

L'esercizio VQR 2004-2010 e la valutazione dei dipartimenti

- *Gli indicatori della VQR 2004-2010 – breve sommario*
- *Perché il Gruppo di lavoro CRUI ha proposto all'ANVUR un nuovo modo per calcolare indicatori della qualità della ricerca dei dipartimenti (secondo VQR)*
- *L'approccio concettuale proposto e il significato dell'indicatore*
- *Le proprietà matematiche del “Voto Standardizzato”*
- *I risultati dell'indicatore sul sistema universitario italiano*
- *Il Voto Standardizzato e il suo possibile utilizzo per l'erogazione delle risorse non dipendenti dal costo della ricerca*
- *Qualche aggiornamento*

L'esercizio VQR 2004-2010 e la valutazione degli atenei (dette Strutture dall'ANVUR)

La VQR 2004-2010 ha consentito di raccogliere una immane quantità di informazioni che si prestano a diversi livelli di lettura ed analisi.

L'ANVUR nel rapporto Finale ha pubblicato l'indicatore IRFS1 di Struttura (leggi: Ateneo) composto da 7 indicatori legati alla ricerca stabiliti in ambito VQR (i cui valori venivano espressi come % del valore complessivo dell'Area).

Per tale indicatore l'ANVUR ha anche suggerito l'utilizzo di pesi di Area CUN riferibili al "costo della ricerca".

Il 90% della parte premiale del Fondo di Finanziamento Ordinario 2013 è stata assegnata agli Atenei mediante IRFS1 depurato dai costi della ricerca.

Vediamo brevemente i 7 indicatori di Ateneo.

L'esercizio VQR 2004-2010 e la valutazione degli atenei

I sette indicatori di ricerca introdotti dall'ANVUR per gli Atenei con il peso con cui compaiono in IRFS1:

- 1. **Indicatore di qualità della ricerca** (IRAS1, peso maggiore, 0.5);*
- 2. **Indicatore di attrazione risorse** (IRAS 2, peso 0.1)*
- 3. **Indicatore di mobilità** (intesa come reclutamento ndr) (IRAS3, peso 0.1)*
- 4. **Indicatore di internazionalizzazione** (IRAS4, peso 0.1)*
- 5. **Indicatore di alta formazione** (IRAS5, peso 0.1)*
- 6. **Indicatore di risorse proprie** (IRAS6, peso 0.05)*
- 7. **Indicatore di miglioramento** (confronto VQR/VTR) (IRAS7, peso 0.05)*

E per i dipartimenti cosa ha proposto l'ANVUR, non avendo un mandato preciso da parte del bando ministeriale?

L'esercizio VQR 2004-2010 e la valutazione dei dipartimenti

Per la valutazione dei Dipartimenti l'ANVUR ha suggerito, e poi riportato nei rapporti finali dei singoli Atenei, l'indicatore IRFD che riproduce parzialmente, a livello di Dipartimento, l'indicatore di Ateneo IRFS1, utilizzando gli indicatori IRDX che partono dalla stessa informazione degli indicatori IRASY, declinata ove possibile a livello di dipartimento.

La corrispondenza IRASY – IRDX è la seguente:

- 1. IRD1 peso 0.5 ← Indicatore di qualità della ricerca (IRAS1, peso maggiore, 0.5);**
- 2. IRD2 peso 0.2 ← Indicatore di attrazione risorse (IRAS 2, peso 0.1)**
- 3. Indicatore di mobilità (intesa come reclutamento ndr) (IRAS3, peso 0.1)*
- 4. IRD3 peso 0.2 ← Indicatore di internazionalizzazione (IRAS4, peso 0.1)**
- 5. IRD4 ← Indicatore di alta formazione (IRAS5, peso 0.1)**
- 6. Indicatore di risorse proprie (IRAS6, peso 0.05)*
- 7. Indicatore di miglioramento (confronto VQR/VTR) (IRAS7, peso 0.05)*

L'indicatore IRD4, definito nel Rapporto generale non è stato poi utilizzato nell'indicatore di Dipartimento IRFD; conseguentemente i pesi sono stati ri-normalizzati a somma 1. Inoltre nei rapporti di Ateneo sono stati introdotti i costi della ricerca secondo ANVUR.

L'esercizio VQR 2004-2010 e la valutazione dei dipartimenti

La tabella IRFD di UniTus (con inclusi i costi della ricerca stimati dall'ANVUR):



90.2.2 La tabella sinottica conclusiva sugli indicatori di dipartimento legati alla ricerca

Nella Tabella 90.15 sono riportati, per ognuno dei dipartimenti post Legge 240 dell'Università degli Studi della Tuscia, i valori dei tre indicatori IRD legati alla ricerca descritti nella Sezione 5.1, già pesati e sommati sulle Aree e il valore dell'indicatore finale IRFD. La tabella contiene anche la quota di prodotti attesi su l'università di singoli dipartimenti.

Tabella 90.15 Valori delle somme pesate sulle Aree dei tre indicatori di Area VQR legati alla ricerca nei dipartimenti post 240, la percentuale dei prodotti attesi sul totale della struttura e il valore dell'indicatore finale IRFD per l'Università degli Studi della Tuscia.

Dipartimento L.240	IRD1 x w x 100	IRD2 x w x 100	IRD3 x w x 100	% Prodotti attesi sul totale struttura	IRFD x 100
n.d.	1,714	0,000	0,000	3,509	1,476
Economia e Impresa	2,076	0,756	0,000	10,401	2,049
Innovazione nei sistemi biologici, agroalimentari e forestali	14,589	39,494	30,270	18,797	36,611
Istituzioni linguistico-letterarie, comunicazionali e storico-giuridiche dell'Europa	5,610	0,420	0,000	12,782	4,978
Scienze dei beni culturali	4,102	3,057	3,717	8,271	5,869
Scienze e tecnologie per l'agricoltura, le foreste, la natura e l'energia	13,392	18,659	19,804	21,679	24,792
Scienze ecologiche e biologiche	7,104	3,846	31,785	12,281	18,399
Scienze umanistiche, della comunicazione e del turismo	4,952	0,000	4,526	12,281	5,826

L'esercizio VQR, la valutazione dei dipartimenti e la CRUI

Ben prima della conclusione dell'esercizio VQR 2004-2010, la CRUI aveva avviato una riflessione all'interno della sua Commissione Ricerca istituendo un Gruppo (Gruppo CRUI nel seguito) dedicato alla ricerca di metodi condivisi sulla valutazione interna della ricerca.

Tale iniziativa corrispondeva all'esigenza, già presente in molte le università, di sviluppare analisi su vari indicatori a livello di Ateneo, che potessero:

- a) Tener conto del nuovo assetto organizzativo (L.240/2010), riaggregando le informazioni secondo la nuova composizione dei Dipartimenti (livello SSD);*
- b) Utilizzare un metodo condiviso e valido per tutte le Aree con lo scopo di individuare le performance di ciascun dipartimento, qualunque fosse la sua composizione in termini di SSD.*
- c) Consentire di utilizzare (generalmente per una quota circoscritta) tale metodo per la ripartizione interna di risorse o per attribuzione di punti organico a livello di Dipartimento.*

L'esercizio VQR 2004-2010 e la valutazione dei dipartimenti

Il Gruppo CRUI, prima ancora di elaborare una precisa proposta di indicatori aggiuntivi, ha immediatamente ravvisato l'opportunità di contattare l'ANVUR, anticipatamente rispetto all'uscita del Rapporto, per sollecitare un'analisi dei dati sulla ricerca effettuata anche per settore scientifico disciplinare (SSD).

L'ANVUR ha subito risposto positivamente a questa richiesta, tant'è che le tabelle finali del Rapporto effettivamente riportano tale aggregazione.

Il Gruppo CRUI ha poi condiviso al suo interno la necessità di:

- Sviluppare una metodologia di valutazione incentrata esclusivamente sulla valutazione dei "prodotti" della ricerca a livello dipartimentale presentati alla VQR, cercando di operare nella maniera applicabile il più possibile a tutte le realtà scientifiche.*

Per comprendere meglio l'approccio seguito occorre compiere una breve analisi critica dell'indicatore IRFD dell'ANVUR

L'indicatore IRFD di Dipartimento dell'ANVUR-VQR

L'indicatore IRFD ANVUR non è particolarmente adatto per distribuire risorse non dipendenti dal costo della ricerca:

- *Presenza di voci non applicabili a tutte le Aree e/o SSD*
- *Modesta correzione degli indicatori rispetto alle variabilità di Area, spesso molto rappresentate nello stesso dipartimento (terzo indicatore R dell'ANVUR-VQR)*
- *Presenza di pesi riferiti al costo della ricerca nell'indicatore, oltretutto basato sulla combinazione delle valutazioni di Area e non di SSD.*

Su questo ultimo punto, se c'è tempo, torneremo alla fine

Dipartimenti e indicatori VQR

Ci sarebbe un'altra domanda fondamentale da porsi: come si colloca, nel panorama internazionale, un certo SSD nazionale? Nel rapporto finale ANVUR - Parte Terza viene toccato il punto (per le Aree CUN "bibliometriche"); forse esso merita, prima o poi, un approfondimento.

In questo documento tale aspetto non è trattato; ciò non significa che non sia importante, anzi. Il prossimo esercizio VQR dovrebbe affrontare con decisione questo tema.

Di ciò in particolare andrebbe tenuto conto per erogare risorse legate alle attività di ricerca.

Recenti contatti con l'ANVUR ci confortano: il prossimo esercizio VQR dovrebbe sviluppare e approfondire questa aspetto essenziale

Un indicatore non dipendente dal costo della ricerca

*Ogni Ateneo dovrebbe essere interessato ad avere i migliori professori e ricercatori che afferiscono ai vari SSD presenti al suo interno → per distribuire ai dipartimenti risorse non dipendenti dal costo della ricerca (**posti** espressi in punti organico) sarebbe opportuno:*

- 1. Utilizzare esclusivamente la valutazione dei prodotti della ricerca a livello di SSD (informazione più dettagliata rispetto all'AREA CUN)*
- 2. Non utilizzare la capacità di attrarre risorse, di compiere trasferimento tecnologico, ovvero di tutto ciò che non sia applicabile in maniera significativa a tutti i SSD;*

E poi, soprattutto individuare un metodo per confrontare fra di loro e ordinare secondo la qualità della ricerca dipartimenti aventi membri appartenenti a SSD molto diversi e con numerosità differenti.

L'idea di base: il criterio del percentile superiore

L'approccio suggerito si ispira alla classificazione che le università statunitensi ci richiedono quando un nostro ex-studente fa domanda di iscriversi presso di loro:

- in quale “top%” dei nostri voti si colloca lo studente (ovvero in quale percentile superiore della distribuzione dei nostri voti, ovvero qual è la probabilità di osservare voti inferiori rispetto a quello da lui ottenuto)?*

Questa classificazione rimuove la dipendenza esplicita dalla modalità di valutare del docente

Per effettuare la classificazione occorre solo avere conservato la distribuzione dei voti che abbiamo assegnato agli studenti di quell'insegnamento

Classificare votazioni di SSD con i percentili superiori

Con questo stesso metodo possiamo confrontare le valutazioni VQR in due SSD completamente differenti. A seconda del percentile della distribuzione delle votazioni nazionali dei rispettivi SSD nel quale esse si collocano, decidiamo quale delle due valutazioni è migliore o peggiore

Per essere applicato, questo criterio richiede esclusivamente di conoscere le distribuzioni dei voti assegnati dalla VQR ai vari membri dei due SSD considerati (che sono disponibili sul sito ANVUR-VQR)

Utilizzando questo metodo scopriremo subito che la stessa valutazione numerica nell'esercizio VQR 2004-2010 in due SSD distinti si colloca quasi sempre in percentili superiori differenti

Qualcosa di simile accade anche per R, il terzo indicatore VQR

Come classificare i dipartimenti?

E per confrontare dipartimenti, come fare?

Occorre definire un indicatore “collettivo” di dipartimento (fattibile), crescente con i voti dei membri

Purtroppo è però impossibile valutare il percentile in cui esso si colloca

Infatti, per farlo, occorrerebbe disporre di un numero molto ampio di dipartimenti composti in termini di SSD esattamente come il nostro. Ovviamente ciò non si verifica mai e ogni dipartimento costituisce un caso a sé stante

Impossibile davvero?

Come classificare i dipartimenti?

No, perché possiamo ragionare così:

Per ogni dipartimento reale (DR) dell'ateneo definiamo un Dipartimento Virtuale o "Dipartimento Specchio" ad esso associato (DVA nel seguito): questo è un dipartimento ipotetico composto da membri (fittizi) afferenti ai vari SSD del nostro dipartimento reale DR, con la loro stessa numerosità.

Basandoci sulle valutazioni reali riportate nell'esercizio VQR da tutti i membri dei rispettivi SSD, calcoliamo ora tutti i valori possibili che l'indicatore può assumere nel DVA e studiamone la distribuzione.

Facciamo un esempio concreto di questo approccio utilizzando un possibile indicatore (R, il terzo indicatore VQR)

Come classificare i dipartimenti?

Come indicatore utilizziamo l'Indicatore di Dipartimento Virtuale Associato (IDVA), definito come segue:

$$IDVA_d = \frac{1}{N} \sum_{ssd=1}^{NSSD} R_{ssd} \cdot n_{ssd}$$

$$R_{ssd} = \frac{I_{ssd}}{\langle I_{ssd} \rangle}$$

R_{ssd} è il cosiddetto "terzo indicatore" della VQR; rapporto fra il voto medio dei prodotti di ricerca nel SSD del dipartimento rispetto a quello medio nazionale.

IDVA rappresenta il valore di R mediato sui SSD presenti del dipartimento. L'indicatore compensa (solo) per i differenti valori medi diversi nei vari SSD. Come vedremo, ne possiamo trovare di migliori. Va comunque bene per illustrare il metodo.

Un esempio di calcolo di IDVA

$$IDVA_d = \frac{1}{N} \sum_{SSS=1}^{NSSD} R_{SSD} \times n_{SSD}$$

SSD	prodotti	Voto SSD nel Dip	Voto Nazionale	R	R*n
VET/01	90	0.75	0.73	1.027	92.466
VET/03	120	0.81	0.84	0.964	115.714
AGR/19	9	0.56	0.58	0.966	8.690
Totale	219				216.870

$$IDVA = (92.466+115.714+8.690)/219 = 0.990$$

Il criterio per classificare i dipartimenti

*L'indicatore $IDVA_d$ assume per il DR il valore $IDVA^*_d$ (notare *).*

*Nel DVA, l'indicatore può assumere numerosissimi valori $IDVA_d$ (niente *), tutti quelli ottenibili dalle combinazioni dei voti nazionali dei SSD; il calcolo può essere effettuato in modo efficiente con metodi di simulazione (Metodo Montecarlo)*

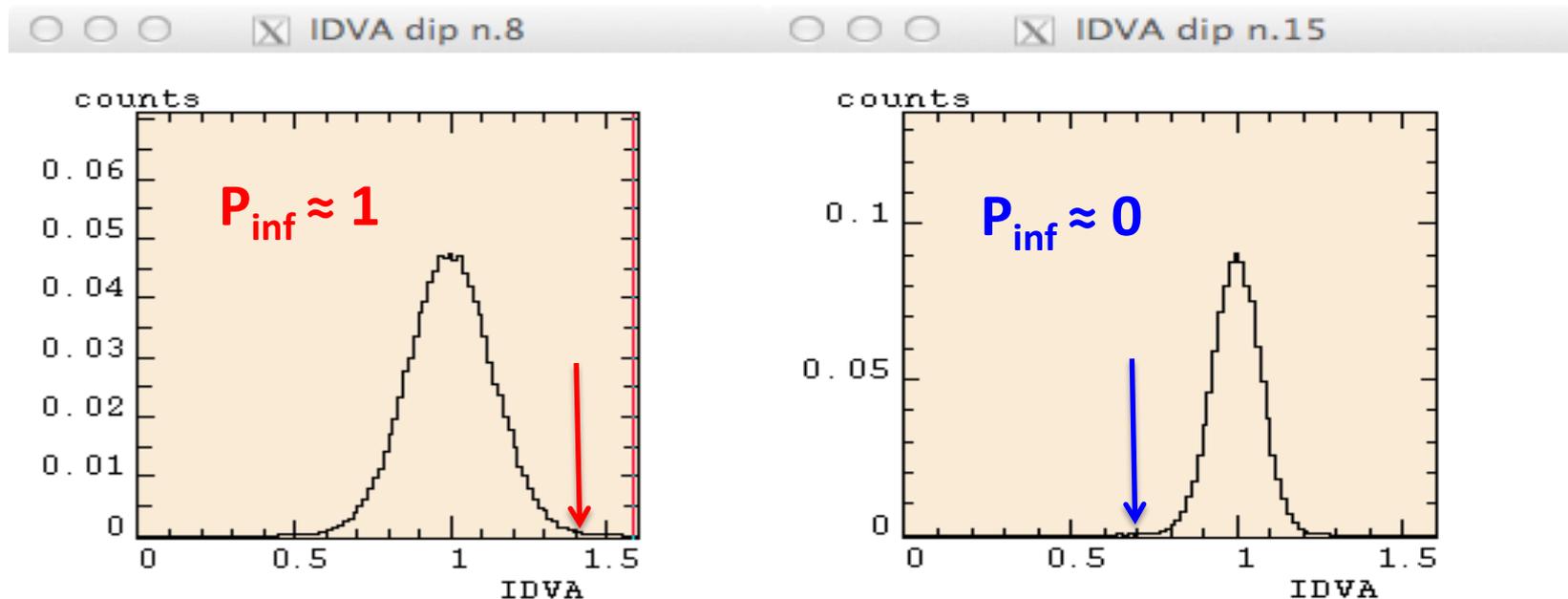
*Contando il numero di casi in cui il valore dell'indicatore $IDVA_d$ è inferiore a $IDVA^*_d$ si determina la probabilità $P_{inf}(IDVA^*_d)$ di ottenere un $IDVA_d$ nel DVA con un valore dell'indicatore inferiore a quello di DR.*

*Sinteticamente possiamo dire che se $P_{inf}(IDVA^*_d)$ è prossimo a 1 (percentile superiore molto piccolo) il DR è molto buono; se viceversa $P_{inf}(IDVA^*_d)$ è prossima 0 (percentile superiore prossimo a 100), il DR è molto modesto.*

Il confronto con il Dipartimento Virtuale Associato

*Si confronta il valore di $IDVA^*_d$ sul DR con la distribuzione di tutti i valori possibili che $IDVA_d$ può assumere nel DVA (Simulazioni Montecarlo).*

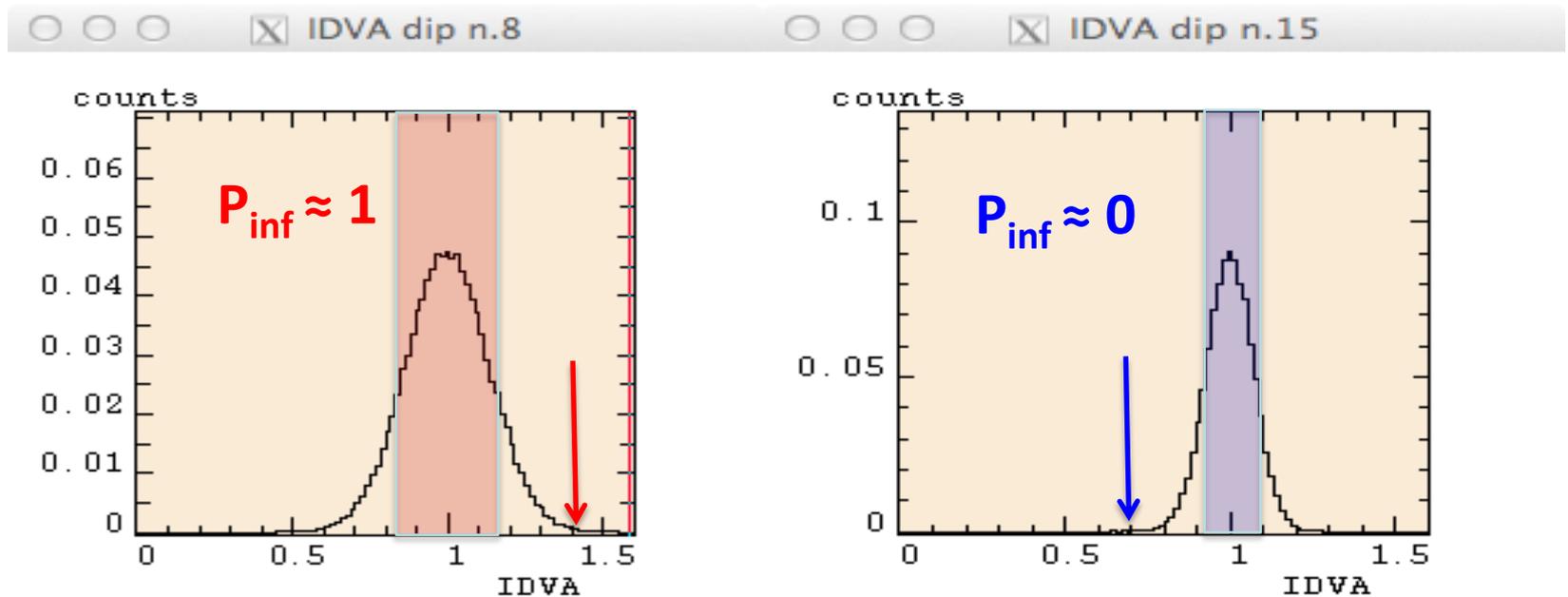
*Un esempio reale: **un dipartimento buono a sinistra** e un **dipartimento modesto a destra**: le frecce indicano i valori di $IDVA^*_d$, cioè i valori dell'indicatore sui DR, gli istogrammi la distribuzione di tutti i valori possibili di $IDVA_d$ sui DVA pe i due casi.*



Nel caso di sinistra praticamente tutti i valori di $IDVA_d$ del DVA sono inferiori a quello reale; l'opposto per il dipartimento di destra. Notare che le due "campane" hanno larghezze diverse.

Il confronto con il Dipartimento Virtuale Associato

In entrambi i casi i valori “normali” sono attorno a 1; le larghezze delle distribuzioni per i due dipartimenti sono assai differenti. Per comprendere se la qualità di ricerca (VQR) di un dipartimento è buona o non buona conta quanto il valore dell’indicatore calcolato per il DR è lontano dai valori “normali” che il DVA può assumere.



Il confronto con il Dipartimento Virtuale Associato

*La conoscenza di $P_{inf}(IDVA^*_d)$ permette di definire un metodo operativo per confrontare la qualità VQR di qualunque insieme di dipartimenti fra di loro.*

*Migliori saranno quelli che sono caratterizzati da valori di $P_{inf}(IDVA^*_d)$ maggiori*

Problemi di applicabilità:

Montecarlo su 850 dipartimenti nazionali – un po' laborioso, possibile, ma poco trasparente (pochi addetti ai lavori utilizzano questo metodo)

La necessità del Montecarlo è tanto maggiore quanto maggiore è la variabilità valutativa dei GEV coinvolti. Da questo punto di vista il terzo indicatore (R e quindi anche IDVA che è un "R mediato sui SSD") lascia a desiderare.

La soluzione proposta per risolvere il problema: il Voto Standardizzato.

http://www.anvur.org/index.php?option=com_content&view=article&id=609&Itemid=563&lang=it9

Il Voto Standardizzato di dipartimento e P_{inf}

Dal campione nazionale completo delle valutazioni del SSD s estraiamo la votazione media $\langle I_s \rangle$ dei soggetti valutati e la deviazione standard campionaria σ_s che misura la “larghezza della distribuzione”.

Esaminando i vari SSD nazionali, si trova che i valori di $\langle I_s \rangle$ e σ_s cambiano fortemente da SSD a SSD.

Spesso piccoli valori di $\langle I_s \rangle$ si accompagnano a grandi valori di σ_s e viceversa.

Definiamo allora, per ciascun membro l del dipartimento d afferente al SSD s il suo voto standardizzato $u_{l,s,d}$:

$$u_{l,s,d} = \frac{V_{l,s,d} - \langle I_s \rangle}{S_s}$$

In maniera un po' semplificata:

il voto standardizzato compensa per i diversi voti medi dei SSD (la differenza a numeratore); compensa parzialmente anche per le diverse larghezze delle distribuzioni dei voti nazionali (rapporto con la deviazione standard campionaria del SSD).

Il Voto Standardizzato di dipartimento e P_{inf}

Mentre le distribuzioni delle valutazioni dei vari SSD hanno diversi valori medi $\langle I_s \rangle$ e diverse deviazioni standard σ_s , le corrispondenti distribuzioni del VS, qualunque sia il SSD esaminato, hanno tutte lo stesso valor medio (0) e la stessa deviazione standard (1) -- per gli esperti: solo i momenti di ordine superiore sono, in genere, differenti.

Altro vantaggio (matematicamente correlato al precedente): partendo dai VS di singolo membro è possibile costruire, semplicemente sommandoli e normalizzandoli, l'indicatore Voto Standardizzato di Dipartimento U_d^ che gode di alcune importanti e utili proprietà.*

Il Voto Standardizzato di dipartimento e P_{inf}

Il VS di dipartimento U_d è definito come la media dei VS di singolo membro, pesata con il numero di prodotti che i vari soggetti valutati dovevano presentare:

$$U_d = \frac{1}{\sqrt{\dot{\bar{a}}_{s=1}^{NSSD} \dot{\bar{a}}_{l=1}^{N_{s,d}} np_{l,s,d}^2}} \dot{\bar{a}}_{s=1}^{NSSD} \dot{\bar{a}}_{l=1}^{N_{s,d}} u_{l,s,d} \times np_{l,s,d}$$

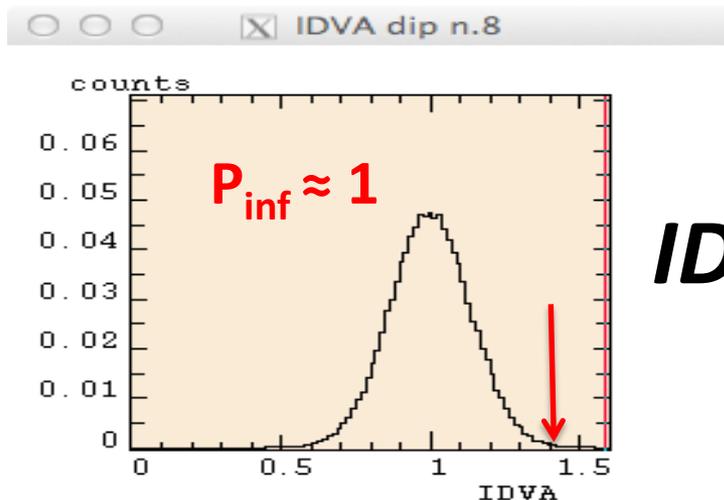
Il denominatore assicura che l'insieme dei valori $\{U_d\}$ calcolati per il DVA abbia valore medio pari a 0 e deviazione standard pari a 1. Assicura inoltre che l'indicatore sia depurato dall'effetto dimensione.

Si può dimostrare che sussistono tutte le condizioni matematiche che assicurano che la distribuzione dei valori di $\{U_d\}$ sul DVA, qualunque sia il DR che stiamo esaminando, sia sempre la stessa: una gaussiana standardizzata.

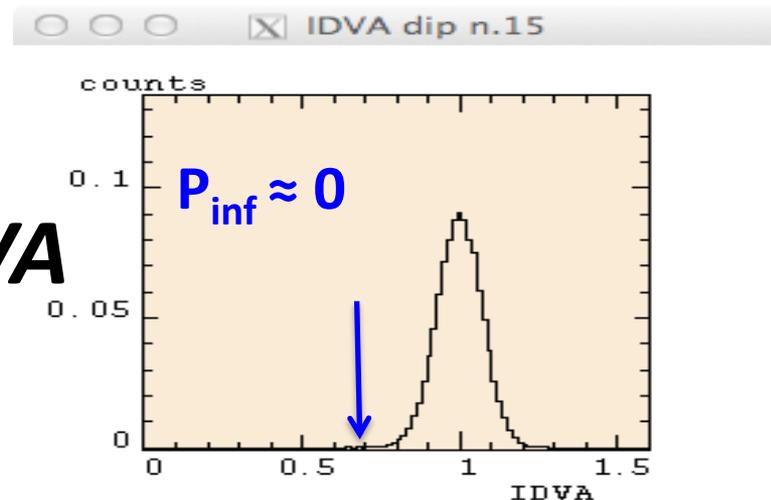
L'uguaglianza di tutte le distribuzioni di $\{U_d\}$ semplifica enormemente il calcolo, altrimenti complesso e laborioso (ricorso al Montecarlo), di $P_{inf}(U_d^)$: il calcolo è lo stesso per tutti i DR.*

Il Voto Standardizzato di dipartimento e P_{inf}

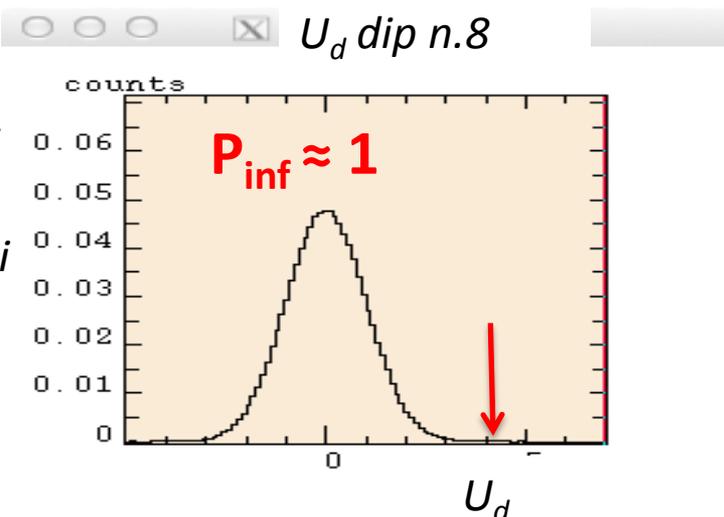
Calcolo di P_{inf} utilizzando $IDVA_d$. Distribuzioni diverse per dipartimenti diversi



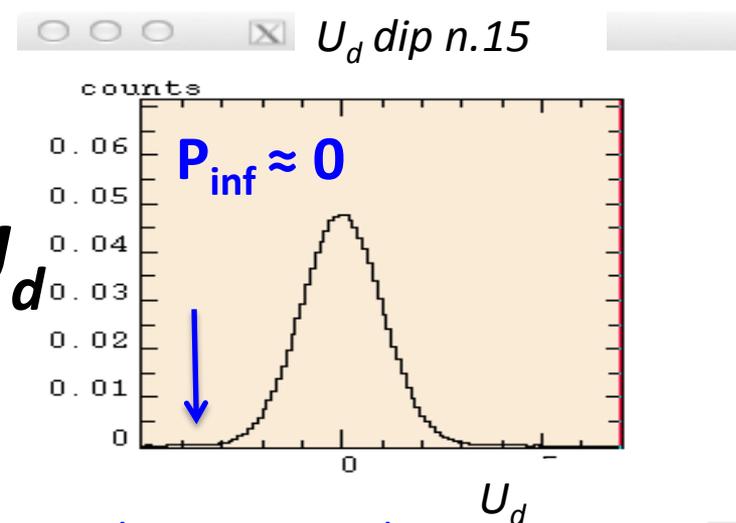
IDVA



Calcolo di P_{inf} utilizzando il VS di dipartimento U_d . Distribuzioni uguali per dipartimenti diversi.

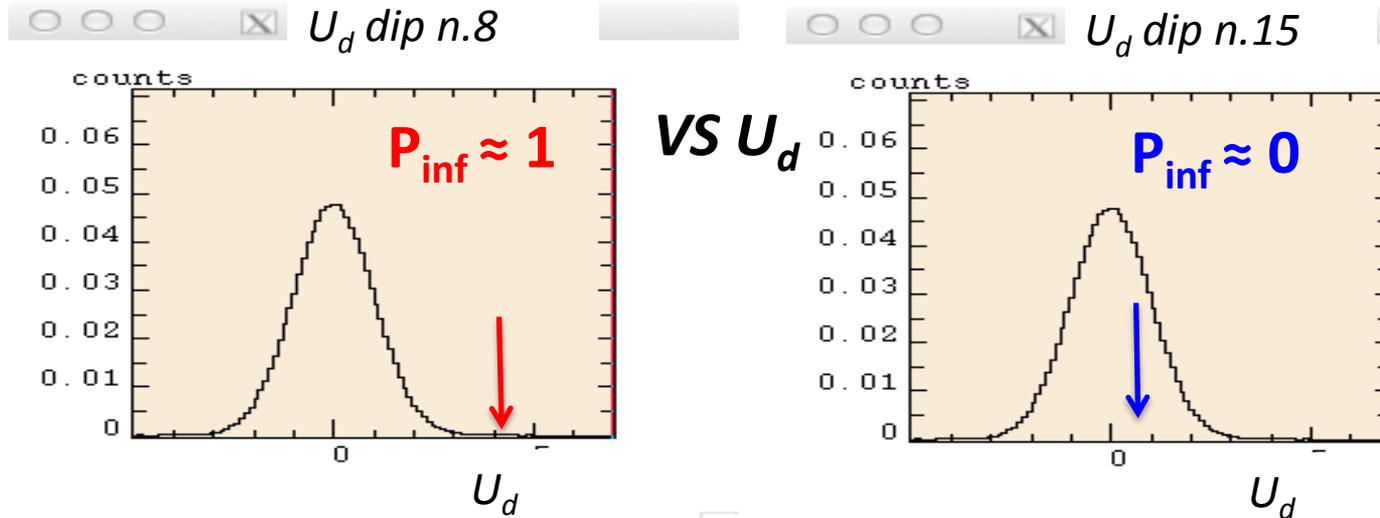


U_d



Notare che ora le distribuzioni di destra e sinistra (VS) sono divenute uguali

Il Voto Standardizzato di dipartimento e P_{inf}



VS U_d

$$P_{inf} = \frac{1}{\sqrt{2\rho}} \int_{-\infty}^{U_d} e^{-\frac{x^2}{2}} dx$$

Calcolo diretto di $P_{inf}(U_d^*)$ per qualunque dipartimento, noto U_d^* : notevole semplificazione. Nessuna necessità di ricorrere al Montecarlo. Basta fare un calcolo da foglio elettronico: $P_{inf}(U_d^*) = \text{DISTRIB.NORM.}(U_d^*, 0, 1, 1)$.

Uguaglianza di tutte le distribuzioni di U_d sui DVA garantita dal Teorema del Limite Centrale (TLC). Rigorosamente ciò vale solo nel caso di un numero infinito di membri del dipartimento.

Studiando i circa 900 dipartimenti nazionali: la massima differenza fra il calcolo che assume la validità del TLC e il calcolo da simulazione Montecarlo ammonta al massimo al 3.5% (un Dipartimento del Politecnico di Milano).

Il Voto Standardizzato di dipartimento e P_{inf}

Limiti noti del VS:

- *La compensazione non perfetta per i diversi voti medi e le diverse larghezze;*
- *Infatti i valori possibili (accessibili) del VS non sono gli stessi per tutti i SSD;*
- *In pratica ciò significa che in certi SSD si possono ottenere valori per il VS di singolo membro inaccessibili per altri SSD (esempio SECS-P/07, 08, 09 contro FIS/04, M-PSI/02);*
- *La presenza di inattivi (-0.5 per ogni prodotto mancante) ha effetti “locali” che diventano sempre più importanti quanto più è alto il voto medio del SSD.*

Tuttavia:

- *L'utilizzo della probabilità $P_{inf}(U_d^*)$ come misuratore della qualità neutralizza in pratica questi effetti spuri residuali nel valore di U_d^* ;*
- *Su altre critiche apparse in rete, ho cercato di rispondere puntualmente, spero con garbo sufficiente (quasi sempre);*
- *Non pochi atenei già utilizzando P_{inf} per distribuire risorse*

Una occhiata alle Tabelle

Le tabelle sono disponibili al sito ANVUR dedicato alla collaborazione ANVUR-CRUI:

http://www.anvur.org/index.php?option=com_content&view=article&id=609&Itemid=563&lang=it

Andamenti generali:

- *distribuzione nazionale U_d^* e presenza di segnale*
- *Compensazione effetto dimensione*
- *Pesanti e preoccupanti differenze Nord /Centro /Sud /Isole*
- *Ogni dipartimento, almeno per i SSD con più di 10 prodotti, può vedere dalle tabelle quali sono i SSD che maggiormente hanno contribuito al risultato finale.*

La distribuzione delle risorse

Una domanda ricorrente: come utilizzare questi indicatori per determinare l'erogazione di risorse non dipendenti dal costo della ricerca?

Molto dipende dal fatto se questi indicatori devono servire a stabilire un fattore saliente in un modello di Ateneo di assegnazione di risorse oppure se l'Ateneo vuole utilizzarli per guidare scelte politiche, riconoscimenti di qualità, definizione di severi criteri di selezione del personale docente, in breve per "azioni politiche" di accompagnamento a modelli già esistenti. Su questo ultimo aspetto non entriamo.

Proviamo a fare un esempio di un possibile utilizzo del Voto Standardizzato ovvero di P_{inf}

La distribuzione delle risorse

Una maniera semplice per utilizzare P_{inf} è quella di definire per ognuno degli ND dipartimenti di ateneo (ognuno avente N_d membri) un peso normalizzato:

$$w_d = \frac{P_{inf,d} \times N_d}{\sum_d^{ND} P_{inf,d} \times N_d}$$

e con esso assegnare al dipartimento una quota $Ris_{prem,d}$ dell'ammontare totale Ris_{prem} di risorse premiali:

$$Ris_{prem,d} = w_d \times Ris_{prem}$$

La distribuzione delle risorse

Il peso w_D è quindi un indicatore quali-quantitativo (come molti di quelli definiti dall'ANVUR). Esso ha quindi la proprietà di

- assegnare risorse premiali in ragione della qualità –ovvero del valore di P_{inf} a parità di dimensione*
- assegnare risorse premiali in ragione della dimensione a parità di P_{inf}*

Si tenga presente che, al fine di guidare le scelte successive in capo ai dipartimenti (per esempio: proposte di bandi sui vari SSD), la pubblicazione dei valori P_{inf} del dipartimento sono accompagnati, nel secondo foglio delle tabelle pubblicate, dal voto standardizzato dei singoli SSD interni al dipartimento (se il numero dei prodotti ≥ 10).

Grazie

(a meno che non ci siano ancora due minuti)

Aggiornamento: il Voto Standardizzato e la selezione dei prodotti

Recentemente il Gruppo CRUI ha richiesto all'ANVUR il calcolo del Voto Standardizzato di dipartimento effettuando alcune selezioni particolari dei prodotti.

Le selezioni richieste, ulteriori rispetto a quelle standard già pubblicate:

- Soggetti reclutati nel periodo 2004-2010*
- Eliminazione soggetti inattivi o parzialmente inattivi*
- “Migliori prodotti” (richiesta da La Sapienza)*
- Inclusione del costo della ricerca calcolato per Area CUN dall'ANVUR, con annesso calcolo dei costi C_d della ricerca associati ai dipartimenti.*

Aggiornamento: il Voto Standardizzato e la selezione dei prodotti

C_d potrebbe essere usato per pesare il P_{inf} per erogare risorse dipendenti dai costi della ricerca, mediante una modifica dei pesi prima introdotti:

$$W_d = \frac{P_{inf,d} \times C_d \times N_d}{\sum_d^{ND} P_{inf,d} \times C_d \times N_d}$$

Potrebbe però anche servire per un'azione complementare, ovvero ...
... per depurare gli indicatori IRFD pubblicati dall'ANVUR dall'effetto "costo della ricerca" e permettere un possibile utilizzo dell'indicatore IRFD di dipartimento per stimare il contributo che ogni singolo dipartimento di Ateneo ha fornito alla quota premiale ricerca dell'FFO2013.

Quando i dati saranno disponibili gli Atenei, se interessati, potranno richiederli alla CRUI stessa.

Grazie

(anche per i due minuti in più)

Verbale della riunione tenutasi il giorno del 07 luglio 2014

Il giorno 7 luglio 2014, alle ore 11.30, presso la sala “ Altiero Spinelli” del Rettorato dell’Università degli Studi della Tuscia (Via S. Maria in Gradi, 4 - Viterbo), si è tenuta una audizione nell’ambito della collaborazione tra ANVUR e CRUI sull’affinamento degli indicatori VQR per la valutazione della "qualità" della ricerca dipartimentale, al fine di illustrare i nuovi indicatori proposti dalla Commissione Ricerca della CRUI e messi a punto, in particolare, dal prof. Giacomo Poggi dell’Università di Firenze.

Partecipano alla riunione il Rettore, il Direttore Generale, i membri della Commissione Ricerca di Ateneo, i membri del Presidio Qualità, la Dott.ssa Paganelli e la Dott.ssa Brenno in rappresentanza della Fondazione CRUI. Il Rettore ricorda l’importanza, per il nostro Ateneo, dell’applicazione di meccanismi valutativi in linea con gli standard nazionali e internazionali, mettendo in risalto la necessità di avere con continuità una fornitura da parte dell’ANVUR di risultati valutativi per facilitare il loro utilizzo anche in fase di programmazione. Il Presidente della CRA afferma l’importanza di conoscere approfonditamente l’approccio seguito dell’ANVUR per valutarne limiti e potenzialità di applicazione. Infine il Presidente del PQA pone in evidenza che per atenei di piccole dimensioni, come il nostro, si ha una scarsa possibilità di utilizzare i risultati della VQR a causa del numero esiguo di ricercatori e professori per alcuni settori scientifici e quindi auspica un miglioramento in tal senso nei meccanismi di fornitura delle informazioni.

Il prof. Poggi illustra nel dettaglio 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 (procedure analoghe possono essere applicate anche aggregando i prodotti per Area CUN). 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. Il prof. Poggi ricorda che il 90% della parte premiale del fondo di finanziamento ordinario 2013 è stato assegnato agli atenei mediante IRFS1 depurato dai costi della ricerca. Si ricorda che l’IRFS1 è composto da 7 indicatori legati alla ricerca ed è riferito all’Ateneo.

Ben prima della conclusione dell’esercizio VQR 2004-2010, la CRUI aveva avviato una riflessione all’interno della sua Commissione ricerca istituendo un gruppo dedicato alla ricerca di metodi condivisi sulla valutazione interna della ricerca. Tale iniziativa corrispondeva all’esigenza già presente in molte università di utilizzare un metodo condiviso e valido per tutte le Aree con lo scopo di individuare le *performance* di ciascun dipartimento, qualunque fosse la sua composizione in SSD, nonché di utilizzare tale metodo per la ripartizione interna delle risorse o per attribuzione di punti organico a livello di Dipartimento.

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 da un numero variabile di membri appartenenti a diversi SSD. Al fine di poter superare queste criticità sono stati introdotti due indicatori il “dipartimento virtuale associato” e l’ “indicatore di voto standardizzato”. Entrambe si rifanno all’approccio che sta alla base del criterio del percentile

superiore. Quest'ultimo si ispira alla classificazione che le università statunitensi richiedono quando un nostro ex-studente fa domanda di iscriversi presso di loro: in quale "top%" dei nostri voti si colloca lo studente (ovvero in quale percentile superiore della distribuzione dei nostri voti, ovvero qual è la probabilità di osservare voti inferiori rispetto a quello da lui ottenuto). Questa classificazione rimuove la dipendenza esplicita dalla modalità di valutazione del docente. Per effettuare la classificazione occorre solo avere conservato la distribuzione dei voti che vengono assegnati agli studenti di quell'insegnamento. Attraverso questo metodo è possibile confrontare le valutazioni VQR in due SSD completamente diversi. A seconda del percentile della distribuzione delle votazioni nazionali dei rispettivi SSD nel quale esse si collocano, è possibile visualizzare quale delle due valutazioni è migliore o peggiore.

Per essere applicato, questo criterio richiede esclusivamente di conoscere le distribuzioni dei voti assegnati dalla VQR ai vari membri dei SSD considerati (che sono disponibili sul sito ANVUR-VQR). Manifestamente questa metodologia di confronto richiede che le valutazioni VQR nei rispettivi SSD siano state compiute con equità e ragionevolezza, 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. Accettato questo principio per il confronto fra membri di SSD differenti, il problema è quello di vedere se è possibile estendere il metodo al caso, indubbiamente più complesso, del confronto fra dipartimenti non mono settoriali. La difficoltà del problema 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 modo omogeneo in termini di SSD e che tutti fossero stati valutati nella VQR. Ovviamente ciò non si verifica mai e ogni dipartimento costituisce un caso a sé stante. Per ovviare a tale problematica, è utile introdurre, per ogni dipartimento dell'ateneo, un Dipartimento Virtuale o "Dipartimento Specchio" ad esso associato (DVA nel seguito): si tratta di un dipartimento ipotetico (ovvero inesistente ma perfettamente definibile in termini operativi) composto da una distribuzione di membri (fittizi) 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 dipartimento reale, la VQR non ha assegnato alcun voto ai membri (anch'essi virtuali) del DVA. Per procedere oltre, occorre quindi individuare un Indicatore 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. Come indicatore è possibile utilizzare l'Indicatore di Dipartimento Virtuale Associato (IDVA), definito come segue:

$$IDVA_d = \frac{1}{N} \sum_{ssd=1}^{N_{SSD}} R_{ssd} \cdot n_{ssd}$$

$$R_{ssd} = \frac{I_{ssd}}{\langle I_{ssd} \rangle}$$

R_{ssd} è il cosiddetto "terzo indicatore" della VQR; rapporto fra il voto medio dei prodotti di ricerca nel SSD del dipartimento rispetto a quello medio nazionale.

IDVA rappresenta il valore di R mediato sui SSD presenti del dipartimento.

Nel DVA, l'indicatore può assumere numerosissimi valori di $IDVA_d$, tutti quelli ottenibili dalle combinazioni dei voti nazionali dei SSD, il calcolo può essere effettuato in modo efficiente con metodi di simulazione (metodo Montecarlo). Tuttavia per questo indicatore esistono problemi di applicabilità poiché è un metodo molto laborioso e poco trasparente, infatti la necessità del metodo Montecarlo è tanto maggiore quanto maggiore è la variabilità valutativa dei GEV coinvolti. Da questo punto di vista quindi l'indicatore R e IDVA lascia a desiderare. La soluzione proposta per risolvere il problema è il voto standardizzato di dipartimento. Se dal campione nazionale completo delle valutazioni del SSD si estrae la votazione media dei soggetti valutati I_s e la deviazione standard campionaria σ_s che misura la larghezza della distribuzione, è possibile definire per ciascun membro l del dipartimento d afferente al SSD s il suo voto standardizzato dato dalla seguente espressione:

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

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. Partendo dai voti standardizzati del singolo membro è possibile costruire semplicemente sommandoli e normalizzandoli, l'indicatore "Voto standardizzato di dipartimento" U_d che gode di alcune importanti ed utili proprietà. Nello specifico il voto standardizzato di dipartimento è definito come la media dei voti standardizzati di singolo membro pesata con il numero di prodotti che i vari soggetti valutati dovevano presentare:

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

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 SSDs nel dipartimento d ; ovviamente nel dipartimento d alcuni SSD saranno privi di membri e per questi $N_{s,d}$ sarà pari a zero. Il denominatore assicura che l'insieme dei valori U_d calcolati per il DVA abbia valore medio pari a zero e deviazione standard pari ad 1. Assicura inoltre che il denominatore sia depurato dall'effetto dimensione. Si può dimostrare inoltre che sussistono tutte le condizioni matematiche che assicurano che la distribuzione dei valori di U_d sul DVA, qualunque sia il DR che si sta esaminando, sia sempre la stessa: una gaussiana standardizzata. L'uguaglianza di tutte le distribuzioni di U_d semplifica enormemente il calcolo altrimenti laborioso e complesso del metodo Montecarlo, infatti il calcolo è lo stesso per tutti i DR.

L'Anvur ha fornito per tutti i dipartimenti nazionali ex legge 240/20110 le tabelle contenenti le informazioni necessarie al calcolo degli indicatori sopradescritti. Si allegano pertanto al presente verbale, quelle relative all'Università della Tuscia ed estrapolate dal sito Anvur sotto la voce "Collaborazione Anvur-CRUI". La tabella, 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 stesso, qualora esso sia chiamato a distribuire al proprio interno risorse assegnategli dall'Ateneo mediante il voto standardizzato di dipartimento. La riunione ha termine alle ore 13.00.

Federica Balletti