



Appendice metodologica

Appendice Metodologica

Premessa

Gli obiettivi dell'indagine e la popolazione di riferimento sono stati descritti nel capitolo I. L'attuale rapporto – come del resto gli studi degli anni precedenti - presenta un grado di complessità non irrilevante per la presenza di dati relativi ad altre indagini simili condotte da altre università italiane, che vengono integrati con l'indagine STELLA. Per tale motivo, di seguito, vengono nuovamente descritte tutte le procedure e le scelte metodologiche adottate per garantire qualità e rigore scientifico a tale importante integrazione.

In riferimento ai **laureati di I livello (definiti triennali negli scorsi rapporti)** questa edizione dell'indagine ha visto una sinergia fra le università del gruppo STELLA e le altre università lombarde (IULM e l'Università Cattolica del Sacro Cuore) interessate ad analizzare comparativamente i dati relativi al placement dei laureati triennali ad un anno dal conseguimento del titolo. La trattazione congiunta di tali dati è presentata nel **secondo capitolo**.

Anche i risultati riguardanti i **laureati magistrali**, data la convergenza nella definizione e nell'identificazione della popolazione obiettivo, sono presentati in un unico capitolo, il **terzo**.

Sia per il secondo sia per il terzo capitolo si è resa però necessaria un'integrazione che ha riguardato alcune informazioni deducibili dai questionari e, laddove necessario, è stata effettuata la normalizzazione delle risposte fornite dagli intervistati.

Il capitolo dei laureati a **ciclo unico**, il **quarto**, riguarda solo i laureati della popolazione degli atenei che aderiscono all'iniziativa interuniversitaria STELLA.

Gli Atenei di Brescia, Milano, Pavia e Sant'Anna di Pisa hanno deliberato di intervistare tutti i propri laureati (effettuando così un censimento). Gli Atenei di Bergamo, Milano-Bicocca, Palermo e Pisa hanno scelto di effettuare un campionamento.

Il disegno campionario

- Modalità di rilevazione

La popolazione di riferimento è costituita dai laureati nel 2011 intervistati a circa un anno dal conseguimento del titolo di studio.

Il gruppo STELLA ha intervistato i laureati in due momenti successivi dell'anno per tenere conto delle diverse sessioni di laurea e mantenere più o meno costante e pari a 12 mesi, l'intervallo di tempo dalla laurea all'intervista.

La modalità di rilevazione adottata è stata CATI.

- Piano di campionamento

Il gruppo STELLA, ad eccezione delle Università degli Studi di Milano, Brescia, Pavia e la Scuola Superiore Sant'Anna che hanno scelto di intervistare tutti i loro laureati, ha adottato un piano di campionamento di tipo stratificato non proporzionale (con estrazione casuale delle unità da ogni strato). Gli strati relativi alle classi di laurea 21 – Scienze e tecnologie chimiche, 25, L-30 – Scienze e tecnologie fisiche e 32, L-35 – Scienze matematiche, mentre per i laureati magistrali alle classi 20/S, LM-17 Fisica, 45/S, LM-40 Matematica, 62/S, LM-54 Scienze chimiche e 81/S Scienze e tecnologie della chimica industriale– le cosiddette Scienze Dure – e tutti quelli con dimensione complessiva inferiore a 30 sono stati censiti.

Le variabili di stratificazione scelte sono: Ateneo, Tipo di Corso, Classe di Laurea e Corso di Studio.

L'obiettivo è stato quello di proporre stime significative anche ai livelli Ateneo, Tipo di Corso, Classe di Laurea e Corso di studio. Le modalità delle variabili di stratificazione rappresentano quindi i domini di studio.

Il numero di strati da cui è stato estratto il campione è determinato dalle modalità congiunte delle variabili di stratificazione introdotte. Il monitoraggio della numerosità degli strati, è stato fondamentale per tenere sotto controllo la significatività dei risultati campionari.

Lo stimatore del totale del carattere y associato a questo piano di campionamento è dato da:

$$\hat{t}_y = \sum_{h=1}^L N_h \bar{y}_h$$

dove

N_h $h=1,2,\dots,L$ sono le dimensioni degli strati nella popolazione

obiettivo ($\sum_{h=1}^L N_h = N$).

\bar{y}_h rappresenta la media campionaria calcolata sulle n_h unità statistiche campionate all'interno dello strato h ed è quindi definita come

$$\bar{y}_h = n_h^{-1} \sum_{j=1}^{n_h} y_{hj}.$$

Possiamo quindi scrivere:

$$\hat{t}_y = \sum_{h=1}^L \sum_{j=1}^{n_h} \tilde{w}_h y_{hj}$$

$$\tilde{w}_h = \frac{N_h}{n_h}$$

Ovviamente, per gli strati censiti i pesi sono pari a 1.

In previsione della possibilità di mancate risposte, e quindi di ottenere un campione effettivo più piccolo di quello programmato, con conseguente diminuzione della qualità dei risultati, è stato estratto per ciascuno strato un campione di riserva di numerosità uguale a quella del campione "obiettivo".

Per gli strati censiti, ovviamente non esiste alcuna possibilità di individuare riserve e quindi le unità non rispondenti non possono essere sostituite. Per strati di modeste dimensioni, inoltre, in alcuni casi le riserve rappresentano semplicemente il complemento alla numerosità complessiva, meno che doppia rispetto alla dimensione complessiva del campione.

Per il piano di campionamento relativo alla popolazione dei laureati triennali della Libera Università di Lingue e Comunicazione si rinvia alle note metodologiche del rapporto Almalaurea.

L'Università Cattolica del Sacro Cuore ha scelto di intervistare tutti i laureati di I livello e specialistici, con la modalità CAWI.

Date le scelte metodologiche effettuate dai singoli atenei, la popolazione base per l'indagine è costituita da **46.585** laureati, di cui il 60,3% composti dai laureati di I livello e il 29,8% da laureati specialistici. I residui, pari al 9,9%, costituiscono la popolazione dei laureati a ciclo unico.

- Tassi di risposta e dimensione del campione

La dimensione del campione teorico risultante dalla procedura descritta è pari a **38.027** laureati, ovvero l'81,6% della popolazione. Il tasso di risposta all'indagine è risultato pari al 70% (per i laureati STELLA si raggiunge il 78,1%) con percentuali vicine all' 80% per gli atenei che hanno applicato la modalità del campionamento per corso di studio (come Bergamo e Milano-Bicocca) (*Tabella VI.1.1.*). Rileviamo, rispetto alla scorsa rilevazione, una riduzione del tasso di risposta.

I tassi di risposta di IULM sono tratti dal sito www.almalaurea.it.

VI.1.1 Tassi di risposta per atenei e corso di studio

STELLA, Cattolica e IULM

Ateneo	LT	LM	CU	Totale
Univ. Studi di Bergamo (***)	94,5%	79,8%	100,0%	88,2%
Univ. Studi di Brescia (**)	77,6%	69,9%	72,1%	74,9%
Univ. Cattolica del S. Cuore	34,0%	51,9%	39,8%	40,5%
Univ. IULM (*)	81,7%	81,9%	-	81,7%
Univ. Studi di Milano-Bicocca (***)	89,9%	80,0%	100,0%	86,6%
Univ. Studi di Milano (**)	74,4%	70,2%	70,0%	72,7%
Univ. Studi di Palermo (***)	86,7%	73,7%	96,3%	83,4%
Univ. di Pisa (***)	88,8%	73,0%	89,6%	82,4%
Univ. Studi di Pavia (**)	75,4%	72,7%	69,3%	73,6%
Sc. Sup. Univ. S.Anna Pisa (**)	66,7%	60,0%	-	65,4%
Totale STELLA	81,2%	73,3%	76,6%	78,1%
TOTALE	55,0%	61,2%	54,6%	69,6%

(*) Dati tratti dal sito Almalaurea

(**) Viene effettuato il censimento dei laureati

(***) Tasso calcolato sulla dimensione del campione

Per una informazione più completa, utile soprattutto per gli atenei che hanno scelto di effettuare il censimento si riporta anche il tasso di contatto, ovvero il numero di laureati effettivamente raggiunti (contattati) - indipendentemente dall'esito dell'intervista stessa - sul numero totale di laureati. Il tasso di contatto, che in alcuni casi raggiunge quasi il 100% dei laureati, è rimasto invariato rispetto allo scorso anno. La tabella successiva mostra quindi quanta parte della popolazione di laureati è stata comunque raggiunta:

VI.1.2 Tassi di contatto per atenei e corso di studio

STELLA

Ateneo	LT	LM	CU	Totale
Univ. Studi di Bergamo	74,4%	86,7%	79,3%	78,3%
Univ. Studi di Brescia (**)	95,9%	94,8%	96,5%	95,7%
Univ. Studi di Milano-Bicocca	71,7%	83,2%	71,4%	75,2%
Univ. Studi di Milano (**)	94,5%	93,4%	94,3%	94,1%
Univ. Studi di Palermo	76,5%	88,0%	70,5%	78,5%
Univ. di Pisa	75,9%	80,9%	73,1%	77,3%
Univ. Studi di Pavia (**)	95,2%	94,4%	95,4%	95,0%
Sc. Sup. Univ. S.Anna Pisa (**)	100,0%	80,0%	-	96,2%
TOTALE	83,0%	88,2%	84,5%	84,7%

(**) Viene effettuato il censimento dei laureati

La normalizzazione e l'uso condiviso dei dati

La raccolta delle informazioni delle varie indagini è avvenuta attraverso l'uso di questionari diversificati a seconda della popolazione di riferimento (alcune università hanno deciso di adottare questionari distinti per i laureati triennali e per i laureati magistrali), della popolazione obiettivo e delle esigenze conoscitive specifiche di ogni singola università. Prima di procedere all'elaborazione dei risultati, per garantire la possibilità di un'elaborazione congiunta, è stato svolto uno studio comparato dei diversi questionari adottati nell'intento di individuare un insieme di domande specifiche che potessero consentire tramite opportune riclassificazioni di ricostruire variabili omogenee e comparabili. A tal fine si è tenuto conto della struttura delle domande (o sequenze di domande) e dei criteri di classificazione delle risposte utilizzati nelle diverse rilevazioni. In alcuni casi, per uniformare i dati raccolti si è proceduto ad una riclassificazione (sia delle domande sia delle risposte).

Le domande integrate si riferiscono alle informazioni base sul gruppo disciplinare, sulla facoltà di appartenenza e sulle strade intraprese dopo il conseguimento del titolo. Per gli occupati, in particolare, si è proceduto a normalizzare le informazioni relative alla tipologia di lavoro svolto (prevalentemente riferite al contratto di lavoro, al settore, all'area funzionale, al numero di addetti nell'azienda). Per quanto riguarda coloro che sono in cerca di lavoro la normalizzazione delle informazioni ha riguardato prevalentemente il tipo di lavoro cercato e la modalità di ricerca del lavoro. Per coloro ancora in fase di formazione la normalizzazione ha puntato a standardizzare le informazioni relative al tipo di percorso formativo scelto.

I risultati del presente rapporto sono frutto del cambiamento di una parte delle domande del questionario effettuato nel 2012 allo scopo di uniformare la rilevazione STELLA (che in questi anni ha seguito un approccio più funzionale alle Università e ai laureati) con quella di Almalaurea (che ha seguito un'ottica istituzionale) e quindi fornire al Ministero dell'Università, Istruzione e Ricerca dati equiparabili e confrontabili. Partendo dalle definizioni usate dall'ISTAT nelle rilevazioni sulle forze lavoro e sull'inserimento professionale dei laureati, sono state inserite nel questionario STELLA alcune domande atte ad individuare e distinguere meglio il collettivo dei laureati occupati che include, in questa nuova ottica, coloro che svolgono anche un'attività in formazione¹. Ciò ha comportato, come è ovvio, un aumento degli occupati e un calo delle altre NFL.

¹ **Laureato occupato secondo l'indagine sull'inserimento professionale dei laureati:** sono occupati coloro che svolgono una qualsiasi attività lavorativa ad esclusione di quanti sono in attività di formazione seppure retribuita. **Laureato occupato secondo l'indagine ISTAT sulle Forze di lavoro (rilevazione continua):** sono considerati occupati tutti coloro che dichiarano di svolgere un'attività, anche di formazione, purché retribuita.

Più specificatamente, al fine di calcolare i tassi di occupazione e disoccupazione in maniera analoga a quanto prodotto da Almalaurea, si è provveduto alla modifica del questionario STELLA includendo le domande per l'individuazione rigida di coloro in cerca di occupazione (domande C2.1a, C2.1b), di coloro che studiano (domande C3.1) ma che potrebbero essere anche occupati (domanda C3.1.a).

La procedura di normalizzazione delle domande e delle relative risposte è stata realizzata prevalentemente mediante la collaborazione del Comitato Scientifico STELLA con la IULM (università aderente al consorzio ALMALAUREA).

La capacità di analisi congiunta del questionario, se necessario, può presentare ampi margini di miglioramento. Con le dovute cautele i risultati pubblicati nel presente volume, pertanto, possono essere trattati congiuntamente con quelli di Almalaurea.

Classificazioni

La variabile *Livello sociale della famiglia*

Il livello sociale è stato ricavato tenendo conto sia della professione svolta che del titolo di studio di entrambi i genitori.

Dopo avere assegnato il livello sociale per ciascuno dei genitori in base al criterio esplicitato nello schema sottostante è stato assegnato come livello della famiglia il livello massimo tra quello assegnato al padre e alla madre.

Il titolo di studio dei genitori è stato trasformato in anni di scolarità totali (8 anni per scuola dell'obbligo, 13 per diploma scuola media superiore, 17 per laurea), con valori che andavano quindi da 16 (entrambi i genitori con la sola scuola dell'obbligo) a 34 (entrambi i genitori laureati).

Professioni dichiarate	Livello assegnato
Imprenditori, dirigenti e legislatori Professioni intellettuali, scientifiche ad elevata specializzazione	Alto
Professioni tecniche Impiegati	Medio
Professioni qualificate Artigiani, operai spec e agricoltori Conduttori impianti e operai semiqualeficati	Basso
Professioni non qualificate Forze Armate Nessuna Non risposta Non voglio dirlo	Non assegnabile

Gli anni di scolarità di entrambi i genitori sono stati utilizzati per attribuire i livelli sociali intermedi o assegnare il livello sociale, secondo lo schema seguente:

Max Livello tra padre e madre	anni di scolarità	Livello assegnato
Alto	26-31	Medio alto
	< 26	Medio
Medio	16	Medio basso
	> 26	Medio alto
Basso	> 16	Medio basso
	>26	Medio
Non assegnabile	27-31	Medio
	16	Basso
	17-25	Medio basso

Indicatori CI per la sintesi delle valutazioni espresse dai laureati

Come nelle passate edizioni dell'indagine STELLA, anche in questo rapporto vengono proposti gli indicatori CI per la sintesi delle valutazioni soggettive espresse dai laureati occupati in merito ai livelli di soddisfazione del lavoro svolto.

Nel questionario le opinioni dei laureati vengono raccolte utilizzando una scala a quattro gradi, con modalità 1="molto", 2="abbastanza", 3="poco", 4="per nulla".

La considerazione che una sintesi di tali valori attraverso la media aritmetica dei punteggi implicherebbe o l'assunzione di equipartitura tra i quattro gradi o l'attribuzione ad ogni modalità di risposta di punteggi scelti più o meno arbitrariamente, unitamente ad altre più strettamente attinenti al problema della misurazione, ha condotto all'utilizzo di **indici basati sulle distribuzioni osservate delle risposte**.

Questa famiglia di indici assegna un punteggio numerico di sintesi alla dimensione valutata da parte di ogni gruppo d'analisi. Esso è ottenuto come somma algebrica di due indici; il primo esprime il punteggio conseguito nel semipiano delle valutazioni positive (modalità 1 e 2), il secondo, il punteggio conseguito nel semipiano delle valutazioni negative (modalità 3 e 4). Essi assumeranno pertanto valori compresi tra -100 (quando tutte le risposte sono concentrate nella modalità 4="per nulla" e, quindi nel caso di valutazione massimamente negativa) a +100 (quando tutte le risposte sono concentrate nella modalità 1="molto", e, quindi, nel caso di valutazione massima positiva).

Al generico gruppo di laureati i , in conseguenza della distribuzione delle risposte da essi fornite dai laureati all'item h_i , corrisponde la distribuzione $N_i^h(1) N_i^h(2) N_i^h(3) N_i^h(4)$ delle frequenze associate ai quattro gradi della scala. Si indichino:

- x_i^h la percentuale di giudizi positivi per l'item h espressi dai laureati appartenenti al gruppo i :

$$x_i^h = \frac{(N_i^h(1) + N_i^h(2)) \times 100}{\sum_{j=1}^4 N_i^h(j)}$$

- y_i^h la percentuale di giudizi molto positivi per l'item h calcolati sul totale dei giudizi positivi espressi dai laureati appartenenti al gruppo i :

$$y_i^h = \frac{N_i^h(1) \times 100}{(N_i^h(1) + N_i^h(2))}$$

- x_i^h la percentuale di giudizi negativi sempre per l'item h espressi dai laureati appartenenti al gruppo i :

$$x_i^h = \frac{(N_i^h(3) + N_i^h(4)) \times 100}{\sum_{j=1}^4 N_i^h(j)} = 100 - x_i^h$$

- y_i^h la percentuale di valutazioni molto negative sul totale delle valutazioni negative:

$$y_i^h = \frac{N_i^h(3) \times 100}{(N_i^h(3) + N_i^h(4))}$$

Con riferimento all'item h , nel versante delle valutazioni positive, l'unità i è rappresentata dal punto $P_i^h(x_i^h, y_i^h)$ appartenente al quadrato di lato 100 situato nel primo quadrante (spazio delle valutazioni positive); nel versante delle valutazioni negative, invece, la stessa unità i è rappresentata dal punto $Q_i^h(x_i^h, y_i^h)$ appartenente al quadrato di lato 100 situato nel terzo quadrante (spazio delle valutazioni negative). Il posizionamento dei singoli gruppi d'analisi all'interno delle due aree fornisce una visualizzazione immediata del posizionamento di ogni gruppo. Come la figura sottostante evidenzia immediatamente, al posizionamento del punto $P_i^h(x_i^h, y_i^h)$ sul vertice superiore destro corrisponde un gruppo i di laureati che hanno risposto tutti "molto" all'item h e quindi che hanno espresso la valutazione massima.

Analogamente, al posizionamento del punto $P_i^h(x_i^h, y_i^h)$ nell'origine degli assi corrisponderà un gruppo i di laureati che hanno risposto solo "poco" e/o "per nulla" e quindi il minimo giudizio.

Per qualunque altra situazione, stabilire se al gruppo 1 è associato un giudizio più positivo che al gruppo 2 (e quindi un valore numerico dell'indicatore per il gruppo 1 maggiore di quello del gruppo 2) dipende dal giudizio di valore dello "sperimentatore/decision maker", dipende cioè dal grado di importanza che egli vuole attribuire alla quota di giudizi molto positivi (e, analogamente, molto negativi). Sulla base di queste considerazioni, si è costruito, con riferimento al quadrante delle valutazioni positive, l'indice CI_i^{h+} , avente la seguente espressione:

$$CI_i^{h+} = \frac{100(x_i^h + ky_i^h)}{\max(CI_i^{h+})}$$

$$\max(CI_i^{h+}) = 100(1 + k).$$

Del tutto analogamente si procede per il quadrante negativo definendo l'indice:

$$CI_i^{h-} = \frac{-100(x_i^h + ky_i^h)}{\max(CI_i^{h-})}$$

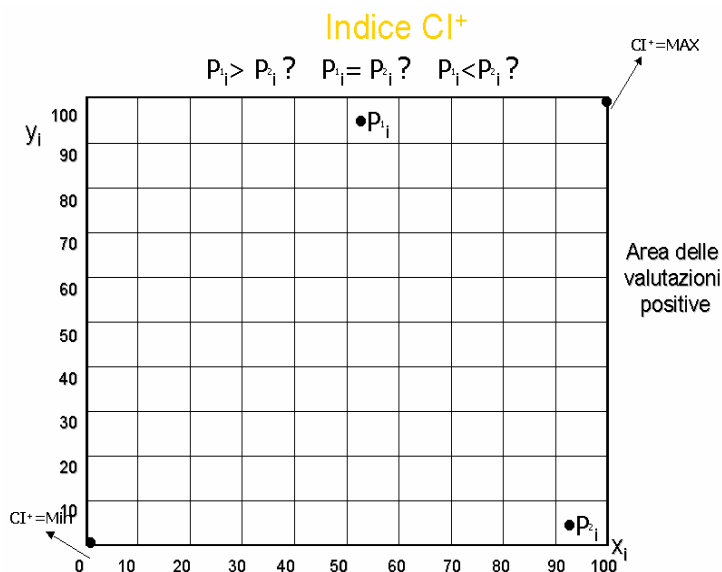
dove $\max(CI_i^{h-}) = 100(1+k)$

Si arriva quindi all'indice:

$$CI_i^h = CI_i^{h+} + CI_i^{h-}$$

con $-100 \leq CI_i^h \leq 100$.

$0 \leq k \leq 1$ è il parametro, scelto dalla "sperimentatore", che esprime appunto il livello di importanza da attribuire ai giudizi "molto positivi".



La scelta di $k=0$, equivale a ridurre la scala a quattro gradi a una scala dicotomica e, quindi, a scegliere di non attribuire alcun peso alle risposte "molto" e "per nulla". In questa applicazione si è scelto $k=0,5$, che, con riferimento al quadrante delle valutazioni positive, implica il considerare equivalenti ad esempio il livello di coerenza di un primo gruppo di laureati che hanno tutti risposto "abbastanza" con quello di un secondo gruppo in cui solo il 50% ha risposto "molto" e nessuno "abbastanza". Il valore dell'indice dei due gruppi, ovviamente, differirà in dipendenza di come, nel secondo gruppo, ha risposto il restante 50% (che ha fornito risposte con modalità negative).

È opportuno richiamare l'attenzione sul fatto che per confrontare correttamente i diversi gruppi sulla base dei valori assunti dall'indice CI è necessario che per ciascuno di essi la numerosità di rispondenti sia molto simile o, quantomeno, che non ci siano gruppi con un numero di rispondenti molto piccolo.

In questo contesto, poiché il calcolo degli indici è stato condotto sulle distribuzioni di risposte riportate all'universo, il problema su accennato non assume particolare rilevanza.

Nel volume i risultati degli indici di sintesi sono stati rappresentati graficamente suddividendo, per ogni aspetto del lavoro considerato, l'area dei giudizi positivi e l'area dei giudizi negativi.

NOTA

Nel rapporto molte tabelle presentano le distribuzioni di frequenza di un carattere condizionate alle modalità di un altro.

Qualora la dimensione della sottopopolazione intervistata e individuata dal condizionamento sia inferiore a 30, i valori delle percentuali sono stati scritti in grigio per evidenziare la cautela necessaria nell'interpretazione di stime soggette ad un'elevata variabilità campionaria.

In alcune tabelle del presente volume, inoltre, i totali, sia nei valori assoluti sia nei valori percentuali, potrebbero non corrispondere alla somma dei singoli valori pubblicati o, nel secondo caso, potrebbero non essere pari a 100. Ciò non costituisce un errore ma è semplicemente dovuto alle necessarie approssimazioni a valori interi dei risultati del coefficiente di espansione calcolato per ottenere la stima del totale della popolazione indagata.