionali dei rispettivi SSD. Nonostante questo esercizio appaia francamente estenuante, viste le numerosissime combinazioni di voto che saremmo stati costretti a identificare e calcolare, supponiamo di averlo compiuto.

Per procedere oltre, occorre ora individuare un Indicatore Ind_d che si possa calcolare in maniera perfettamente definita in termini matematici sia per il DR, partendo dai voti dei membri del nostro dipartimento reale, sia per tutte le possibili combinazioni di voti del DVA. Assumiamo anche che Ind_d sia una funzione crescente con il voto di ogni singolo membro.

Supponiamo di aver definito tale indicatore, per esempio combinando in maniera opportuna i voti ottenuti dai singoli membri dei SSD presenti nel dipartimento. Da ora in poi il valore dell'indicatore Ind_d calcolato sul DR sarà accompagnato con un asterisco (Ind^*_d) mentre i valori che lo stesso indicatore assumerà per le combinazioni di voti del DVA non avranno alcun asterisco (Ind_d); entrambi manterranno l'indice *d,* comune a DR e DVA, perché entrambi i dipartimenti hanno la stessa identica composizione interna. È importante osservare che mentre Ind^*_d assume un valore unico, vale a dire quello calcolato partendo dalla valutazione VQR reale dei membri del DR, Ind_d rappresenta un insieme, anche molto numeroso, di valori: precisamente tutti quelli possibili (e diversi fra loro) che si ottengono associando ai membri del DVA tutte le possibili combinazioni dei voti presenti nelle valutazioni nazionali dei rispettivi SSD. Per maggior chiarezza, quando ci riferiremo esplicitamente all'insieme di questi valori useremo la notazione tipografica $\{Ind_d\}$. Chiamiamo inoltre N_{DVA} il numero di modi con cui si possono combinare i voti nazionali sui SSD del DR e quindi del DVA.

Ora, disponendo di Ind^*_d e della distribuzione di Ind_d, (ovvero di {Ind_d}) possiamo verificare come si colloca Ind^*_d nella distribuzione complessiva di {Ind_d}. Precisamente, contando il numero di casi in cui Ind_d del DVA è superiore rispetto al valore Ind^*_d di DR e dividendo questo numero per N_{DVA}, determiniamo la probabilità $P\_sup(Ind^*_d)$ di ottenere un DVA con un valore dell'indicatore superiore a quello di DR, ovvero superiore a Ind^*_d. Definita così $P\_sup(Ind^*_d)$, è immediato calcolare la probabilità di trovare DVA con valutazione inferiore: $P\_inf(Ind^*_d) = 1 - P\_sup(Ind^*_d)$. Sinteticamente possiamo dire che se $P\_sup(Ind^*_d)$ è molto minore di 1 (ovvero prossimo a zero) il DR è molto buono; se viceversa $P\_sup(Ind^*_d)$ è circa 1, il DR è molto modesto.

A questo punto possiamo estendere in maniera molto naturale quanto visto prima per due votazioni ottenute in SSD diversi, quando per trovare un modo per confrontarle, indipendentemente dalle specificità di comportamento e valutazione dei GEV coinvolti, siamo andati a determinare la probabilità di individuare, nelle rispettive valutazioni nazionali, votazioni di valore superiore. Infatti, supponiamo di trovarci in presenza di due dipartimenti qualsiasi — DR_1 e DR_2 — comunque composti in termini di SSD, con associati i relativi *DVA*. Per i due DR abbiamo i due valori Ind^*_1 e Ind^*_2 e per entrambi immaginiamo, seguendo la procedura appena descritta, di calcolare i due insiemi di valori {Ind_1} e {Ind_2}, e poi di calcolare, sempre come detto sopra, $P\_sup(Ind^*_1)$ e $P\_sup(Ind^*_2)$. A questo punto siamo in grado, coerentemente con il nostro approccio, di confrontare la qualità VQR dei due dipartimenti reali DR, come segue:

- se $P\_sup(Ind^*_1) = P\_sup(Ind^*_2)$, i due dipartimenti reali D_1 *e* D_2 sono di pari qualità;
- se $P\_sup(Ind^*_1) > P\_sup(Ind^*_2)$, allora D_1 è di qualità peggiore di D_2, in quanto per D_1 è più elevata, rispetto a D_2, la probabilità di trovare un DVA con un valore superiore dell'indicatore.

Poiché i due dipartimenti sono qualsiasi, il ragionamento e la procedura si possono estendere a qualunque insieme di dipartimenti, comunque essi siano composti. Pertanto abbiamo definito un metodo operativo per confrontare la qualità VQR di qualunque insieme di dipartimenti. Ovviamente i criteri, sopra riportati, per confrontare i dipartimenti possono essere formulati in maniera equivalente facendo ricorso a $P\_inf(Ind^*_d)$.

Resta un problema che, se non risolto, renderebbe il metodo in pratica inapplicabile o applicabile solo a costo di un imponente impegno di calcolo. Infatti il calcolo di tutti gli insiemi {Ind_d} risulterebbe estremamente laborioso, dato il numero dei membri di ognuno dei circa 900 dipartimenti universitari italiani, che la legge impone non inferiore a 40 per la maggioranza di essi.

Viceversa saremmo a posto se riuscissimo a dimostrare, o quanto meno a giustificare con argomenti solidi, che è possibile definire un indicatore direttamente calcolabile IND_d, comunque funzione crescente del voto medio di ogni membro dei DR, con la proprietà di permettere l'identificazione univoca del $P\_sup(IND^*_d)$, nello stesso modo per tutti i *DR*; in altre parole, *a valori diversi di* IND^*_d *devono corrispondere valori diversi* $P\_sup(IND^*_d)$ *e questa corrispondenza deve essere la stessa quale che sia il DR che stiamo considerando*. Infatti, in questa ipotesi, basterebbe calcolare IND^*_d per determinare la probabilità di trovare valori di IND_d superiori nel suo DVA; e quindi il calcolo di IND_d sarebbe sufficiente per confrontare tutti i dipartimenti secondo qualità.

Si può far vedere che esiste un'ampia classe di indicatori con queste caratteristiche. Uno di essi è l'indicatore $\Delta=(IDVA-1)/\sigma$ elaborato inizialmente dal gruppo di lavoro presso la CRUI. Un altro indicatore, anche più adatto e più coerente con la metodologia suggerita, è il <u>*Voto standardizzato di dipartimento*</u> U_d, di cui è stato chiesto il calcolo all'ANVUR e i cui valori per tutti i dipartimenti italiani sono oggi messi a disposizione degli Atenei. Ricordiamo, per inciso, che il calcolo dettagliato di questi indicatori può essere effettuato solo dall'ANVUR per via della impossibilità di accesso pubblico ai dati per aggregazioni di meno di 10 prodotti.

Vale la pena notare esplicitamente che gli indicatori IND_d, essendo funzioni crescenti del voto medio dei membri del DR, possono essere utilizzati direttamente come parametro quantitativo per erogare risorse, con le modalità illustrate nella terza parte di questo stesso documento[3].

2. **Gli indicatori possibili: *Δ = (IDVA-1)/σ* e *U* = Voto standardizzato di dipartimento**

In questa seconda parte si illustra brevemente la tecnica con la quale è possibile individuare due indicatori che — con un maggiore o minor livello di accuratezza — soddisfano le condizioni richieste di corrispondenza univoca con il *P_sup*.

Facciamo dapprima riferimento all'indicatore $\Delta=(IDVA-1)/\sigma$, partendo dalla definizione dell'indicatore "primario" *IDVA*. Si tratta dell'indicatore individuato all'interno del gruppo CRUI per illustrare all'ANVUR il nostro approccio. Esso è scelto anche in questo documento come primo esempio di applicazione della metodologia proposta a causa della sua stretta parentela con *R*, il cosiddetto "terzo indicatore" dell'ANVUR-VQR, certamente familiare a chi ha analizzato i risultati di tale esercizio.

Dato un dipartimento *d* con N_d membri totali distribuiti in gruppi di n_s membri per SSD *s*, detto R_s il valore del terzo indicatore VQR (il rapporto fra il voto medio dei membri del SSD *s* del dipartimento e quello nazionale), l'indicatore $IDVA_d$ vale:

$$IDVA_d = \frac{1}{N_d}\sum_{s=1}^{NSSD} R_s \cdot n_s = \sum_{s=1}^{NSSD} R_s \cdot \pi_s = \langle R_d \rangle \, ,$$

dove *NSSD* è il numero totale di SSD, e $\pi_s = n_s / N_d$ è la percentuale di membri del dipartimento *d* afferenti al SSD *s.* In altre parole, $IDVA_d$ è un *R* medio di dipartimento; infatti esso consiste nella media pesata sul dipartimento dei rapporti R_s di SSD.

Si trova che $IDVA_d$ possiede, calcolato sui DVA, valore atteso pari a 1; inoltre esso è crescente con i valori delle valutazioni di qualunque membro. Si trova anche che se andassimo a guardare la distribuzione di $\{IDVA_d\}$ calcolate sui DVA troveremmo delle distribuzioni tutte gaussiane (questo è un risultato comune a molti degli indicatori definibili sui dipartimenti[4]) con lo stesso valore centrale (il valore atteso, pari a 1), ma aventi diverse larghezze (deviazioni standard) $\sigma(IDVA_d)$, dipendenti dalla specifica suddivisione in settori del dipartimento *d*. La deviazione standard di questa gaussiana si può stimare proprio partendo dalla distribuzione delle valutazioni nazionali dei vari SSD presenti nel dipartimento:

$$\sigma^2\left(IDVA_d\right) = \sum_{s=1}^{NSSD} \pi_s^2 \sigma^2\left(R_s\right) = \frac{1}{N_d}\sum_{s=1}^{NSSD} \pi_s \frac{\sigma_s^2}{\langle I_s \rangle^2},$$

---

[3] Va sottolineato tuttavia che tale indicatore non tiene conto della numerosità del dipartimento, contrariamente a quanto fanno invece gli indicatori IRD di dipartimento della VQR. Nella terza parte del documento si mostrerà come esso possa essere opportunamente modificato al fine di renderlo utilizzabile per la distribuzione di risorse finanziare.

[4] La ragione di ciò va ricercata nel cosiddetto Teorema del Limite Centrale (TLC); esso assicura che una variabile aleatoria data dalla somma di numerose variabili aleatorie indipendenti con valore medio e varianza definite tende ad assumere un andamento gaussiano all'aumentare del numero degli addendi. I nostri indicatori, costruiti sul DVA, sono assimilabili a variabili aleatorie indipendenti dal punto di vista della statistica; esse sono anche molto numerose, proprio perché ad ogni dipartimento per la legge 240/2010 devono afferire non meno di 40 membri. È stato verificato mediante simulazioni Montecarlo che effettivamente le distribuzioni di IDVA sono con buona approssimazione gaussiane, nonostante che le valutazioni dei singoli prodotti siano discrete e distribuite in maniera assolutamente non gaussiana.

dove σ_s^2 rappresenta la varianza campionaria delle valutazioni nazionali del SSD. La seconda uguaglianza discende dal fatto che R_s è il rapporto fra il voto medio di n_s membri del SSD s rispetto al voto medio nazionale $\langle I_s \rangle$ del *SSD* e quindi la sua deviazione standard è data da:

$$\sigma^2\left(R_s\right) = \frac{\sigma_s^2}{n_s \cdot \langle I_s \rangle^2} ,$$

dove la presenza al denominatore di n_s deriva dalla riduzione della varianza dovuta alla presenza del voto mediato su n_s membri, mentre σ_s^2 è la varianza delle votazioni dei singoli membri.

In conseguenza di ciò, la conoscenza di $IDVA^*_d$ permetterebbe di calcolare la probabilità di ottenere un valore di $IDVA_d$ maggiore di $IDVA^*_d$, in quanto sappiamo che la distribuzione di $IDVA_d$ è gaussiana di valore atteso 1 e deviazione standard $\sigma(IDVA_d)$. Cambiando però dipartimento (*d'* invece che *d*), pur restando il valore atteso pari a 1, cambia la deviazione standard di $IDVA_{d'}$ rispetto quella di $IDVA_d$, per cui la relazione fra indicatore e *P_sup* non è univoca e dipende dal dipartimento considerato. È però semplicissimo ottenere l'univocità voluta: basta passare dall'indicatore $IDVA_d$ all'indicatore

$$\Delta_d = \frac{\left(IDVA_d - 1\right)}{\sigma\left(IDVA_d\right)} ,$$

che ha la proprietà di essere distribuito come una gaussiana centrata intorno a 0 e con deviazione standard *sempre pari a 1* (gaussiana standardizzata).

In questo modo, comunque sia composto il dipartimento reale DR e quindi il suo DVA, gli elementi dell'insieme dei valori dell'indicatore $\{\Delta_d\}$ sono distribuiti sempre nello stesso modo (gaussiana standardizzata); di conseguenza, noto Δ^*_d è univocamente determinato il $P\_sup(\Delta^*_d)$ e possiamo confrontare la qualità VQR in qualunque insieme di dipartimenti.

Come anticipato prima, pur potendosi definire più di un indicatore che soddisfi alle condizioni richieste di univoca relazione con il *P_sup*, l'indicatore teoricamente e matematicamente più soddisfacente e che oltretutto può essere applicato in maniera particolarmente diretta e semplice a qualunque aggregazione di membri valutati nell'esercizio VQR (volendo anche presi da dipartimenti diversi), è il ***Voto standardizzato di dipartimento***, indicato nel seguito con U_d. Un indicatore molto simile è stato suggerito da C. Mortarino della Università di Padova; la principale differenza fra il suo e quello qui proposto è che quest'ultimo elimina ogni residuo effetto di dipendenza impropria dell'indicatore dalla numerosità dei membri nelle aggregazioni.

Riportiamo qui solo la formula che definisce U_d, rimandando gli interessati alla quarta parte di questo documento per la descrizione di come è stata ricavata. Prima di tutto, per ciascun membro *l* (elle) del dipartimento *d* afferente al SSD *s* indichiamo con $u_{l,s,d}$ il suo *voto standardizzato*

$$u_{l,s,d} = \frac{v_{l,s,d} - \langle I_s \rangle}{\sigma_s}$$

dove $v_{l,s,d}$ rappresenta la votazione media del singolo membro, $\langle I_s \rangle$ è la votazione media del SSD *s* a livello nazionale, e σ_s^2 la varianza campionaria delle valutazioni nazionali del SSD già introdotta. Il voto standardizzato di un membro rappresenta quindi la valutazione media dei prodotti attesi dal membro depurata dalle differenze introdotte dalle diverse metodologie di valutazione usate dai vari GEV.

Il voto standardizzato di dipartimento è una combinazione dei voti standardizzati dei suoi membri, opportunamente pesati tenendo presente il numero di prodotti che ciascun membro doveva presentare, calcolata come segue:

$$U_d = \frac{1}{\sqrt{\sum_{s=1}^{NSSD} \sum_{l=1}^{N_{s,d}} np_{l,s,d}^2}} \sum_{s=1}^{NSSD} \sum_{l=1}^{N_{s,d}} u_{l,s,d} \cdot np_{l,s,d}$$

dove $np_{l,s,d}$ rappresenta il numero di prodotti che il membro *l* del dipartimento *d afferente al* SSD *s* doveva presentare, e $N_{s,d}$ rappresenta il numero di membri afferenti al SSD *s* nel dipartimento *d*; ovviamente nel dipartimento *d* alcuni SSD saranno privi di membri e per questi $N_{s,d}$ sarà pari a zero.

Nonostante che la distribuzione del voto standardizzato di singolo membro (una variabile discreta) sia spesso assai asimmetrica e che quindi essa non sia e non possa essere gaussiana, la distribuzione di *U*, calcolata per il *DVA* come somma pesata (e normalizzata per avere varianza unitaria) di numerose variabili aleatorie indipendenti, è con buona approssimazione una gaussiana standardizzata, ovvero $\{U_d\}$ su *DVA* ha media nulla e deviazione standard pari a uno; essendo la distribuzione gaussiana, la conoscenza di U^*_d permette di determinare univocamente il $P\_sup(U^*_d)$, esattamente come fatto per *Δ.* A differenza dell'indicatore *Δ*, la correzione per la variabilità dei voti e delle loro distribuzioni tra i vari GEV in *U* è effettuata sul singolo membro e non come aggiustamento finale sull'indicatore; ciò conferisce all'indicatore *U* caratteristiche teoriche migliori rispetto a *Δ.* D'altra parte, un'analisi condotta su tutti i dipartimenti mostra che, per ogni dipartimento, i valori numerici degli indicatori Δ^*_d e U^*_d sono estremamente simili, e quindi forniscono risultati analoghi nel momento in cui sono applicati per confrontare la qualità dei dipartimenti. La nostra preferenza va all'indicatore di Voto Standardizzato esclusivamente per la sua maggiore robustezza teorica e per la sua maggiore generalità, apprezzabile soprattutto quando si voglia estendere l'indicatore a qualunque insieme di membri dei vari SSD.

3. <u>**Le tabelle prodotte dall'ANVUR e l'utilizzo degli indicatori**</u>

Le tabelle contengono le informazioni e gli indicatori per tutti i dipartimenti nazionali ex-lege 240/2010. Pertanto mancano gli Atenei che non hanno inviato all'ANVUR l'informazione relativa. Inoltre i "dipartimenti" indicati con -999 rappresentano le aggregazioni di membri dell'ateneo corrispondente che, pur valutati nell'esercizio VQR, sono cessati dal servizio prima della costituzione di tali dipartimenti.

Le tabelle sono organizzate in due fogli:

1. Il primo foglio contiene due gruppi di colonne.
    a. Il primo gruppo a sinistra mostra, dopo gli identificatori degli Atenei e dei Dipartimenti, i dettagli sul numero di prodotti e di soggetti valutati dei dipartimenti stessi, nonché nella terz'ultima colonna l'indicatore "*IDVA = R medio del dipartimento*", nella penultima la deviazione standard di *IDVA* e infine nell'ultima il "*Voto Standardizzato di dipartimento*"[5]. Si ricorda che l'indicatore *IDVA* non corregge né gli effetti di variabilità di valutazione dei GEV, né la dipendenza dell'indicatore dalla numerosità dei membri nelle aggregazioni. Tali effetti sarebbero corretti, anche se in maniera meno puntuale rispetto a quanto ottenibile con all'indicatore di Voto standardizzato, dall'indicatore Δ=(IDVA-1)/σ che non è qui riportato esplicitamente in quanto

---

[5] Lo stesso indicatore, già in fase di utilizzo e studio da parte di alcuni Atenei facenti parte del Gruppo di lavoro CRUI (ad esempio *La Sapienza*), è denominato con il simbolo Q (come qualità).

sostituito dal voto standardizzato. Il valore di σ nella penultima colonna è la deviazione standard che compare nella formula precedente. ***Si sconsiglia in ogni modo l'utilizzo diretto di IDVA per l'erogazione di risorse ai dipartimenti;*** esso è riportato esclusivamente perché potrebbe essere utile nel determinare l'escursione massima alla quale riportare il parametro ($1+\alpha\times IPR^*_d$, definito nel seguito) destinato alla erogazione delle risorse.

b. Le tre colonne a destra riportano le informazioni più direttamente utilizzabili per l'erogazione di risorse non dipendenti dal costo della ricerca: la prima colonna contiene l'ambito di *P_sup (Top% stimato)* dove si colloca il dipartimento, il $P\_inf(U^*_d)$ e infine l'indicatore IPR^*_d suggerito per l'erogazione delle risorse direttamente legato alla collocazione nel percentile, descritto di seguito.

2. Il secondo foglio, oltre ai soliti identificativi di Ateneo e dipartimento, fornisce l'informazione disaggregata dei SSD del dipartimento, sia per quanto riguarda il numero di prodotti e di soggetti valutati, sia per il voto standardizzato (terz'ultima colonna). Nelle ultime due colonne, in corrispondenza di ogni riga sono riportate la valutazione media e la deviazione standard dell'insieme nazionale del SSD corrispondente. Tutte queste informazioni, nel rispetto della riservatezza cui l'ANVUR è tenuta, sono riportate solo quando il numero di prodotti presentati per quel SSD e quel dipartimento sono almeno pari a 10. Il valore del Voto standardizzato per SSD nel dipartimento può aiutare le decisioni del dipartimento qualora esso sia chiamato a distribuire al proprio interno risorse assegnategli dall'Ateneo mediante il voto standardizzato di dipartimento, riportato nel primo foglio. Infatti i valori di voto standardizzato di SSD nel dipartimento consentono di comprendere quanto i vari SSD al suo interno hanno contribuito al valore del voto standardizzato di dipartimento, come quantificato dalle formule della quarta parte di questo documento.

Esiste anche un secondo file, sostanzialmente identico al primo come struttura, nel quale però le elaborazioni sono effettuate al livello delle 14+2 Aree CUN invece che a livello di SSD. Esso è fornito per venire incontro a quegli Atenei che durante l'incontro del 9/1/2014 presso la CRUI hanno espresso il desiderio, in continuità con il bando VQR, di procedere a livello di Area CUN. Non ci siamo soffermati su questo aspetto, ma è evidente che le considerazioni finora svolte operando sulle valutazioni di SSD possono applicarsi immediatamente, insieme con gli indicatori e il formalismo adottato, alle valutazioni di Area CUN.

Trattiamo ora brevemente il tema delle modalità di utilizzo dell'indicatore come parametro numerico per assegnare risorse.

L'indicatore *IDVA,* che come detto altro non è che un *R* medio sul dipartimento, assume valori centrati attorno al valore 1 e quindi potrebbe essere usato direttamente come fattore moltiplicativo della dimensione N_d del dipartimento al fine di ottenere un indicatore quali-quantitativo, come molti di quelli prodotti dall'ANVUR, se non soffrisse dei difetti già segnalati (non correzione delle difformità di valutazione dei vari GEV, e dipendenza dalle dimensioni degli insiemi esaminati).

Il voto standardizzato U_d rimuove molti di questi difetti, in quanto confronta i dipartimenti sulla base della probabilità di avere DVA con votazioni superiori a quelle osservate per i DR. Sarebbe quindi possibile, partendo da U_d, costruire un indicatore di dipartimento US_d (come U_d scalato) che assume valori centrati attorno al valore 1 e che variano in un intervallo di larghezza determinata da un parametro α. Basterebbe infatti porre

$US^*_d = 1 + U^*_d \cdot \alpha$.

Meglio sarebbe riferirsi esplicitamente alla grandezza che ci ha guidato per confrontare i dipartimenti, ovvero la probabilità che il DVA assuma valori superiori o inferiori a quelli osservati per il DR; ricordiamo che valori di U^*_d elevati comportano una probabilità molto piccola $P\_sup(U^*_d)$ che il DVA assuma valori maggiori di U^*_d, ovvero una probabilità $P\_inf(U^*_d) = 1 - P\_sup(U^*_d)$ molto alta che il DVA assuma valori inferiori a U^*_d. Tale probabilità si può stimare sfruttando la gaussianità del voto standardizzato di dipartimento calcolato sui DVA:

$$P_{\inf}(U^*_d) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{U^*_d} e^{-\frac{x^2}{2}} dx$$

Per coloro che sono abituati ad utilizzare fogli elettronici, questa probabilità è esprimibile come DISTRIB.NORM.(U^*_d, 0,1,1). Proprio le variazioni di questa probabilità rispetto a 0.5 potrebbero essere usate per costruire l'indicatore per distribuire risorse non dipendenti dal costo della ricerca, similmente a quanto accennato prima:

$$US^*_d = 1 + \left(P_{\inf}\left(U^*_d\right) - 0.5\right) \cdot \alpha$$

Rispetto all'utilizzo diretto di U^*_d, questo indicatore offre il vantaggio di premiare (penalizzare) in maniera simile tutti quei dipartimenti che hanno avuto un voto standardizzato positivo e significativamente maggiore di uno (negativo e significativamente minore di uno). In altre parole, quando si tratta di erogare risorse, è sostanzialmente indifferente sapere se il dipartimento si colloca nel top 2% (*$U^*_d = 1.65$, $(P\_inf(U^*_d) - 0.5) = 0.451$*) oppure nel top 1% (*$U^*_d = 2.3$, $(P\_inf(U^*_d) - 0.5) = 0.489$*) oppure nel top 0.1% (*$U^*_d = 3.1$, $(P\_inf(U^*_d) - 0.5) = 0.499$*); si tratta in ogni caso di un dipartimento con ottime valutazioni VQR.

Detto quindi IPr^*_d l'indicatore da utilizzare per l'erogazione delle risorse (sulla base di quanto detto sopra si suggerisce $IPr^*_d = P\_inf(U^*_d) - 0.5$, abbiamo:

$$US^*_d = 1 + IPR^*_d \cdot \alpha$$

L'indicatore $IPr^*_d = P\_inf(U^*_d) - 0.5$ è quello riportato nell'ultima colonna di destra nel primo foglio delle tabelle.
In generale, la determinazione di α è affidata a scelte dell'Ateneo, assunte per esempio sulla base di una decisione politica che definisca l'ammontare della dinamica premiale. Se l'Ateneo decide di distribuire le risorse con criteri prevalentemente premiali associati alla VQR, adotta valori di α elevati; se invece predilige una distribuzione più *a pioggia* utilizza valori di α piccoli. Infatti, una volta definito tale indicatore US^*_d, l'Ateneo distribuisce le risorse totali *RIS* fra i suoi *NDIP* dipartimenti, assegnando al dipartimento *d* le risorse RIS_d definite dalla seguente relazione che sfrutta l'indicatore in maniera quali-quantitativa[6]:

$$RIS_d = RIS \cdot \frac{US^*_d \cdot N_d}{\sum_{d=1}^{NDIP} US^*_d \cdot N_d} .$$

Quando α=0 il valore di RIS_d dipende esclusivamente dalle dimensioni relative dei dipartimenti, cioè si sta effettuando una distribuzione a pioggia; al contrario, più alto è il valore di α, maggiore è l'influenza della valutazione della qualità sulla distribuzione delle risorse.

---

[6] Una formula analoga in cui si ponga $N_d = 1$ per ogni dipartimento permette di distribuire risorse indipendenti dalle dimensioni dei dipartimenti in base solo alla qualità relativa dei risultati della VQR.

È anche possibile distinguere una politica premiale nei confronti dei dipartimenti che presentano un voto standardizzato positivo, da una politica di penalizzazioni nei confronti dei dipartimenti con voto standardizzato negativo. Per far ciò si può usare l'indicatore USV_d definito come segue:

$$USV_d^* = \begin{cases} 1 + IPR_d^* \cdot \alpha & se\ IPR_d^* \geq 0 \\ 1 + IPR_d^* \cdot \beta & se\ IPR_d^* < 0 \end{cases}$$

dove α ≥ 0 è un parametro che determina la dinamica premiale, mentre β ≥ 0 è un parametro che determina la dinamica delle penalizzazioni. Per esempio, prendendo α > 0 e β = 0 si ottiene una distribuzione delle risorse che non penalizza i dipartimenti sotto media mentre premia (in proporzione ad α) i dipartimenti sopra la media nazionale. Questo metodo, pensando alla assegnazione di punti organico, troverebbe applicazione se l'Ateneo stabilisse una quota aggiuntiva di risorse destinata unicamente ai dipartimenti con valori positivi di IPR^*_d e operasse con i dipartimenti caratterizzati da valori negativi di IPR^*_d raccomandando o in qualche caso (valori di IPR^*_d molto negativi) addirittura imponendo procedure concorsuali realmente selettive, magari affidate a commissari esterni al dipartimento, opportunamente selezionati.

A mero titolo di esempio illustriamo un possibile metodo per determinare α (o per determinare α e β), che non dipende da considerazioni politiche, ma fa piuttosto riferimento diretto al valore di *R*, vale a dire all'indicatore "principe" usato dalla VQR, o più precisamente al valore di *IDVA,* cioè al valore medio di *R* del dipartimento. Infatti, scegliendo

$$\alpha = \frac{\max dip(IDVA) - \min dip(IDVA)}{\max dip(IPR) - \min dip(IPR)},$$

dove max*dip(X)* e min*dip(X)* sono il massimo e il minimo dei valori osservati dell'indicatore *X* sull'insieme dei dipartimenti dell'Ateneo, facciamo in modo che l'indicatore *US* abbia un'escursione, centrata in 1, entro gli stessi limiti di *IDVA.*

In maniera analoga, volendo distinguere la dinamica premiale dalla dinamica di penalizzazione, un modo per scegliere i parametri α e β così da mantenere l'escursione di *USV* entro gli stessi limiti di *IDVA* è il seguente:

$$\alpha = \frac{\max dip(IDVA) - 1}{\max dip(IPR)}, \quad \beta = \frac{\min dip(IDVA) - 1}{\min dip(IPR)}.$$

Infine, sempre sul tema delle politiche di erogazione delle risorse, un Ateneo potrebbe decidere di non adottare alcun criterio premiale o limitativo delle risorse nei confronti di dipartimenti che l'Ateneo considera "normali" ovvero caratterizzati da valori di $P\_inf(U^*_d)$ che si collocano attorno a 0.5 e viceversa utilizzare l'indicatore *IPR* come detto sopra solo per i dipartimenti al di fuori di questo intervallo.

4. L'indicatore di voto standardizzato — le formule

Richiamiamo dapprima alcune proprietà delle somme pesate di variabili aleatorie standardizzate.

Dette u_i delle variabili aleatorie indipendenti e standardizzate, costruiamone la media pesata *Vp* con pesi definiti con criteri "esterni" e poniamoci il problema di come fare poi a rendere tale somma pesata anch'essa una variabile standardizzata. Indicando con p_i il peso associato alla variabile aleatoria u_i, la media pesata *Vp* è data da

$$Vp = \frac{1}{\sum_{i=1}^{N} p_i} \sum_{i=1}^{N} p_i \cdot u_i = \frac{1}{P} \sum_{i=1}^{N} p_i \cdot u_i \ ,$$

dove N è il numero delle variabili aleatorie, e $P = \sum_{i=1}^{N} p_i$. Manifestamente *Vp* non è standardizzata in quanto, pur avendo valore medio 0, la sua varianza non è unitaria ma vale:

$$\sigma^2(Vp) = \frac{1}{\left(\sum_{i=1}^{N} p_i\right)^2} \cdot \sum_{i=1}^{N} p_i^2 = \frac{\sum_{i=1}^{N} p_i^2}{P^2} \ .$$

Per ottenere una variabile *Up* standardizzata occorre pertanto dividere *Vp* per la sua deviazione standard:

$$Up = \frac{Vp}{\sigma(Vp)} = \frac{\frac{1}{P} \sum_{i=1}^{N} p_i \cdot u_i}{\sqrt{\frac{\sum_{i=1}^{N} p_i^2}{P^2}}} = \frac{1}{\sqrt{\sum_{i=1}^{N} p_i^2}} \cdot \sum_{i=1}^{N} p_i \cdot u_i \ .$$

Il caso di media con pesi tutti uguali fornisce immediatamente:

$$U = \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} u_i \ .$$

Veniamo ora al caso di nostro interesse, ovvero la costruzione del voto standardizzato di membro, di *SSD* e di dipartimento, con riferimento alle valutazioni VQR dei prodotti di ricerca. Concettualmente continuiamo a fare riferimento al dipartimento virtuale associato (DVA); ovvero immaginiamo qui e nel seguito di popolare i SSD all'interno del DVA estraendo a caso i membri del DVA dall'insieme dei membri dei SSD nazionali, della cui distribuzione nazionale dei voti abbiamo preventivamente calcolato voto medio e deviazione standard.

Per quanto riguarda i simboli: *s* è l'indice di SSD, *i* conta i prodotti presentati dal singolo membro, identificato dall'indice *l* (elle) che varia fra 1 e N_S (numero membri del SSD nazionale); *i* corre fra *1* e $np_{l,s}$, con $np_{l,s}$ dato da 1 o 2 o 3. Quando occorrerà saranno introdotti gli indici *d* di dipartimento e *a* di ateneo.

Il voto medio $v_{l,s}$ del singolo membro è ottenuto come media, sugli $np_{l,s}$ prodotti che doveva presentare, delle valutazioni $v_{l,s,i}$ dei singoli prodotti[7]:

$$v_{l,s} = \frac{1}{np_{l,s}} \sum_{i=1}^{np_{l,s}} v_{l,s,i} \ .$$

Il voto medio nazionale vale quindi

$$\langle I_s \rangle = \frac{1}{N_s} \sum_{l=1}^{N_s} v_{l,s} \ .$$

[7] La procedura rigorosa per trattare in modo coerente le medie e le deviazioni standard delle votazioni dei membri sarebbe quella di mantenere separati fin dall'inizio, all'interno di ciascun SSD, i membri che dovevano presentare 1, 2 o 3 prodotti, rispettando poi tale specificità anche in fase di popolazione del DVA. Abbiamo verificato che la complicazione dell'approccio non è ripagata da alcun cambiamento significativo dei risultati e quindi abbiamo preferito procedere semplificando la procedura per il calcolo del valor medio e della deviazione standard nazionale.

Si noti che il voto medio nazionale di SSD è stato calcolato tenendo presenti i voti medi di tutti i membri degli SSD appartenenti agli Atenei, senza considerare i membri appartenenti agli Enti di Ricerca o ai consorzi. Inoltre, per omogeneità sono stati considerati solo i prodotti di competenza degli atenei, escludendo quindi gli ulteriori prodotti che i ricercatori associati a Enti di Ricerca hanno dovuto presentare.

Analogamente, la deviazione standard della distribuzione nazionale dei voti medi dei membri del SSD nazionale è la radice quadrata della varianza campionaria delle votazioni medie dei singoli membri del SSD:

$$\sigma_s^2 = \frac{1}{N_s} \sum_{l=1}^{N_s} \left(v_{l,s} - \langle I_s \rangle \right)^2 .$$

La varianza campionaria è calcolata sulle votazioni medie di singolo membro in coerenza con il fatto che il popolamento del DVA avviene immaginando di estrarre a caso i membri del SSD nazionale, e non i singoli prodotti. Ciò al fine di tenere in considerazione le correlazioni, ampiamente osservate e attese, fra i voti dei prodotti dei singoli membri.

Definite queste grandezze possiamo introdurre il voto standardizzato del singolo membro:

$$u_{l,s,d} = \frac{v_{l,s,d} - \langle I_s \rangle}{\sigma_s}.$$

Prima di andare oltre notiamo subito che il voto standardizzato di singolo membro è un indicatore che rimuove molte delle criticità associate alla variabilità del *modus operandi* dei GEV: corregge infatti il diverso voto medio e la diversa larghezza delle distribuzioni prodotte dai valutatori, portandole rispettivamente a 0 (il voto medio) e a 1 (la larghezza della distribuzione, misurata dalla deviazione standard).

Per comporre il voto standardizzato di SSD nel dipartimento, dobbiamo innanzitutto sommare i voti standardizzati, pesandoli in maniera proporzionale al numero di prodotti presentati, seguendo in ciò l'approccio seguito dall'ANVUR, che per i suoi indicatori ha lavorato a livello di prodotti e non di membri. Una volta introdotto questo sistema di pesi possiamo quindi definire, in coerenza con quanto visto prima, il voto standardizzato $U_{s,d}$ del SSD *s* nel dipartimento *d*:

$$U_{s,d} = \frac{1}{\sqrt{\sum_{l=1}^{N_{s,d}} np_{l,s,d}^2}} \sum_{l=1}^{N_{s,d}} u_{l,s,d} \cdot np_{l,s,d}$$

Per inciso notiamo che, costruendo per ogni aggregazione di membri degli indicatori che a loro volta sono voti standardizzati (come abbiamo appena fatto con $U_{s,d}$), rimuoviamo la dipendenza dell'indicatore dalla numerosità dei membri nelle aggregazioni.

Passando poi al voto standardizzato U_d di dipartimento e procedendo in maniera analoga abbiamo:

$$U_d = \frac{1}{\sqrt{\sum_{s=1}^{NSSD} \sum_{l=1}^{N_{s,d}} np_{l,s,d}^2}} \sum_{s=1}^{NSSD} \sum_{l=1}^{N_{s,d}} u_{l,s,d} \cdot np_{l,s,d}$$

che può essere espresso ricorrendo ai voti standardizzati dei SSD al suo interno come segue:

$$U_d = \frac{1}{\sqrt{\sum_{s=1}^{NSSD} \sum_{l=1}^{N_{s,d}} np_{l,s,d}^2}} \sum_{s=1}^{NSSD} U_{s,d} \cdot \sqrt{\sum_{l=1}^{N_{s,d}} np_{l,s,d}^2} .$$

Può essere utile notare che se il numero di prodotti da presentare fosse stato lo stesso per ogni membro di Ateneo le espressioni di sopra si sarebbero semplificate nel modo seguente:

$$U_{s,d} = \frac{1}{\sqrt{N_{s,d}}} \sum_{l=1}^{N_{s,d}} u_{l,s,d}$$
$$U_d = \frac{1}{\sqrt{N_d}} \sum_{s=1}^{NSSD} \sum_{l=1}^{N_{s,d}} u_{l,s,d}$$
$$U_d = \frac{1}{\sqrt{N_d}} \sum_{s=1}^{NSSD} U_{s,d} \cdot \sqrt{N_{s,d}}$$

Un’ultima nota, per altro ovvia, riguardo all’estensione al voto standardizzato di Ateneo. Se applichiamo questa procedura all’insieme di tutti i membri dell’Ateneo otteniamo

$$U_a = \frac{1}{\sqrt{\sum_{d=1}^{NDD} \sum_{s=1}^{NSSD} \sum_{l=1}^{N_{s,d}} np_{l,s,d}^2}} \sum_{d=1}^{NDD} \sum_{s=1}^{NSSD} \sum_{l=1}^{N_{s,d}} u_{l,s,d} \cdot np_{l,s,d} =$$
$$= \frac{1}{\sqrt{\sum_{d=1}^{NDD} \sum_{s=1}^{NSSD} \sum_{l=1}^{N_{s,d}} np_{l,s,d}^2}} \sum_{d=1}^{NDD} U_d \cdot \sqrt{\sum_{s=1}^{NSSD} \sum_{l=1}^{N_{s,d}} np_{l,s,d}^2} ,$$

dove *NDD* è il numero di dipartimenti dell’ateneo.

Ricordiamo infine che il calcolo di U^*_d e quindi di $\{U_d\}$ approda a valori estremamente simili — in alcuni casi identici — a quelli cui si giunge con $\Delta_d = \frac{(IDVA_d - 1)}{\sigma(IDVA_d)}$. La preferenza verso l’utilizzo di U^*_d è motivata dal fatto, già prima accennato, che la rimozione della variabilità delle votazioni dei GEV interviene *ab initio* e prosegue automaticamente ad ogni passo aggregativo di membri (SSD prima, dipartimento poi e finalmente Ateneo, se del caso) cancellando anche l’effetto dimensione; in Δ_d, al contrario, la correzione avviene solo alla fine della combinazione dei contributi ed è conseguentemente meno efficace.