

Analisi della transizione università-lavoro in Italia: obiettivi e prospettive dei laureati in Scienze Statistiche

Domenica Fioredistella Iezzi¹

*Dipartimento di Statistica, Probabilità e Statistiche Applicate
Università “La Sapienza” - Roma*

Riassunto. L’attuale mercato del lavoro richiede che un laureato possieda non solo le conoscenze e le competenze del titolo acquisito, ma anche una serie di abilità di tipo trasversale. In questo contesto, diventa rilevante valutare l’efficacia esterna dei processi formativi dei diversi corsi di laurea. Attualmente molte indagini si occupano di misurare la situazione occupazionale e l’ingresso nella vita lavorativa delle nuove leve della forza lavoro. L’obiettivo di questo lavoro è di esaminare la transizione-università lavoro dei neo-laureati italiani, con particolare riferimento al caso di Scienze Statistiche. A questo scopo si utilizzano in maniera congiunta diverse tecniche di analisi multidimensionale (cluster analysis, scaling multidimensionale e regressione logistica multinomiale).

Parole chiave: transizione università-lavoro, cluster analysis (CA), scaling multidimensionale (MDS), regressione logistica multinomiale (RLM).

1. Premessa

Il mercato del lavoro vive attualmente una rapida e radicale trasformazione in conseguenza del passaggio dai tradizionali particolarismi regionali e nazionali ad un contesto di ampio respiro internazionale. L’autonomia didattica degli Atenei avviata con il D.M. 3 novembre 1999 n. 509 è un processo aperto che ha come obiettivo quello dell’“affermazione dello spazio europeo dell’istruzione superiore” (Lisbona, 2000). L’obiettivo sembra essere ancora lontano per l’Italia, in quanto molti aspetti critici, come gli alti tassi di abbandono, il basso numero di studenti attivi, i tempi eccessivamente lunghi per il conseguimento dei titoli di studio, la bassa mobilità nazionale ed internazionale degli studenti e dei docenti, la mancanza di incontro tra for-

¹ Il presente lavoro è stato finanziato nell’ambito del progetto “OUTCOMES. Modelli di analisi della transizione università-lavoro”, cofinanziato dal MIUR. Coordinatore nazionale è L. Fabbris.

mazione e fabbisogni del mercato del lavoro rendono il nostro sistema universitario spesso inadeguato per un immediato inserimento professionale (CNVSU-MIUR, 2004; Iezzi D.F., 2003). Il cammino verso la “globalizzazione” della formazione ha avuto come suo indiscusso nocchiero *Internet*, il cui utilizzo è ampiamente diffuso tra i giovani neolaureati che si confrontano per la prima volta con il mondo del lavoro. Alla fine di un percorso di studi molti laureati in cerca di occupazione inviano i *curricula* a banche dati on-line o creano pagine web personalizzate. Il neo dottore è cosciente che per avere opportunità di lavoro non solo deve dimostrare di possedere conoscenze e competenze tipiche del proprio corso di laurea (CDL), ma soprattutto di essere rapido nell’acquisire informazioni e flessibile nell’apprendere nuovi sistemi.

Le rilevazioni condotte sui laureati e sul mondo del lavoro sono in numero elevato e vi sono significative esperienze sia a livello locale che nazionale.

Tra i principali studi si devono ricordare le indagini sui laureati svolte da:

- AlmaLaurea (<http://www.almalaurea.it>);
- CampusOne (<http://www.campusone.it>);
- Diogene(http://www.diogenet.net/vbd_index.jsp);
- Excelsior (<http://excelsior.unioncamere.net>);
- Istituto Tagliacarne (<http://www.starnet.unioncamere.it>);
- ISTAT, Rilevazione Trimestrale delle Forze Lavoro e l’inserimento professionale dei laureati (<http://www.istat.it>);
- Università degli studi di Firenze, Valutazione della didattica (http://www.unifi.it/aut_dida/indexval.html);
- Università degli studi Padova, Sportello per le Professioni, Osservatorio PHAROS - *Pursuing Home-market Accessibility and Raise of Occupational Standing* (<http://www.unipd.it/sportelloprofessioni/>);
- VULCANO (<http://www.cilea.it>);

L’obiettivo di questo lavoro è di descrivere la situazione occupazionale dei neolaureati, in Italia, con particolare riferimento a quella degli statistici e di rispondere ai seguenti interrogativi:

- i) chi sono i laureati in scienze statistiche;
- ii) quali sono le professioni che svolgono a tre anni dalla laurea;
- iii) quali sono le potenzialità che stanno sviluppando.

L’interesse verso questo piccolo gruppo di professionisti è mosso dal fatto che, pur avendo i laureati in statistica ottime opportunità di lavoro, il numero di immatricolati diluisce di anno in anno. In questo studio, si utilizzano le ultime informazioni raccolte dall’indagine ISTAT sull’inserimento professionale dei laureati e dalla rilevazione CNVSU-MIUR. Il campione ISTAT è composto da 20.000 laureati che hanno conseguito il diploma di laurea nel corso dell’anno 1999 in tutte le sedi universitarie italiane. Le variabili oggetto di studio sono 166 e riguardano i profili dei neolaureati, i percorsi lavorativi e di formazione intrapresi nei tre anni successivi al conseguimento della laurea. Il dilatarsi dei tempi di ingresso nel mercato del lavoro

evidenzia i limiti di un'indagine circoscritta ai tre anni successivi alla conclusione degli studi. Il passaggio ad un intervallo temporale di osservazione più ampio permetterebbe, infatti, di esaminare non solo la prima fase di impatto con il mercato del lavoro, ma anche i cambiamenti successivi in termini di condizione e qualifica professionale, nonché gli ulteriori percorsi di formazione post-laurea, spesso obbligatori per chi intende esercitare una professione regolamentata. Fra il terzo ed il quinto anno, infatti, i laureati del gruppo giuridico e del gruppo medico vedono crescere la loro quota di occupati in modo consistente. Rimane ancora elevata la quota di laureati del gruppo medico che, a cinque anni, prosegue la formazione post-laurea, causa soprattutto la lunga durata delle specializzazioni e la non immediata consequenzialità fra conseguimento della laurea ed ammissione alla specializzazione. Similmente, e ciò si applica a tutte le tipologie di laureati, un posto di lavoro stabile o una qualifica professionale adeguata al titolo di studio si ottiene soltanto dopo qualche anno di esperienza lavorativa.

2. Il profilo dei neo laureati

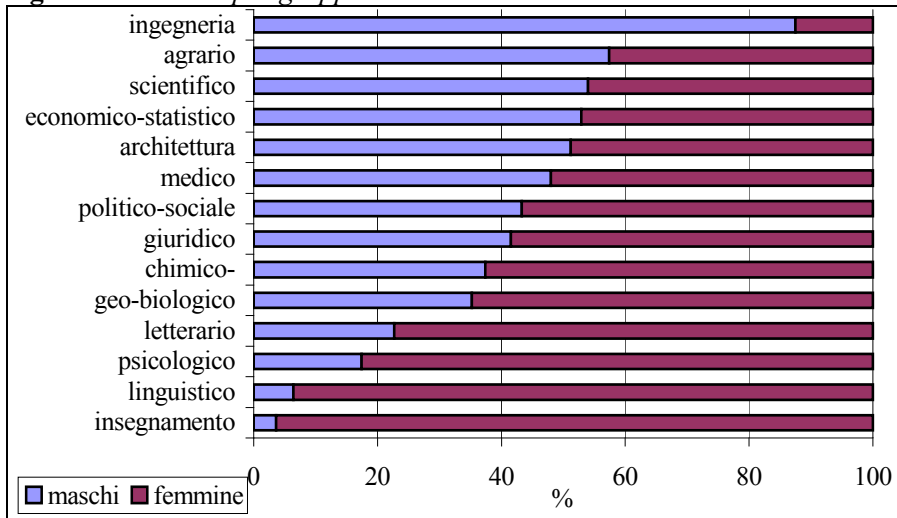
In base all'indagine ISTAT sull'inserimento professionale dei laureati, si rileva che, in Italia, per conseguire la laurea si impiegano in media 7,6 anni, ottenendo nel 19,2% dei casi il massimo della votazione (110 e lode). A tre anni dal conseguimento del titolo il 63,2% dei laureati svolge un lavoro continuativo (iniziato dopo la laurea) ed il 45,9% degli occupati ha un impiego per il quale era richiesta la laurea. Il guadagno mensile medio netto di chi lavora in modo continuativo (a tempo pieno) è di poco superiore ai 1.190 euro.

L'analisi della distribuzione per sesso e per macrogruppi di CDL² mostra una presenza femminile prevalente: le donne con il 55,5% di laureate superano i colleghi maschi di oltre 10 punti percentuali. La fig. 1 evidenzia che alcuni gruppi di CDL sono stati frequentati prevalentemente da uomini (ingegneria 85,56%); mentre altri da donne (insegnamento 90,48%, letterario 91,90% e psicologico 82,33%).

Se si considera il grafico a dispersione (fig. 2) dei laureati per genere (in percentuale sul totale), è evidente una disposizione lineare dei vari gruppi. Il baricentro (39,58; 59,35) si trova in prossimità dei CDL chimico-farmaceutico, geo-biologico, giuridico e politico-sociale. Al di sotto del baricentro vi è una marcata presenza maschile, mentre al di sopra una femminile. Le neolaureate, pur essendo in numero maggiore, hanno una ridotta presenza nelle discipline più richieste dal mercato del lavoro.

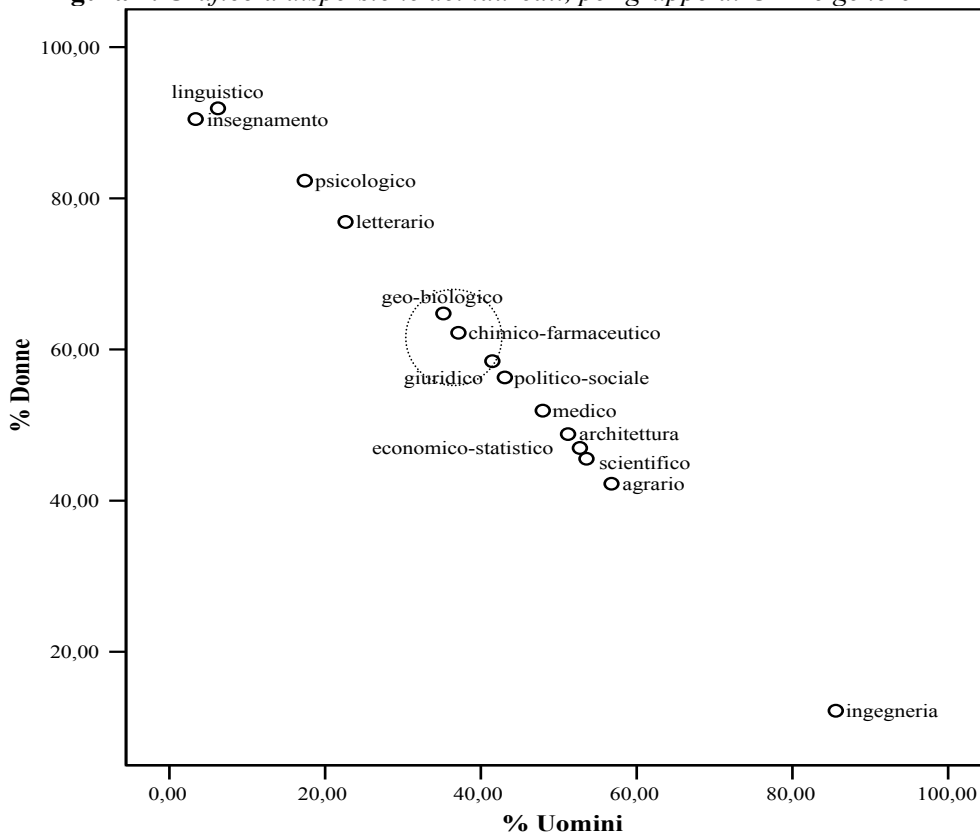
² L'ISTAT aggrega i corsi in 14 macrogruppi: scientifico, chimico-farmaceutico, geo-biologico, medico, ingegneria, architettura, agrario, economico-statistico, politico-sociale, giuridico, letterario, linguistico, insegnamento e psicologico.

Figura 1. Laureati per gruppo di CDL e sesso



Fonte: Elaborazioni proprie su dati ISTAT

Figura 2. Grafico a dispersione dei laureati, per gruppo di CDL e genere



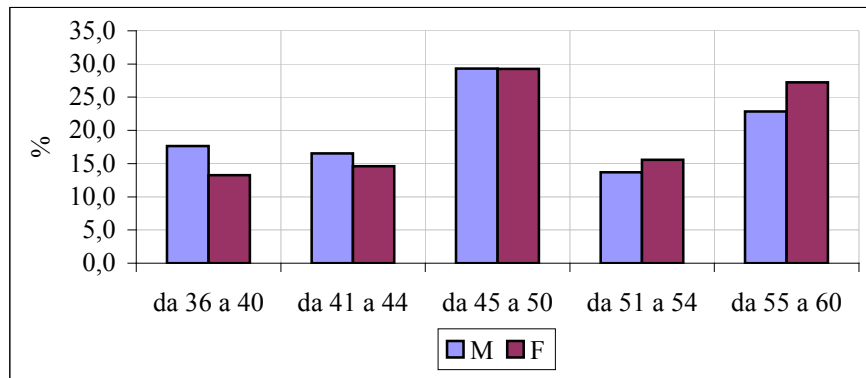
Fonte: Elaborazioni proprie su dati ISTAT

Vi è una significativa associazione tra il diploma conseguito e le scelte universitarie (Chi quadrato di Pearson $p < 0,000$): una forte attrazione nei confronti di coloro che provengono dal liceo artistico o dall'istituto d'arte, si riscontra nel gruppo di architettura, che ha raccolto anche una quota rilevante di diplomati degli istituti tecnici; quasi il 50%, invece, degli studenti che hanno conseguito la laurea nel gruppo insegnamento è in possesso del diploma magistrale (64%). La maturità classica è molto diffusa nei gruppi giuridico (30,1%), letterario (21,6%) e medico (9,7%); una importante quota di diplomati del liceo scientifico è presente in quasi tutti i gruppi di laurea e, in particolare, nei gruppi economico-statistico (19,2%), ingegneria (17,4%) e geo-biologico (6,0%). Nel gruppo economico-statistico la maggior parte dei laureati proviene da un istituto tecnico commerciale (52,4%); una maturità, quella tecnica, molto diffusa anche nei gruppi agrario (15,9%) e politico-sociale (11,7%). Appare evidente come le possibilità di riuscita all'università dipendano anche dalla scuola superiore frequentata e, quindi, la preparazione di base rimane un valido parametro per stimare il tasso di successo³ degli studenti⁴.

La media del voto di maturità conseguito dai laureati italiani è di quasi 47/60, con variazioni più o meno significative in relazione al sesso e al tipo di diploma. La distribuzione per classi di voto evidenzia come complessivamente il 15,2% dei laureati si collochi nella classe dei voti più bassi (36-40); la metà degli studenti non vada oltre i 48/60, mentre 1/4 superi i 55/60, classe, questa, che comprende anche il voto con la massima frequenza 60/60, ottenuto in media dal 12% dei laureati. Il tasso di successo è particolarmente basso nei gruppi di CDL geo-biologico e giuridico (rispettivamente 36% e 37%). Le donne presentano rendimenti migliori anche prima dell'iscrizione all'università. Infatti, se si considerano cinque classi di voto di diploma è possibile rilevare come la quota di donne nei primi due gruppi, quelli che presentano i voti più bassi, è sempre inferiore a quella degli uomini, e ciò è vero fino al

³ Quando si utilizza il tasso di successo nel conseguimento della laurea bisogna considerare che l'indicatore sovrastima le possibilità di riuscita nei CDL che prevedono il numero di studenti programmato, registrano molti trasferimenti in entrata in anni successivi al primo, mentre sottostima le possibilità di riuscita per i gruppi che registrano molti trasferimenti in uscita. Se dunque si esclude il gruppo architettura, il cui successo dipende dall'introduzione del numero programmato a partire dal 1993/94, il gruppo medico si distingue da tutti gli altri per l'alta percentuale di esiti positivi (soltanto 12 immatricolati su 100 abbandonano gli studi).

⁴ Dalla rilevazione MIUR-CNVSU del 2004 risulta che coloro i quali provengono da un liceo ottengono i risultati migliori: su 100 immatricolati, 62 riescono a laurearsi; invece le maggiori difficoltà le incontrano i ragazzi che conseguono un diploma tecnico o professionale. Sul totale dei laureati, una quota superiore ai due terzi risulta in possesso di licenza liceale (37,1% liceo scientifico, 20,6% liceo classico, 4,2% liceo linguistico e 1,3% liceo artistico), poco meno di un terzo proviene dagli istituti tecnici (15,4% istituto tecnico commerciale, 3,6% istituto tecnico per geometri, 5,1% istituto tecnico industriale e 3,1% altri istituti tecnici), il 6,6 % dagli istituti magistrali o dalla scuola magistrale e il 2,7% da quelli professionali, mentre il restante 0,3% è costituito quasi interamente dai diplomati dell'istituto d'arte.

Figura La distribuzione dei voti al diploma per genere

Fonte: Elaborazioni proprie su dati ISTAT

gruppo 45-50, quando si ha un'inversione di tendenza (fig. 3). Variazioni consistenti del voto si hanno in relazione ai diversi tipi di diploma, valori più alti si registrano tra i diplomati degli istituti d'arte (51,3) e commerciali (50,1), mentre i diplomati delle scuole o degli istituti magistrali conseguono mediamente voti più bassi, nei licei il voto medio oscilla tra il 47 al 49.

La votazione media alla laurea è di 103,1, anche se le donne mediamente ottengono un punteggio di 104,1 contro 101,1 degli uomini. L'analisi dei punteggi delle votazioni di laurea conferma la difformità dei criteri di valutazione adottati nelle diverse facoltà; differenza che diventa un fattore inevitabile di discriminazione nella società, soprattutto laddove la votazione acquisita si traduca in punteggi determinanti per la formazione di graduatorie. La votazione media alla laurea arriva appena a 100,2 tra i laureati del gruppo giuridico e a 100,9 tra quelli del gruppo economico-statistico, mentre i loro colleghi del gruppo letterario raggiungono quasi il punteggio di 108. Questa prima analisi induce erroneamente a considerare che le donne mediamente ottengono un punteggio elevato, poiché scelgono percorsi ove le valutazioni sono naturalmente più alte. Il confronto tra laureate e laureati, invece, evidenzia come le donne tendano ad ottenere voti di laurea più alti in tutti i gruppi di CDL; infatti, in media superano il voto di 104 poco meno del 60% delle laureate rispetto al 42% dei laureati. Tra i laureati che si collocano nella classe di voti più bassa (inferiori a 100) i maschi risultano oltre il 15% in più. Inoltre, ottiene la lode il 24,7% delle donne contro il 16,6% degli uomini.

La regolarità degli studi risulta molto bassa, poiché solo l'8,3% dei laureati ha concluso gli studi entro i tempi previsti dall'ordinamento universitario, mentre il restante 91,7% si è laureato fuori corso. Le donne, con nove laureate in corso su 100, risentono del problema in misura più contenuta degli uomini. Sul dato complessivo incide la diversa distribuzione dei due sessi nei gruppi di corsi a più elevata irregola-

rità e l'obbligo di leva. Una buona parte dei laureati maschi ha svolto il servizio di leva durante gli studi universitari e questo porta ad una dilatazione della durata degli studi per gli stessi.

Si hanno consistenti differenziazioni tra i CDL: la percentuale più elevata di laureati fuori corso si registra nei gruppi linguistico (97,7%), architettura (97,6%), economico-statistico (95,9%) e giuridico (95,6%). I gruppi che presentano minore irregolarità sono, oltre al gruppo medico, che spicca per la bassa quota di laureati fuori corso (58,7%), quello psicologico (77,1%) e il gruppo geo-biologico (84,3%).

Tra i laureati irregolari, inoltre, quasi la metà sono fuori corso di lunga durata, persone cioè che hanno protratto per tre anni o più l'iscrizione oltre il termine previsto. La condizione di fuori corso di lunga durata interessa soprattutto i giovani in uscita dal gruppo architettura, che si caratterizza per un'altissima percentuale (60,6%) di laureati oltre il terzo anno fuori corso al momento della laurea.

Circa il 9% di coloro che hanno conseguito il titolo nel 1999 risulta essere stato interessato da processi di mobilità inter-corso. I ragazzi sperimentano percorsi di studio più frammentari di quelli femminili, modificando meno raramente *in itinere* la scelta formativa. Un fattore questo, che contribuisce insieme agli aspetti visti precedentemente (diversa distribuzione dei due sessi nei gruppi di corsi a più elevata irregolarità e obbligo di leva) a far sì che le donne impieghino mediamente meno tempo rispetto agli uomini a conseguire il titolo.

Ancora più interessante appare lo studio della mobilità all'inizio del percorso universitario in quanto, un avvio critico della carriera accademica, frequentemente si risolve nell'abbandono definitivo degli studi. Tra gli studenti che riescono a conseguire il titolo, al netto di quelli che lasciano l'università, l'incidenza di quelli che hanno cambiato corso è quindi relativamente inferiore. Tra i campione di laureati del 1999 che nel corso dell'iter accademico hanno cambiato corso, particolarmente consistente è la quota di quanti poi hanno conseguito il titolo in discipline dei gruppi letterario (20,9%) e politico-sociale (19,0%). Poco rilevante, al contrario, è il numero di giovani che si reindirizzano verso i gruppi ingegneria, benché questo sia tra i più frequentati in assoluto, agrario e chimico-farmaceutico. Gli studenti che cambiano corso, tendono a permanere nello stesso raggruppamento disciplinare o comunque a reinscrivere a corsi di settori affini. Tale scelta, assicurando un maggiore riconoscimento del curriculum pregresso, tende a minimizzare i costi in termini di allungamento di percorso inevitabilmente derivanti dalla mobilità. Sono frequenti, inoltre, i passaggi dall'area scientifica a quella letteraria o sociale, delineando un cambiamento radicale. Il gruppo letterario, in particolare, svolge una chiara funzione di attrazione nei confronti degli studenti, oltre che dei raggruppamenti dell'area umanistica e sociale, dei settori chimico-farmaceutico, scientifico, geo-biologico e architettura. Il gruppo economico-statistico, invece, raccoglie le nuove iscrizioni di molti giovani provenienti dai raggruppamenti architettura, scientifico ed ingegneria.

La mobilità inter-corso, d'altra parte, non è che uno dei possibili effetti determinati da scelte formative poco conformi con l'effettiva preparazione didattica, nonché con inclinazioni personali dei ragazzi. Il cambiamento di indirizzo di studi, pur dilatando i tempi necessari al conseguimento del titolo, non pregiudica però gli esiti.

L'ultimo aspetto considerato riguarda il settore della formazione post-laurea, soprattutto perché in forte espansione. Dall'anno accademico 1995/96, infatti, si assiste ad un consistente aumento del numero di laureati che proseguono gli studi iscrivendosi a corsi di perfezionamento e scuole di specializzazione, fino ad arrivare a 68.700 unità nel 1999/00, con un incremento del 43% circa rispetto al 1991/92 (http://www.murstat.it/ustat/rilevazioni/aa1998_99/siu/guida.htm; <http://www.cnvsu.it/dati/nuclei/default.asp>). L'effettuazione di un periodo di formazione successivo alla laurea è un'esperienza che interessa una la quasi totalità degli intervistati (circa il 95%).

Tra i vari tipi di formazione, il più diffuso è il tirocinio/praticantato che coinvolge più del 26% del campione. Un'esperienza, questa, che interessa particolari gruppi di laurea, anche in virtù della regolamentazione vigente per l'accesso alle libere professioni. Per quanto riguarda il caso degli avvocati, lo svolgimento di un periodo biennale di tirocinio veniva e viene richiesto prima di sostenere l'esame di stato. Si giustifica, quindi, anche alla luce dell'obbligatorietà del praticantato, l'elevata percentuale di tirocinanti con una laurea in giurisprudenza. Di seguito al tirocinio/praticantato viene l'effettuazione di una formazione legata a corsi di formazione professionale o di aggiornamento e di specializzazione post-laurea, seguite a loro volta da una formazione di tipo universitario (dottorato, master, altro corso di laurea).

Il computer è ormai uno strumento indispensabile per chi lavora. Lo dimostra il fatto che su 100 laureati, 92 utilizzano il computer per lavoro almeno una volta al mese e ben 85 lo usano giornalmente. Come prevedibile, la quota di utilizzatori giornalieri è ancor più elevata per i laureati del gruppo ingegneria (97%) ed economico-statistico (96%), mentre scende tra i laureati del gruppo insegnamento e medico (intorno al 40%). Il pc viene utilizzato soprattutto per scrivere (nell'83% dei casi), inserire dati (70%), spedire posta elettronica (64%) e navigare su Internet (63%).

È piuttosto contenuto, invece, l'utilizzo delle lingue straniere: su 100 laureati, 64 leggono almeno una volta al mese in una lingua straniera per motivi di lavoro, 30 lo fanno quotidianamente. Tali valori scendono a 48 (una volta al mese) e 18 (ogni giorno) su 100 laureati se si considerano quanti scrivono e parlano in una lingua straniera.

La presente ricostruzione dello scenario sui neolaureati per macrogruppi di CDL disegna i contorni di una realtà polimorfica, che non mette in luce le molteplici sfaccettature che contraddistinguono i diversi CDL. Ogni macrogruppo è composta da molti CDL ed ognuno di questi rappresenta un particolare sistema. I 41 CDL del gruppo 37 "Scienze Statistiche" vengono aggregati nella classe discipline economi-

che-statistiche, perdendo, in questo modo, le peculiarità che li contraddistinguono (Aureli, Iezzi, 2004). D'altra parte è evidente che nel dover dare un quadro d'insieme non è possibile creare un gruppo per una popolazione di laureati che nell'a.a. 2002/03, in Italia, contavano 1197 laureati, circa lo 0,6% del totale degli addottorati (<http://www.cnvsu.it/dati/nuclei/default.asp>). Un piccolo gruppo che sarà destinato a diminuire ancora nel corso dei prossimi anni se non ci saranno radicali inversioni di tendenza. Infatti, sempre nel 2004 il numero di immatricolati è di 664 studenti, lo 0,2% degli immatricolati totali. Eppure i dati sono incoraggianti: per conseguire una laurea in uno dei CDL della classe 37 si impiegano in media 5,5 anni, ottenendo nel 16,9% dei casi il massimo della votazione (110 e lode). A tre anni dal conseguimento del titolo il 91,8% dei laureati svolge un lavoro continuativo (iniziato dopo la laurea) ed il 66,4% degli occupati ha un impiego per il quale era richiesta la laurea. Il guadagno mensile medio netto di chi lavora in modo continuativo (a tempo pieno) è di poco superiore ai 1200 euro.

3. Utilizzo sequenziale di tecniche multivariate

L'analisi multidimensionale è stata sviluppata in 3 passi distinti⁵ e solo sul campione dei laureati in Statistica:

1. Cluster Analysis (CA), con l'obiettivo di descrivere i profili del campione dei laureati;
2. Scaling Multidimensionale (MDS) per esplorare le dimensioni latenti della formazione ulteriore, modalità utilizzo PC e sfera soddisfazione professionale.
3. Regressione logistica multinomiale (RLM) per classificare le professioni attualmente svolte dai laureati sulla base dei predittori ottenuti dai passi precedenti.

⁵ La notazione utilizzata nella descrizione di vari passi dell'analisi è la seguente:

$\mathbf{X}=[x_{ij}]$	matrice di dati contenente i soggetti intervistati e j variabili, con $i=1, \dots, n$ e $j=1, \dots, m$
K^A	numero di caratteri quantitativi usati nella procedura di classificazione
K^B	numero di caratteri qualitativi usati nella procedura di classificazione
L_k	numero di categorie per la k^{esima} variabile categorica
R_k	range della k^{esima} variabile continua
N	numero di unità statistiche esaminate
N_k	numero di unità statistiche presenti nel gruppo k
S^2_k	varianza stimata della k^{esima} variabile continua in \mathbf{X} .
S^2_{jk}	varianza stimata della k^{esima} variabile continua nel gruppo j .
N_{jkl}	numero di unità statistiche presenti nel gruppo j da cui la k^{esima} variabile categorica prende l' l^{esima} categoria
$d(j, s)$	distanza tra il gruppo j ed s .
$\langle j, s \rangle$	indice che rappresenta il gruppo formato dalla combinazione delle classi j ed s .

PASSO 1

Al fine di raggruppare gli individui in un numero ristretto di classi, si è utilizzato l'algoritmo *two-step cluster* (Zhang T., Ramakrishnon R., Livny M., 1996; Chiu T., Fang D., Chen J., Wang Y., Jeris C., 2001), che consente di analizzare simultaneamente sia caratteri qualitativi che quantitativi. Inoltre, permette di ottenere una partizione ottimale senza dovere utilizzare metodi gerarchici e non gerarchici in modo sequenziale, ma simultaneamente gli oggetti vengono raggruppati in *cluster* "naturali". Come è noto, la caratteristica fondamentale degli algoritmi gerarchici è produrre una sequenza di partizioni in classi via via più ampie, gli esempi più noti sono quelli prodotti dagli zoologi per classificare le specie, i generi, le famiglie ecc; i metodi non gerarchici, invece, producono direttamente una partizione in un numero prefissato di classi. Questo algoritmo produce entrambe le classificazioni (gerarchica e non gerarchica) in due fasi. Nella prima fase, si calcolano gli indici BIC (Schwarz's Bayesian Information Criterion) e AIC (Akaike's Information Criterion) per un certo numero di gruppi, e si specifica l'intervallo con il numero prefissato di cluster sui quali calcolare le stime iniziali [1] e [2]. Nella seconda fase, le stime iniziali sono ricalcolate per aggregare unità o gruppi più vicini sulla base del criterio della distanza minore.

Gli indici BIC e AIC per J gruppi sono definiti come:

$$BIC(J) = -2 \sum_{j=1}^J \xi_j + m_j \log(N) \quad [1]$$

$$AIC(J) = -2 \sum_{j=1}^J \xi_j + 2m_j \quad [2]$$

con $m_j = 2jK^A + j \sum_{k=1}^{K^B} (L_k - 1)$.

La distanza tra il gruppo j e s è definito come: $d(j,s) = \xi_j + \xi_s - \xi_{\langle j,s \rangle}$, ove

$$\xi_v = -N_v \left[\sum_{k=1}^{K^A} \frac{\log(S_k^2 + S_{vk}^2)}{2} - \sum_{k=1}^{K^B} \left(\sum_{l=1}^{L_k} \frac{N_{vkl}}{N_v} \log \frac{N_{vkl}}{N_v} \right) \right] \quad [3]$$

La logverosimiglianza delle distanze consente di misurare contemporaneamente variabili sia categoriche che continue. Naturalmente nel calcolare la logverosimiglianza è necessario che sia rispettata l'ipotesi di normalità per le variabili continue. Si assume, inoltre, che le variabili e i casi siano indipendenti gli uni dagli altri.

Se S_k^2 è ignorata nell'equazione [3], la distanza tra i gruppi j e s è pari alla diminuzione della logverosimiglianza quando i 2 gruppi sono combinati. Il termine S_k^2 è aggiunto per risolvere il problema causato da $S_{vk}^2=0$, che comporterebbe un risultato indefinito. Questa ipotesi si verifica quando un gruppo contiene un solo caso.

Quando la violazione dalla normalità è forte, risulta difficile sostenere che le variabili possiedano i requisiti necessari per l'applicazione delle analisi parametriche. Allora, si consiglia di trasformare le variabili continue in categoriali ordinabili (Farrington e Loeber, 2000; Hunter e Schmidt, 1990).

Quando il campione è casuale, ogni variabile categoriale, può essere considerata proveniente da una distribuzione multinomiale:

$$pr(Y_1 = y_1, \dots, Y_k = y_k; m, \pi) = \binom{m}{y} \pi_1^{y_1} \dots \pi_k^{y_k} ,$$

dove π_1, \dots, π_k sono le frequenze attribuite ad una popolazione infinita.

PASSO 2

In questa fase, si applica il modello MDS non metrico [4] su 3 sottomatrici della matrice \mathbf{X} , che costituiscono i dati di input di tre differenti analisi (formazione ulteriore, utilizzo del PC, livello di soddisfazione).

$$\delta_{ij} = f\left[\sum_{s=1}^k (x_{is} - x_{js})^2\right]^{1/2} . \quad [4]$$

dove f è una funzione monotona tale che (monotonicità debole): $\delta_{ij} < \delta_{lh} \Rightarrow d_{ij} \leq d_{lh}, \forall i, j, l, h$. In questa specifica procedura di ricerca, le soluzioni si ottengono in maniera iterativa, ossia dopo aver creato una configurazione iniziale casuale dalla matrice di dissomiglianza iniziale, si calcolano le distanze standardizzate tra le n unità e si confronta l'ordinamento delle dissomiglianze con quelle ottenute al passo precedente ed, infine, si calcola una misura di adattamento. Per arrestare l'algoritmo si è scelta la condizione che fosse assicurato un adattamento ottimo, mediante la misura di S-STRESS (Takane et al., 1977) e il quadrato del coefficiente di correlazione lineare (RSQ).

PASSO 3

Nell'ultimo passo dell'analisi si applica il modello di RLM [5] per classificare i soggetti intervistati sulla base dei predittori ottenuti nei passi 1 e 2. Come è noto, questo tipo di regressione è simile alla regressione logistica, ma è più generale in quanto la variabile dipendente non è ristretta in due categorie ma in un numero superiore, e può essere sinteticamente espressa con:

$$\pi_{ik} = \frac{e^{z_{ij}}}{e^{z_{i1}} + e^{z_{i2}} + \dots + e^{z_{ik}}} , \quad [5]$$

dove:

Π_{ik} = è la probabilità che l' i^{esimo} laureato ricada nella k^{esima} categoria professionale;
 Z_{ik} = è il valore della k^{esima} variabile non osservata per l' i^{esimo} individuo;
 $Z_{ik} = b_{k0} + b_{k1}x_{i1} + \dots + b_{kj}x_{ij}$
 x_{ij} = è il valore dell' j^{esimo} predittore per l' i^{esimo} individuo;
 b_{kj} = è il j^{esimo} coefficiente per la k^{esima} variabile non osservata;
 j = è il numero dei predittori.

4. I laureati in Scienze Statistiche

Le variabili utilizzate per disegnare il profilo dei rispondenti sono sette, di cui due assimilabili a continue (voto di maturità e voto di laurea) e le altre categoriche (status familiare, sesso, tipo di maturità, stato civile, classe d'età).

Per verificare la normalità delle variabili continue *voto di maturità* e *voto di laurea* si è utilizzata sia la rappresentazione dei quantili, in modo tale da confrontare i quantili della distribuzione della variabile con i quantili della distribuzione normale, sia l'esame degli indici di asimmetria e di curtosi.

L'esame degli indici di asimmetria e curtosi (tab. 1) e l'analisi grafica dei Q-Q plot (fig. 4) suggeriscono che le distribuzioni presentano votazioni alte con una maggiore frequenza, ed inoltre che le distribuzioni sono tendenzialmente "platicurtiche". Inoltre, il test di normalità di Kolmogorov-Smirnov (tab. 2) conferma quanto emerso dall'esame grafico e dalla lettura degli indici di asimmetria e curtosi, ossia che le variabili voto di maturità e di laurea deviano in maniera significativa dalla ipotesi di normalità. Per questo motivo si è ritenuto opportuno trasformarle da variabili continue a categoriali ordinabili.

Tabella 1. Indici di asimmetria e di curtosi per le variabili "voto di maturità" e "voto di laurea"

Indici	Asimmetria	Curtosi	E. S. Asimmetria	E. S. Curtosi
voto di maturità	-0,22	-0,99	0,13	0,26
voto di laurea	-0,40	-0,91	0,13	0,26

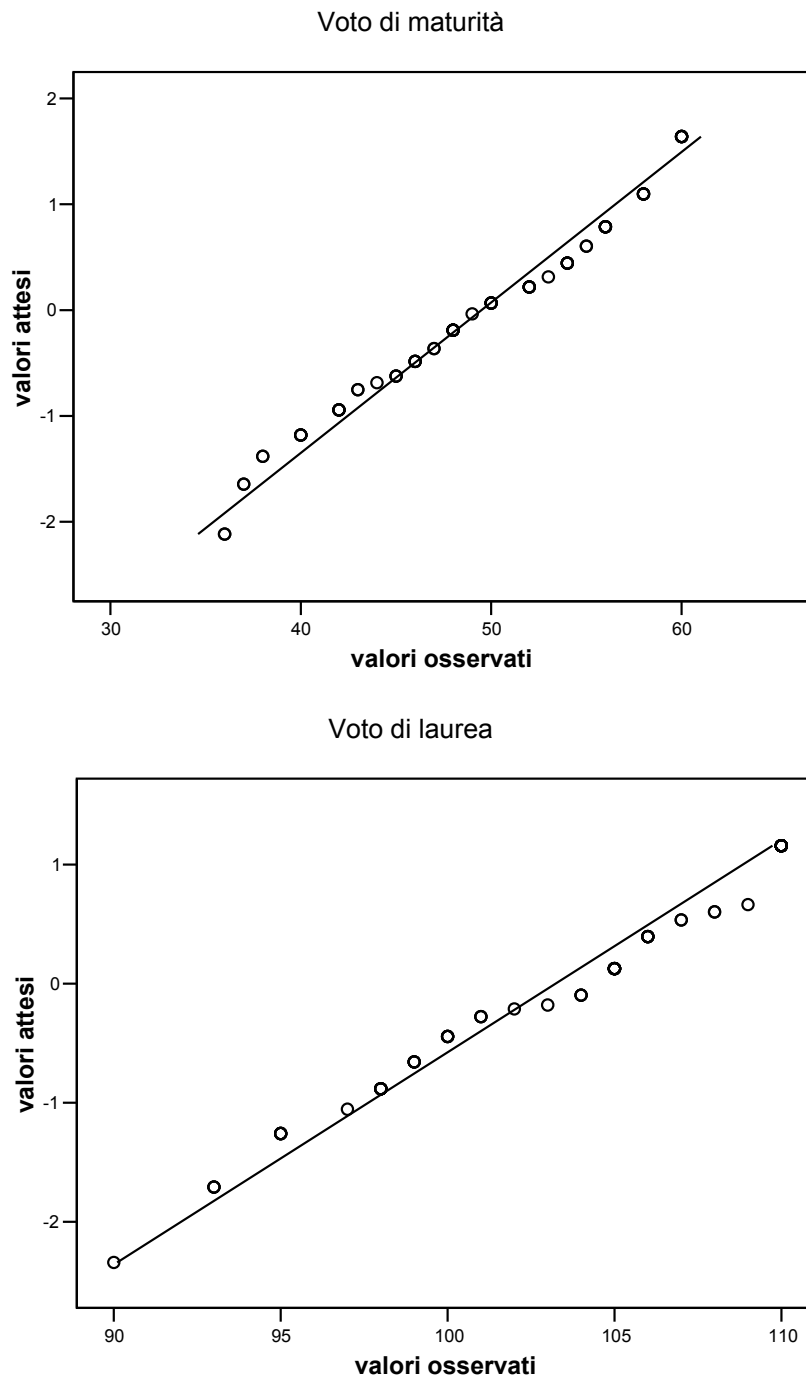
Fonte: Elaborazioni proprie su dati ISTAT

Tabella 2. Test di Normalità di Kolmogorov-Smirov, con la correzione di significatività di Lillefors

variabile	Statistica	df	Sig.
voto di maturità	0,11	339,76	0,00
voto di laurea	0,14	339,76	0,00

Fonte: Elaborazioni proprie su dati ISTAT

Fig. 4. Grafici dei quantili (Q-Q Plot) per la verifica della normalità.



PASSO 1

L'algoritmo di classificazione adottato individua due profili identificabili con lo status familiare dei rispondenti (tab. 3). Il primo gruppo di laureati (il 75% dei casi) composto sia da donne che da uomini (F-M), provenienti da una famiglia con uno status sociale basso/medio, che ha frequentato in prevalenza una scuola superiore di tipo tecnico, ma con un punteggio finale elevato (51-60), con un'età compresa tra i 28 e 30 anni, liberi e coniugati ed un voto di laurea estremamente variabile (91-110); un secondo gruppo (25% dei casi) composto prevalentemente da uomini (M), provenienti da una famiglia con uno status sociale alto, di età mediamente superiore ai 30 anni, proveniente da licei, non coniugati, con un voto di maturità variabile (36-60), ma un voto di laurea generalmente alto (105-110). Coloro i quali provengono da famiglie con uno status sociale basso o medio non hanno frequentato un liceo, ma sono i "bravi" degli Istituti Tecnici. Da questa prima analisi sembra che il bagaglio culturale di tipo tecnico limiti una parte degli studenti nel raggiungimento degli obiettivi finali, infatti soltanto alcuni di loro riescono ad ottenere una votazione elevata. Tale ipotesi sarà poi ulteriormente valutata nelle fase 2 e 3 del modello proposto, in merito ad aspettative e prospettive.

Tabella 3. *Il profilo dei laureati in Scienze Statistiche*

caratteri osservati	gruppo 1	gruppo 2
<i>status familiare</i>	basso-medio	alto
<i>sex</i>	F-M	M
<i>tipo di maturità</i>	Tecnica	licei
<i>età</i>	28-30	oltre 30
<i>stato civile</i>	coniugati-liberi	liberi
<i>voto di laurea</i>	91-110	105-110
<i>voto di maturità</i>	51-60	36-60

Fonte: Elaborazioni proprie su dati ISTAT

PASSO 2

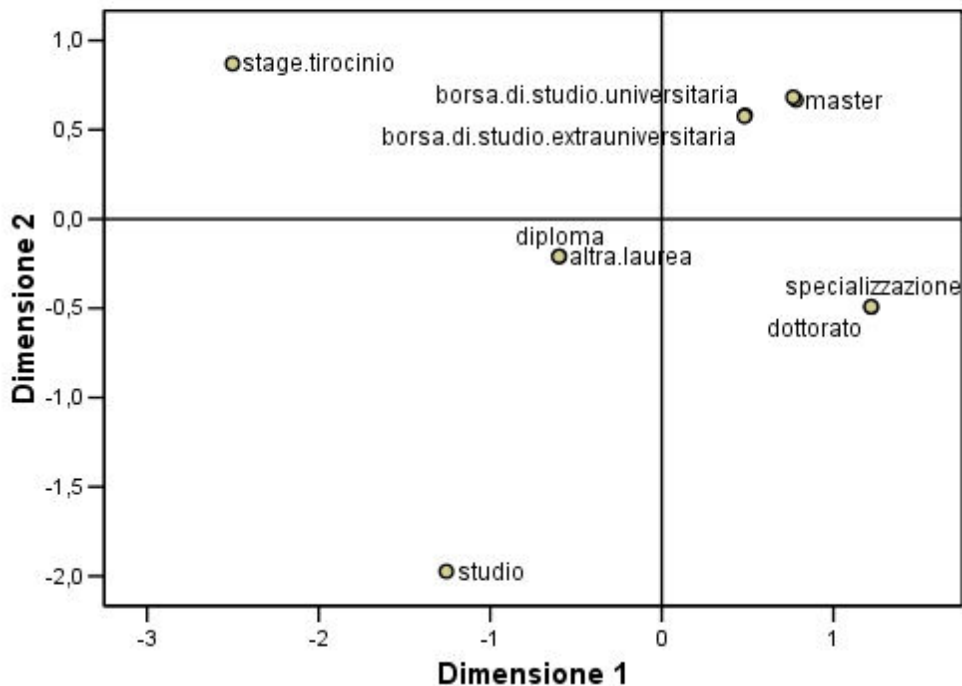
In questo passo, si esaminano le dimensioni latenti di:

- 1) formazione post-laurea;
- 2) grado di soddisfazione nello svolgere l'attuale lavoro;
- 3) utilizzo del personal computer (PC).

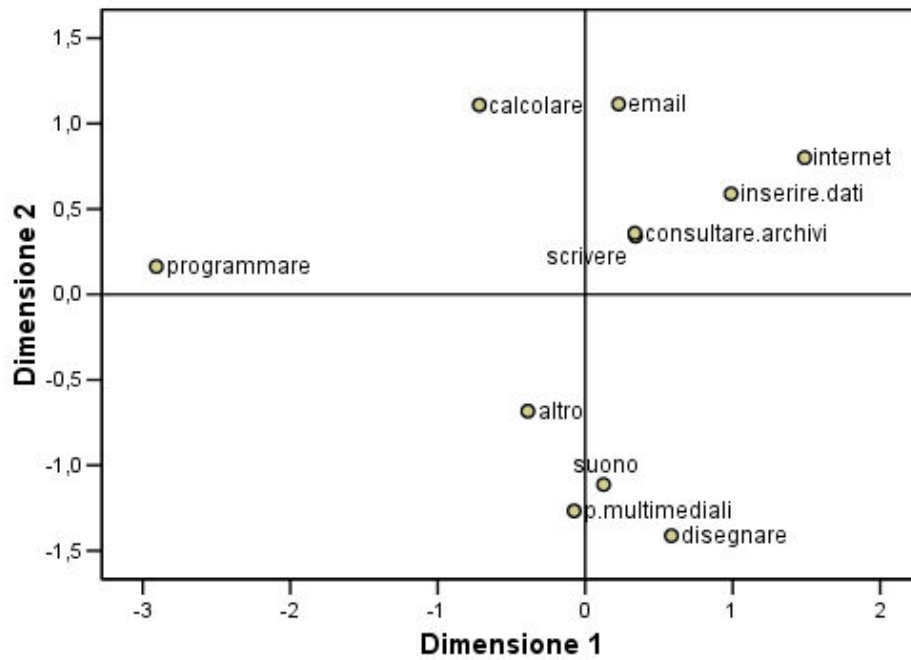
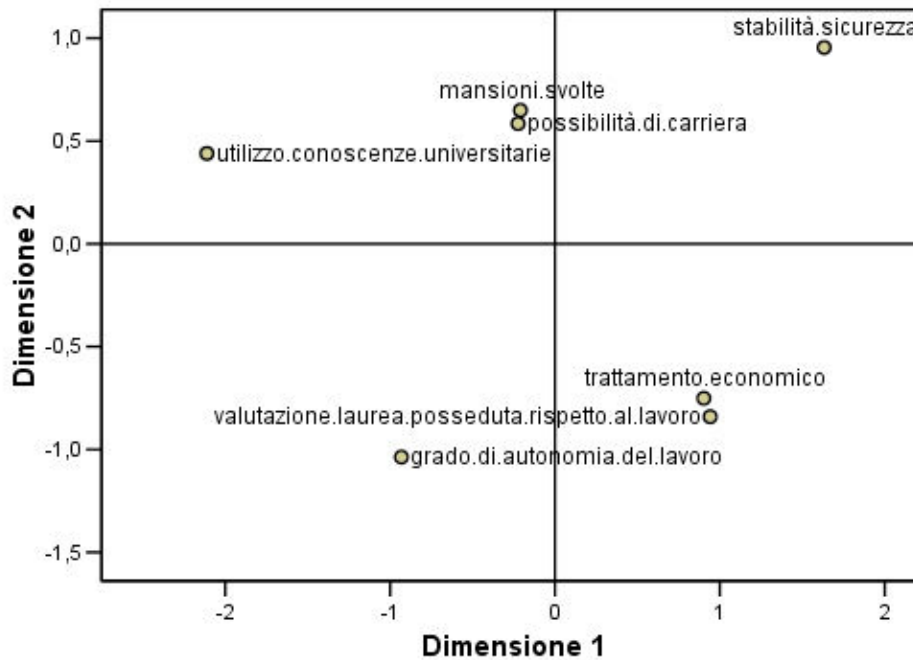
Il punto 1) rappresenta l'acquisizione di competenze e di conoscenze del laureato, rendendolo anche più duttile ai mutamenti del mercato del lavoro; il 2) misura l'appagamento del lavoratore; il 3) le conoscenze e le abilità nei confronti di strumenti necessari a svolgere velocemente e con perizia dei compiti di natura tecnica.

Le dimensioni latenti della formazione post-laurea sono esaminate rispetto a 8 possibili scelte: stage-tirocinio, borsa di studio universitaria, borsa di studio extrauniversitaria, master, diploma-altra laurea, specializzazione, dottorato e studio. Gli indici di stress e di bontà di adattamento confermano che la soluzione in 2 dimensioni è buona (S-Stress=0,00137; Tucker's Coefficient of Congruence 0,99905). La rappresentazione grafica della configurazione (fig. 5) evidenzia l'esistenza di una posizione contrapposta tra la formazione ulteriore del genere *market oriented* (stage-tirocinio) - e quella *individual oriented* (dottorato di ricerca e specializzazione).

Fig. 5. Configurazione in 2 dimensioni delle 8 tipologie di formazione ulteriore



Le dimensioni dell'utilizzo del PC sono esaminate rispetto ad 11 modalità: calcolare, e-mail, interne inserire dati, consultare archivi, scrivere, programmare, altro, suono, programmi multimediali, disegnare. La prima dimensione (fig. 6) è quella del programmatore esperto, mentre la seconda è quella dei utenti: di programmi di videoscrittura o fogli di calcolo (individui posizionati sulla parte superiore dell'asse delle ordinate) o di programmi multimediali (parte inferiore dell'asse delle ordinate). Lo scrivere non è un elemento discriminante in quanto appartiene a tutto il campione intervistato. Anche in questo caso gli indici di stress e di bontà di adattamento confermano che la soluzione in 2 dimensioni è buona (S-Stress=0,002; Tucker's Coefficient of Congruence 0,97803).

Fig. 6. Configurazione in 2 dimensioni delle 11 modalità di utilizzo del PC**Fig. 7.** Configurazione in 2 dimensioni dei 7 livelli di soddisfazione

Le dimensioni della soddisfazione sono analizzate rispetto a 7 livelli: mansioni svolte, stabilità e sicurezza, possibilità di carriera, utilizzo di conoscenze universitarie, trattamento economico, valutazione laurea posseduta rispetto al lavoro e grado di autonomia dal lavoro (fig. 7). A tre anni dalla laurea il trattamento economico si pone in contrapposizione con le mansioni svolte e la possibilità di far carriera, l'autonomia nelle mansioni svolte con la stabilità e la sicurezza. Da questa configurazione emerge che coloro i quali preferiscono la sicurezza o la stabilità devono rinunciare alla autonomia nel lavoro, chi punta ad un adeguato trattamento economico rinuncia alla possibilità di far carriera. Anche in questo caso gli indici di stress e di bontà di adattamento confermano che la soluzione in 2 dimensioni è buona (S-Stress=0,001; Tucker's Coefficient of Congruence 0,98763)

PASSO 3

Nell'ultimo passo si vogliono classificare le professioni svolte dai laureati sulla base dei profili degli stessi e le dimensioni latenti della formazione ulteriore, della soddisfazione e delle modalità di utilizzo del PC, individuate nei passi 1 e 2. I gruppi individuati nel passo 1 della analisi costituiscono il fattore del modello RLM e le dimensioni latenti individuate nel passo 2 dell'analisi le covariate del modello. La variabile dipendente è la categoria professionale di appartenenza. La classificazione predisposta dall'ISTAT è formata da 8 classi: legislatori, dirigenti, imprenditori; professioni intellettuali, scientifiche, di elevata specializzazione; professioni intermedie (tecnici); professioni esecutive relative all'amministrazione e gestione, professioni relative alle vendite ed ai servizi per famiglie, artigiani, operai specializzati, agricoltori, conduttori di impianti, operatori di macchinari, personale non qualificato, forze armate. I laureati in statistica appartengono soltanto a 4 delle 8 categorie sopra elencate, ed in base al modello stimato, che presenta un ottimo adattamento con uno pseudo R^2 di Cox e Snell pari 0.90, vengono classificati come riportato nella tab. 5, che evidenzia un errore di classificazione inferiore al 22%.

Tab. 5. *Classificazione ottenuta dalla applicazione del RLM*

OSSERVATI	PREDETTI				% CASI ESATTAMENTE CLASSIFICATI
	A	B	C	D	
A	5,08	0,00	0,00	0,00	100,00
B	0,00	10,35	16,16	0,00	39,05
C	0,00	6,08	45,32	0,00	88,17
D	0,00	0,00	0,00	20,98	100,00

Legenda: A= legislatori, dirigenti, imprenditori;
 B= professioni intellettuali, scientifiche, di elevata specializzazione;
 C= professioni intermedie (tecnici; professioni esecutive relative all'amministrazione;
 D= professioni esecutive relative all'amministrazione e gestione.

5. Conclusioni e prospettive

L'analisi, effettuata sulla base dei dati dell'Indagine Istat, mostra che alcune scuole superiori preparano meglio di altre ad affrontare il percorso universitario, inoltre, le donne presentano performance preaccademiche ed accademiche migliori rispetto ai loro colleghi maschi, anche se poi scelgono percorsi universitari meno selettivi, ma anche poco richiesti dal mercato. Le motivazioni che spingono le donne a preferire percorsi formativi con limitati sbocchi professionali sono probabilmente da far risalire a modelli familiari tradizionali dove la donna è ancora prevalentemente relegata al ruolo di madre e di educatrice. Una evoluzione si può cogliere nelle maggiori capacità informatiche e nell'impegno della maggior parte dei laureati profuso durante la formazione post-laurea.

I laureati in Scienze Statistiche, al contrario di molti loro colleghi, dopo 3 anni dalla laurea trovano una occupazione soprattutto nel ruolo di tecnici fornendo attività di supporto a molti settori (medico, economico, sociale). Soltanto un piccolo gruppo di costoro è destinato a svolgere una carriera dirigenziale, probabilmente a causa di un bagaglio culturale essenzialmente tecnico. Inoltre, le stesse aspettative dei giovani laureati da un breve lasso di tempo sono poco ambiziose, in quanto una volta conseguito il titolo, trovando subito un lavoro, lasciano poco spazio alla formazione post-laurea.

Ad ogni modo, per avere un quadro completo e dettagliato della transizione università-lavoro potrebbero essere applicati anche modelli di regressione multilivello (Goldstein, H.; 1995, Chiandotto B., Bacci S., 2005), che consentono di descrivere in maniera più compiuta la reale situazione occupazionale e fornire adeguate indicazioni di *policy*. Le tecniche statistiche di analisi multivariata forniscono validi mezzi per esaminare i dati ed indicare possibili decisioni strategiche.

Riferimenti bibliografici

- AURELI E., IEZZI D.F. (2004), Mappe conoscitive per l'analisi dell'offerta formativa nel settore statistico. In: E. AURELI CUTILLO (a cura di) *Strategie metodologiche per lo studio della transizione Università-lavoro*, CLEUP, Padova, **5**: 1-18.
- CHIANDOTTO B., BACCI S. (2005), Un modello multilivello per l'analisi della condizione occupazionale dei laureati. In: C. CROCETTA (a cura di) *OUTCOMES: Modelli di analisi della transizione università-lavoro*, CLEUP, Padova, **6**: 217-240.

- CHIU T., FANG D., CHEN J., WANG Y. and JERIS C. (2001), A Robust and Scalable Clustering Algorithm for Mixed Type Attributes. In “*Large Database Environment*”. *Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining*: 263.
- FARRINGTON D. P. and LOEBER R. (2000), Some benefits of dichotomization. *Psychiatric and criminological research. Criminal behaviour and mental health*, **10**, 100-122.
- GOLDSTEIN H. (1995) *Multilevel Statistical Models*, Kendall’s Advanced Theory of Statistics. Arnold, London / Halsted, New York.
- GOLDSTEIN H. (2003), *Multilevel statistical models*, Arnold Publishers, London.
- HUNTER J. E. and SCHMIDT F. L. (1990) Dichotomous of continuous variables: the implications for meta-analysis. *Journal of Applied Psychology*, **3**: 334-349.
- IEZZI D.F. (2003), *Uno strumento per migliorare la qualità*. Universitas, **90**: 12-17.
- ISTAT (2002), *Indagine 2002 sull’inserimento professionale dei diplomati universitari del 1999*, Istat, Roma.
- ISTAT (2004), *Università e lavoro. Statistiche per orientarsi - 2004*, Istat, Roma.
- TAKANE Y., YOUNG F. W., DE LEEUW J. (1977). Non metric individual differences multidimensional scaling. An alternative least squares method with optimal scaling features, *Psychometrika*, **42**: 7-67.
- ZHANG T., RAMAKRISHNON R. and LIVNY M. (1996). *BIRCH*: An Efficient Data Clustering Method for Very Large Databases. In: *Proceedings of the ACM SIGMOD Conference on Management of Data*, Montreal, Canada: 103-114.

Analysis of the university-work transition: objectives and perspectives of the graduates in Statistical Sciences

Summary: *The labour market requires that a graduate possess not only the knowledge and competencies related to the degree obtained, but also transversal skills. In this context, it is extremely important to evaluate the external effectiveness of the formative process of the different study programmes. Currently many surveys analyse the labour situation and the entry into the work force of new employees.*

The aim of this paper is to review the principal databases that explore the university-work transition and, in particular, to examine the occupational situation of graduates in Statistic Sciences. For this purpose we have applied different techniques of multivariate analysis, jointly (Cluster analysis, multidimensional scaling and multinomial logistic regression).

Keywords: *Databases, cluster analysis, multidimensional scaling, multinomial regression analysis.*