

Dipartimento di:
Impresa e management

Cattedra:
Statistica

**PROFILI DI COMPETITIVITÀ
DELLE REGIONI ITALIANE
ATTRAVERSO LA CLUSTER ANALYSIS**

RELATORE
Prof.ssa Livia De Giovanni

CANDIDATO
Marco D'Epifanio
Matr. 164821

Anno Accademico:
2012-2013

PROFILI DI COMPETITIVITÀ DELLE REGIONI ITALIANE ATTRAVERSO LA CLUSTER ANALYSIS

Indice

Ringraziamenti	5
Presentazione della ricerca	6

CAPITOLO 1 L'ITALIA NEL CONTESTO INTERNAZIONALE

1.1 Il ruolo della domanda estera	9
1.1.1 Performance delle imprese esportatrici.....	10
1.1.2 Specializzazione all'export.....	11
1.1.3 Internazionalizzazione Produttiva	13
1.1.4 Forme di internazionalizzazione.....	15
1.2 Competitività del sistema produttivo italiano	17
1.2.1 Conoscenza e Innovazione	20
1.2.2 Regolamentazione del mercato del lavoro	22
1.2.3 Sistema dei trasporti	23
1.2.4 Giustizia civile	25
1.2.5 Economia sommersa	26

CAPITOLO 2 I DATI DELLE REGIONI ITALIANE: DELINEARE LA COMPETITIVITÀ

2.1 Le Basi di dati	28
2.1.1 Istat	29
2.1.2 Unioncamere	30
2.1.3 ICE.....	31
2.2 Le variabili considerate.....	32
2.2.1 Pil pro capite.....	32
2.2.2 Produttività del lavoro.....	33
2.2.3 Costo del lavoro	34
2.2.4 Tasso di natalità delle imprese	35

2.2.5 Valore aggiunto delle grandi imprese	36
2.2.6 Scambi di merci con il mondo	36
2.2.7 Addetti di imprese estere partecipate da imprese italiane	37
2.2.8 Addetti di imprese italiane a partecipazione estera	38
2.2.9 Spesa in Ricerca e Sviluppo	39
2.2.10 Numero di Brevetti	40
2.2.11 Numero di Laureati	41
2.2.12 Investimenti pubblici	42

CAPITOLO 3

LA METODOLOGIA STATISTICA DELLA CLUSTER ANALYSIS

3.1 Le unità da raggruppare	44
3.2 Similarità delle unità	45
3.3 Metodi di formazione dei gruppi	47
3.3.1 Metodi gerarchici	47
3.3.2 Metodi non gerarchici	48
3.4 Valutazione della partizione	50
3.5 Procedura di formazione dei gruppi	52
3.6 Il software R	53

CAPITOLO 4

PARTIZIONE DELLE REGIONI ITALIANE IN GRUPPI

4.1 La matrice dei dati	54
4.2 Cluster analysis: metodi gerarchici	57
4.2.1 Valutazione della bontà delle partizioni	60
4.3 Cluster analysis: metodo delle k-medie	61
4.3.1 Valutazione della bontà delle partizioni	63
4.4 Scelta della partizione finale	64
4.5 I cluster regionali	66
4.5.1 Gruppo 1: reddito alto, grandi imprese, investimenti all'estero e brevetti	69
4.5.2 Gruppo 2: reddito altissimo, efficienza del lavoro e spesa pubblica	71
4.5.3 Gruppo 3: reddito alto, elevate esportazioni	72
4.5.4 Gruppo 4: reddito medio, profilo medio	73

4.5.5 Gruppo 5: reddito alto, alta natalità delle imprese e spesa in R&S	74
4.5.6 Gruppo 6: reddito basso, numerosi laureati	75
4.5.7 Gruppo 7: reddito bassissimo, profonde lacune	76
4.6 Conclusioni.....	77
Appendice.....	79
Bibliografia.....	83
Siti Internet consultati	84

Ringraziamenti

Al termine di un progetto quale la tesi di laurea, ciò che viene alla mente, oltre al tempo e alle energie spese, sono tutte le persone che hanno contribuito, in modo più o meno consapevole, alla completezza dell'elaborato. Queste sono le persone che mi hanno confortato e motivato a portare avanti il lavoro con il massimo dell'impegno in ognuna delle sue parti, sono coloro che con un atto di amore mi hanno sostenuto anche se non era loro richiesto o dovuto. Rivolgo i miei più sinceri e sentiti ringraziamenti a tutti quelli che hanno condiviso con me questo impegno, che lo hanno reso vivo e reale, a cui pensavo durante il lavoro e a cui dedicavo tutta la mia perseveranza e la mia fatica.

Prima fra tutti, il ringraziamento più speciale va alla prof.ssa Livia De Giovanni che mi ha guidato in questo percorso e ha creduto nelle mie potenzialità. Di lei non scorderò mai la professionalità con cui mi indicava i passi da seguire con la massima cura e meticolosità, e soprattutto la pazienza che ha saputo mantenere anche con il tesista alle prime armi più insicuro. Ringrazio la professoressa anche per la bellissima persona che è, per tutto l'affetto che riesce a donare e per tutta la solarità con cui ha dipinto questo percorso.

Tengo a ringraziare il Centro di ricerca Luiss CEFOP (Centro di Economia e della Formazione delle Professioni), in particolare la cortesia e la disponibilità della sua direttrice, la dottoressa Francesca Sica, che ha partecipato in questo lavoro nel reperimento e nell'analisi dei dati.

Ringrazio poi i miei più cari amici, che sono per me quella parte della vita per cui vale la pena dedicarsi ad un percorso di investimento nella propria carriera. Li ringrazio dal più profondo del cuore per la gioia che mi fanno provare e per l'essenza con cui riempiono la mia vita che mi porta ad avere la forza di compiere dalle azioni più quotidiane ai progetti che richiedono più costanza e cura.

Infine vorrei ringraziare la mia famiglia, che riesce a toccare corde più profonde del cuore, legate a parti di me più essenziali e basilari. In questo periodo il loro amore e la loro presenza mi ha risollevato nei momenti più cupi e ha dato più spessore alle realtà che ho vissuto. Ai miei genitori devo tutto, li ringrazio perché sono riusciti a donarmi l'unica cosa di cui un figlio ha veramente bisogno.

Per il resto, dedico questo lavoro, come ogni altro mio sforzo quotidiano, allo scopo che inseguo ormai da tempo e a cui più ambisco, la ricerca dell'essere. Una vita libera da tutte le trappole della mente, che permette di esprimere pienamente se stessi e di avere quell'apertura in grado di percepire la vita senza fermarsi a quelle realtà fasulle che ci costruiamo. Solo questa è una vita piena, una vita realmente vissuta e felice. A tale vita dedico questo lavoro insieme a tutto me stesso.

Presentazione della ricerca

Il lavoro che si va a presentare si pone al termine del triennio di studi di Economia e Management.

Il cuore della ricerca è la metodologia statistica della *cluster analysis*, ovvero l'analisi dei gruppi, che consente di raggruppare una vasta e variegata popolazione di unità in *cluster* distinti in cui tutti i componenti condividono le medesime caratteristiche. Questa metodologia è indicata per dividere in segmenti il mercato dei consumatori, quello dei prodotti oppure per eseguire una classificazione a livello territoriale.

Il lavoro in questione rientra in quest'ultima specie e considera le regioni italiane. Ognuna di queste sarà identificata attraverso dati raccolti per numerosi indicatori che possano delineare il profilo competitivo del sistema produttivo regionale.

La competitività è intesa dunque come un concetto poliedrico che trova la sua spiegazione attraverso numerosi aspetti economici che rientrano nelle aree di crescita reale, apertura all'estero, conoscenza e fattori amministrativi. Gli indicatori considerati sono la produttività e il costo del fattore produttivo Lavoro, la natalità delle imprese e la loro dimensione, l'ammontare di esportazioni nette o di investimenti esteri di lunga durata, la spesa in Ricerca & Sviluppo, le domande di brevetti, il tasso di laureati e l'ammontare della spesa pubblica.

L'analisi competitiva a livello regionale si colloca in quella più ampia a livello nazionale.

L'Italia registra negli anni a seguito della crisi del 2008-2009 una forte contrazione della domanda interna a favore di quella estera che può quindi rappresentare una nuova opportunità di crescita economica nel medio periodo. Questa possibilità di indirizzare la crescita produttiva ai mercati esteri viene confermata studiando la capacità, attuale e potenziale, di conquistare spazi sui mercati esteri: analizzando la variazione delle vendite da parte delle imprese esportatrici, il ruolo dell'Italia nelle dinamiche mondiali di *export* e infine studiando il grado di apertura internazionale, sia in termini quantitativi che qualitativi.

Le possibilità di crescita per l'Italia sono approfondite tramite l'analisi competitiva del sistema produttivo italiano che considera le precedenti aree economiche per effettuare un *benchmarking* con gli altri paesi europei. Da questa analisi emergono quelli che si possono definire i fattori determinanti per la crescita, come l'investimento in conoscenza ed innovazione, nonché i fattori che in Italia risultano critici per la crescita in quanto ne ostacolano il pieno dispiegamento, quali la regolamentazione del mercato del lavoro, la particolare configurazione del sistema logistico, il funzionamento della giustizia civile e l'ammontare dell'economia sommersa.

All'interno di questo contesto si introducono le varie situazioni regionali. Queste saranno messe a confronto con lo scopo di ricercare *cluster* regionali con un profilo competitivo, definito in tutte le precedenti variabili, che trova la sua manifestazione in un dato livello di ricchezza, ovvero nel reddito medio delle regioni raggruppate.

La presente tesi è organizzata in quattro parti descritte in seguito.

Nel primo capitolo sono presentate le possibilità di crescita dell'Italia all'estero e dunque l'analisi competitiva del sistema produttivo italiano a confronto con gli altri principali paesi europei. A partire da quest'ultima saranno poi approfonditi i fattori determinati e critici per l'espansione produttiva.

Il secondo capitolo riporta, per ogni variabile economica in grado di rappresentare un aspetto economico del sistema produttivo regionale, l'insieme dei dati raccolti degli enti dell'Istat, Unioncamere ed ICE per i quali è stata dedicata una spiegazione delle loro attività e servizi che mettono a disposizione.

Successivamente, nel terzo capitolo, saranno descritti gli aspetti teorici della *cluster analysis*, presentando una pluralità di strumenti statistici integrati in un'unica procedura nonché il *software* utilizzato per l'implementazione dell'analisi.

Infine, nell'ultimo capitolo, sarà discussa la procedura di partizione delle regioni italiane per poi analizzare nel dettaglio i profili competitivi manifestati da ogni gruppo di regioni.

CAPITOLO 1

L'ITALIA NEL CONTESTO INTERNAZIONALE

La nostra ricerca comincia con l'analisi a livello nazionale del sistema produttivo italiano, ovvero con l'insieme aggregato di tutte le regioni nel suo funzionamento complessivo. Una volta definito il profilo competitivo dell'Italia, in relazione agli altri principali paesi europei, procederemo ad individuare quei sotto-insiemi, cioè quei raggruppamenti regionali, che condividono il medesimo livello di competitività.

Il carattere della “competitività” di un sistema economico-produttivo è un concetto poliedrico, ovvero definito attraverso numerosi aspetti e misurato tramite determinati indicatori.

Questi possono riguardare la dinamica della produttività e quindi si rivolgeranno all'efficienza dei fattori di produzione, al funzionamento e alle altre caratteristiche generali delle imprese.

La capacità di un sistema produttivo di guadagnarsi spazi sui mercati esteri è un altro aspetto rilevante per il suo grado di competitività. Esso determina infatti la portata della possibilità di espandersi oltre il livello concesso dalla sola domanda interna nonché la possibilità di ottenere dall'estero una quantità e una qualità di risorse diverse da quelle ottenibili internamente.

Un altro aspetto che determina la qualità competitiva è il livello tecnologico, direttamente influenzato dall'ammontare di investimenti nella ricerca, innovazione e istruzione.

Infine occorre considerare oltre al contesto puramente economico anche quello meta-economico, cioè che fornisce l'ambiente operativo dove il sistema produttivo può svolgere le sue funzioni. Questo è costituito chiaramente dal sistema giuridico e politico che possono favorire l'economia di un paese come anche rallentarla. Dal primo discendono le procedure in materia di commercio dei prodotti, mercato del lavoro, del credito e risoluzione delle controversie. Dall'ambito amministrativo discende l'ammontare e l'efficienza delle infrastrutture, ovvero i beni e servizi che determinano l'utilizzabilità delle aree geografiche (quali i servizi di trasporto) come il benessere dei cittadini, da una parte compratori e dall'altra lavoratori.

Ognuno di questi aspetti registra un valore assoluto in grado di esprimere la quantità generica per una data variabile, tuttavia è il confronto con gli altri sistemi produttivi che permette di capire la rilevanza relativa di ogni aspetto del sistema e dunque di capirne la qualità competitiva.

Attraverso un confronto tra diversi sistemi produttivi con caratteri essenziali in comune, come possono essere gli stati dell'Unione europea, è possibile valutare per ognuno di essi i punti di forza e di debolezza così da poter preservare i primi ed intervenire sui secondi. Attraverso questo genere di paragone è anche possibile una ricerca delle migliori soluzioni, per esempio si può eleggere per ogni dato aspetto di una performance il sistema produttivo più esemplare da utilizzare come riferimento.

In una recente ricerca condotta dall'Istat nel Rapporto annuale 2012, viene affrontata proprio questa analisi rivolta alla competitività dell'Italia paragonata ai principali paesi europei. Questo studio costituisce non solo la base nazionale della nostra analisi a livello regionale, ma anche l'impostazione con cui sarà svolta la rilevazione delle variabili regionali che ne definiscano la competitività.

La ricerca dell'Istat viene esposta nel capitolo del Rapporto annuale intitolato "Apertura internazionale e potenzialità di crescita del sistema produttivo italiano". Già dal titolo si possono capire gli obiettivi della ricerca: verificare se la domanda estera possa sostenere la crescita reale dell'Italia nel medio termine, ovvero rappresentare un possibile indirizzo di sbocco della crescita produttiva.

Andando a ripercorrere i risultati di questo studio, nel prossimo paragrafo sarà spiegato il ruolo della domanda estera come motore della crescita e quindi la capacità di conquistare spazi sui mercati esteri da parte del sistema produttivo italiano.

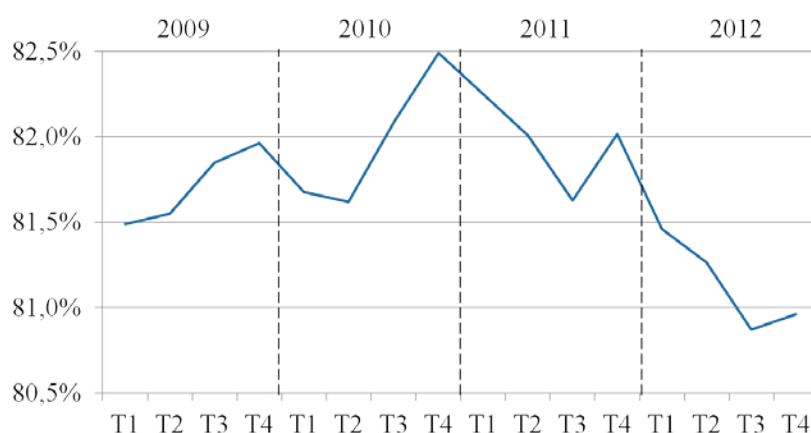
Nel paragrafo ancora successivo saranno discusse le potenzialità di crescita per l'Italia, dunque se ne esporrà il profilo competitivo in base ad un confronto internazionale su tutte le principali caratteristiche di un sistema produttivo. I risultati saranno approfonditi esaminando i fattori determinati e critici per la crescita, in quanto necessari al suo dispiegamento oppure frenanti.

1.1 Il ruolo della domanda estera

Nell'anno seguente alla crisi del 2008-2009, il principale motore della crescita per l'Italia è stata la domanda interna. Dal 2011, tuttavia, si registra un'inversione di tendenza a favore delle esportazioni: i consumi si vanno a contrarre e le esportazioni cominciano ad espandersi. Nel biennio 2011-2012, mentre i consumi interni vedono ridursi la loro quota del Pil, le esportazioni nette registrano un aumento percentuale che mostra un ribaltamento dalla prevalenza delle importazioni a quella delle esportazioni.

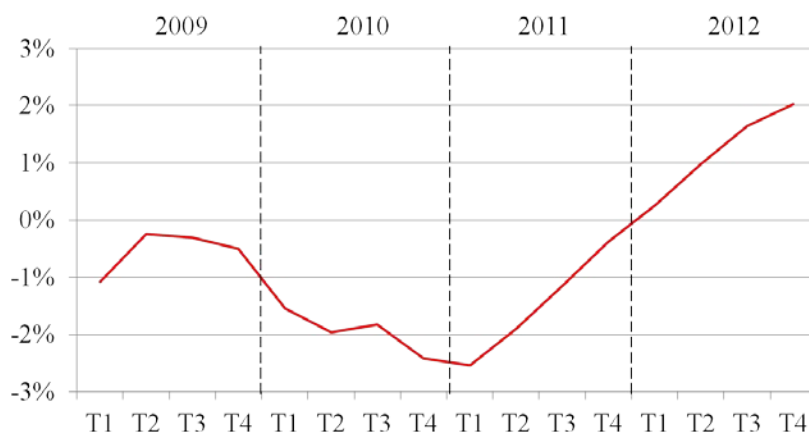
Nelle seguenti figure sono riportati questi valori registrati dall'Istat nei Conti economici nazionali delle risorse e degli impieghi. Gli anni considerati sono dal 2009 al 2012 e sono divisi in trimestri.

Fig. 1.1a – Consumi interni come percentuale del Pil (2009-2012)



Fonte: elaborazione su dati Istat

Fig. 1.1b – Esportazioni nette come percentuale del Pil (2009-2012)



Fonte: elaborazione su dati Istat

Questo scenario postumo alle difficoltà della crisi può spiegarsi, per quanto riguarda i consumi interni, tramite il clima di forte incertezza, gli effetti della politica fiscale restrittiva e naturalmente l'aggiustamento ritardato alle condizioni cicliche da parte del mercato del lavoro.

I mercati esteri si presentano così come una'alternativa per le imprese che vedono contrarsi i mercati nazionali. Possiamo interpretare questo fenomeno come un segnale di cambiamento, ovvero un'opportunità di crescita economica nel medio periodo dovuta all'azione della domanda estera.

Nei seguenti paragrafi andremo a valutare proprio questa ipotesi di trasformazione dello scenario delle opportunità di crescita attraverso varie analisi: la performance delle imprese esportatrici, ovvero la variazione delle vendite per mercati di sbocco e fascia dimensionale delle imprese, l'evoluzione del modello di specializzazione internazionale, cioè se l'Italia sta rafforzando o perdendo il suo ruolo di esportatrice dei suoi prodotti caratteristici, e infine l'apertura internazionale. Quest'ultimo punto sarà studiato in due aspetti, ovvero la sua portata rispetto ad altri paesi e i suoi effetti sulle imprese italiane.

Riguardo la portata, saranno utilizzati diversi indicatori confrontati a livello internazionale per avere un'idea generale dei margini di miglioramento che l'Italia può conseguire nel suo livello di internazionalizzazione. Riguardo il secondo aspetto, si vedranno i risultati di una ricerca attuata su un campione consistente di imprese italiane riguardo gli effetti delle diverse forme di apertura all'estero e con lo scopo di scoprire quali sono le più performanti in termini di crescita economica.

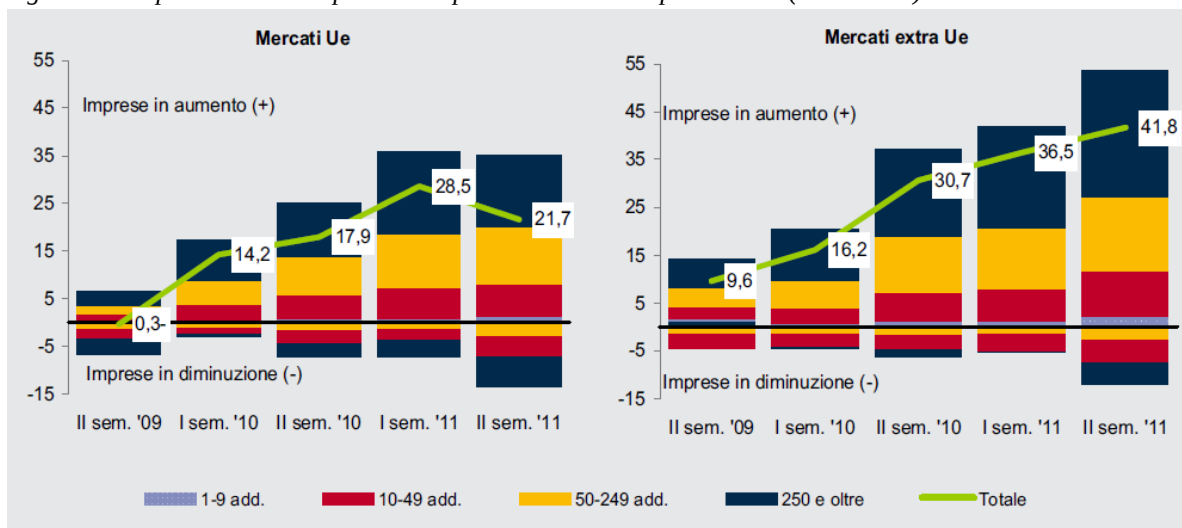
1.1.1 Performance delle imprese esportatrici

Andiamo ad analizzare il sostanziale recupero subito dalle esportazioni nette del 2011 in termini di mercato di destinazione e di dimensioni delle imprese.

Riguardo il mercato di destinazione distinguiamo i mercati europei da quelli extra-europei. Le dimensioni delle imprese sono invece classificate in 4 fasce per numero di addetti: da 1 a 9, da 10 a 49, da 50 a 249 e, infine, da 250 e oltre.

Soffermandoci su questi due aspetti, nel seguente grafico sono riportate le variazioni percentuali delle esportazioni nette nell'intervallo 2009-2011 che mostrano la crescita dell'export distinguendo il contributo positivo delle imprese in crescita da quello negativo delle imprese in calo di vendite.

Fig. 1.1.1 - Esportazioni di un panel di imprese in variazioni percentuali (2009-2011)



Fonte: Istat, rapporto annuale 2012

Confrontando la fase più acuta della crisi con il periodo più recente, ovvero dal primo semestre 2009 al secondo semestre 2011, si nota come due imprese su tre hanno aumentato le loro vendite sui mercati esteri.

La domanda estera che più prevale è quella dei mercati extra Ue dove la crescita dell'export è maggiore di quella rilevata per i mercati Ue per ogni fascia dimensionale di imprese, questo potenziale di crescita è così rivolto ai più grandi come ai più piccoli.

Un altro aspetto interessante sono le peculiarità dimensionali nei due mercati di destinazione: mentre nei mercati mondiali prevalgono le grandi imprese, nei mercati comunitari se la cavano meglio le piccole e medie imprese. Sui mercati extra Ue, infatti, la crescita del 41,8% (dovuta ai flussi di +53,6% e -11,8%) è dovuta per più della metà alle imprese di grandi dimensioni (22% su 41,8%), sui mercati comunitari la crescita del 21,7% (somma di +35,1% e -13,5%) è spiegata dalle piccole e medie imprese rispettivamente per il 4,2% e 7,8% (entrambi su 21,7%).

1.1.2 Specializzazione all'export

Come emergeva nel precedente paragrafo, la quota totale delle esportazioni è aumentata negli anni 2009-2011. Tuttavia nell'ultimo decennio (dal 2000 al 2011) i dati mostrano una progressiva riduzione delle esportazioni italiane sulle esportazioni mondiali: la quota è passata dal 3,8% nel 2000 al 3,1% nel 2011.

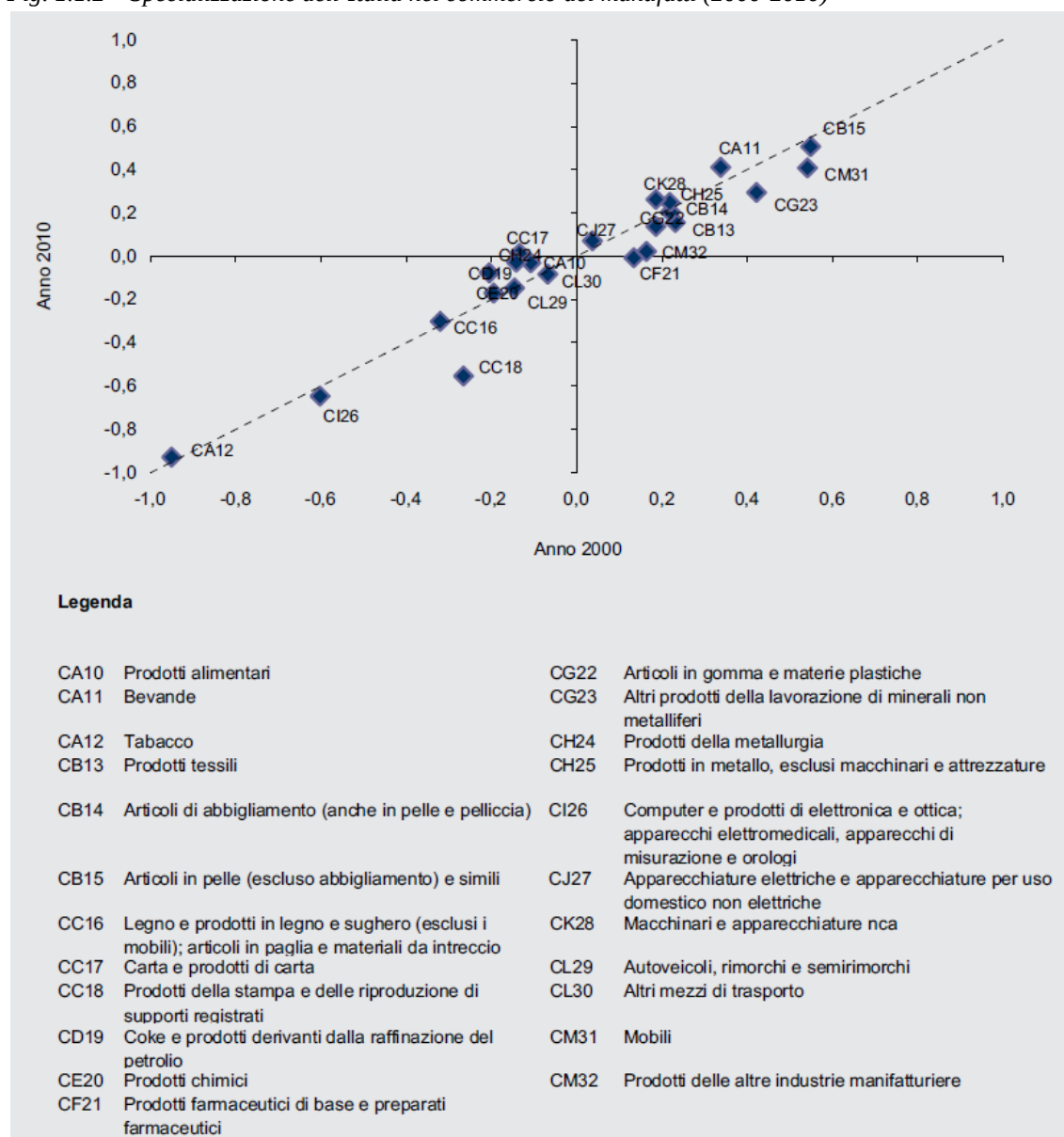
Questa performance in declino va contestualizzata con i cambiamenti dell'economia mondiale, o meglio, con l'evoluzione del modello di divisione internazionale del lavoro in cui ogni paese si specializza in una rosa di prodotti che andrà a scambiare con l'estero. Nell'analisi di come l'Italia si inserisce nelle dinamiche commerciali mondiali, possiamo notare l'esistenza di ampi margini di miglioramento nella quantità di beni da esportare a conferma dell'ipotesi che una maggiore apertura all'estero possa rappresentare una spinta alla crescita del paese.

L'analisi condotta dall'Istat si basa su un **indice di specializzazione delle esportazioni** calcolato come esportazioni nette italiane in un dato settore diviso le esportazioni mondiali nello stesso settore. I settori considerati sono tutti quelli manifatturieri della classificazione Cpa-Ateco con un livello di dettaglio a due cifre.

Calcolando tale indicatore per due istanti di tempo, è possibile verificare in quali settori è avvenuto un aumento relativo delle esportazioni nette (e quindi una specializzazione) e in quali settori è avvenuta una diminuzione relativa di queste (e quindi una despecializzazione): l'indice assume valori nell'intervallo tra -1 e +1, che indicano rispettivamente la massima despecializzazione e la massima specializzazione nel dato settore. Inoltre valori positivi indicano che le esportazioni superano le importazioni, mentre valori negativi indicano il contrario.

Nel seguente grafico si può leggere per ogni settore il valore dell'indice nel 2000 (asse delle ascisse) e nel 2010 (asse delle ordinate). Tutti i settori al di sopra dell'asse delle ascisse (ovvero con l'indicatore maggiore di 0 nel 2010) sono quelli in cui l'Italia è maggiormente specializzata, inoltre la diagonale $x = y$ fornisce un'indicazione aggiuntiva: i settori che giacciono su di essa hanno mantenuto costante il valore dell'indice, quindi non è avvenuto alcun cambiamento nelle proporzioni degli scambi tra Italia e mondo, per quelli al di sopra è aumentato e per quelli al di sotto è diminuito.

Fig. 1.1.2 – Specializzazione dell'Italia nel commercio dei manufatti (2000-2010)



Fonte: Istat, rapporto annuale 2012

I risultati mostrano la persistenza dell'Italia negli stessi settori in cui risultava specializzata ad inizio decennio, più precisamente in 11 settori su 12 ha mantenuto la prevalenza delle esportazioni sulle importazioni nel commercio mondiale. L'unico settore in cui si è despecializzata è quello dei prodotti farmaceutici (CF21) che sono stati "rimpiazzati" dai prodotti di carta (CC17) che, in questo decennio, hanno visto le esportazioni superare le importazioni. Più in dettaglio si può poi notare un rafforzamento della specializzazione nei settori delle bevande, prodotti in metallo, macchinari e apparecchiature meccaniche mentre si registra una riduzione nei settori dei prodotti tessili, prodotti delle altre industrie manifatturiere, mobili e gli altri prodotti della lavorazione dei minerali non metalliferi. La conclusione generale di questa ricerca sull'evoluzione della divisione del lavoro internazionale dimostra come l'Italia tragga profitto dagli scambi con il mondo esportando maggiormente i beni di quei settori che fanno parte della sua tradizione di *export*.

1.1.3 Internazionalizzazione Produttiva

Una caratteristica dell'evoluzione dello scenario internazionale nel decennio 2000-2010 è l'integrazione dei processi produttivi su scala globale. Questo fenomeno di intersezione delle filiere produttive a livello internazionale si manifesta in vari modi.

Ad un livello più superficiale si registra un aumento degli scambi di beni intermedi nei processi di produzione, analizzando più approfonditamente le dinamiche tra sistemi produttivi internazionali si possono poi notare altri flussi. Questi sono di carattere monetario, in quanto si effettuano investimenti da un paese all'altro, e di controllo, in quanto un paese può esercitare attività economiche in altri paesi in cui dispone i mezzi e gli impianti necessari.

La ricerca dell'Istat analizza queste tre forme di internazionalizzazione per l'Italia per dimostrare come la tendenza all'espansione internazionale delle filiere coinvolga anche il nostro paese e come vi siano, in base al confronto con gli altri stati europei, dei margini di miglioramento.

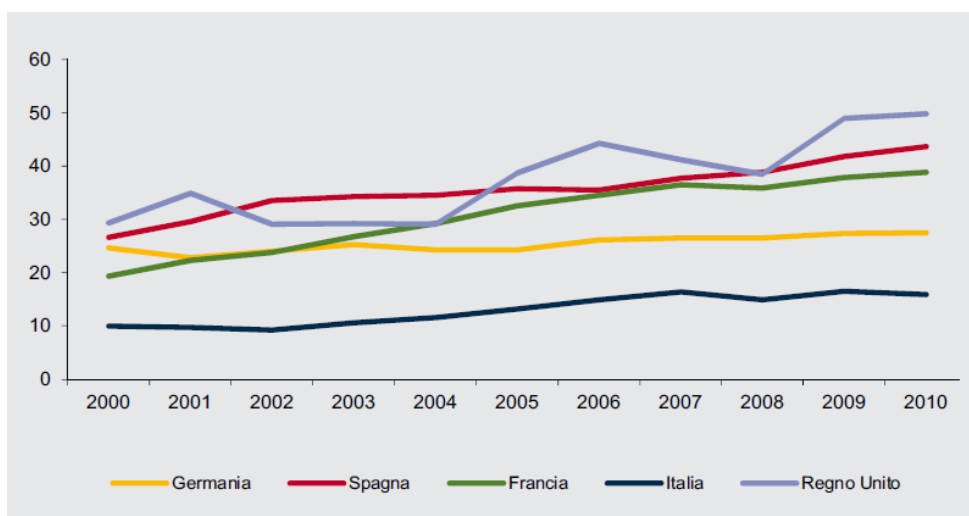
Riguardo la percentuale di beni intermedi nell'insieme dei prodotti non energetici scambiati con l'estero, i dati dell'Istat registrano un aumento sia nelle esportazioni (da 33,5% a 35,9%) che nelle importazioni (da 40,1% a 43%) nel periodo 2000-2011.

I flussi internazionali monetari, meglio conosciuti come **Investimenti diretti esteri** (Ide), sono invece acquisizioni di lungo periodo di partecipazioni in società o di immobili produttivi. Questi ultimi rappresentano un grande punto di forza per la competitività del sistema produttivo compratore in quanto consentono di delocalizzare le attività economiche all'estero risparmiando sui costi di produzione ma anche guadagnando un contatto più stretto con le dinamiche di mercato locali.

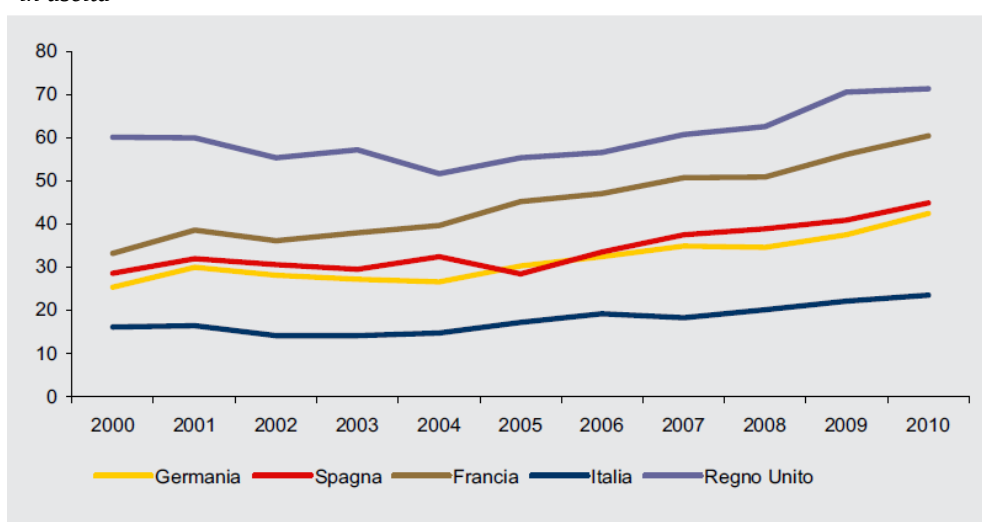
Nell'ambito degli Ide in entrata e in uscita i dati registrano, anche in questo caso, un aumento nel decennio 2000-2010: come percentuali del Pil sono aumentati rispettivamente di 5,9 e di 7,3 punti percentuali.

Il grado di internazionalizzazione italiano risulta tuttavia molto contenuto rispetto agli altri principali paesi comunitari quali Germania, Francia, Spagna e Regno Unito come mostrano i grafici seguenti.

Fig. 1.1.3 - Investimenti diretti esteri in percentuale del Pil (2000-2010)
- in entrata -



- in uscita -



Fonte: Istat, rapporto annuale 2012

La Francia, per esempio, presenta incrementi di circa 20 punti percentuali dello stock di Ide in entrata sul Pil e di 27 punti percentuali dello stock di Ide in uscita sul Pil.

Questo confronto internazionale dimostra come, potenzialmente, l'apertura internazionale sia un'opportunità di crescita economica.

I flussi di controllo, infine, si possono misurare come la percentuale di attività estere sul totale di attività nazionali, flussi in entrata, come attività svolte all'estero sempre sul totale di attività nazionali, flussi in uscita. La rilevanza delle attività delle multinazionali si può misurare in due modi: in termini di addetti controllati sul totale degli addetti, oppure in termini di valore aggiunto controllato sul totale del valore aggiunto. Ovviamente l'indicatore può essere in uscita, e quindi considerare il controllo della nazione verso l'estero, oppure in entrata, considerando il controllo estero verso la nazione. Secondo i dati dell'Istat del 2008 la quota di internazionalizzazione passiva, ovvero in entrata, ammonta al 7,4%, in termini di addetti, e al 13,2%, in termini di valore aggiunto. In entrambe le quote l'Italia viene superata da tutti i principali paesi europei: Regno Unito (rispettivamente 18,1% e 28,7%), Germania (11,7% e 20,5%), Francia (3,2% e 23,2%), Spagna (8,8% e 17,3%) e Polonia (14% e 26,6%).

Per quanto riguarda l'internazionalizzazione attiva, ovvero le attività delle multinazionali italiane all'estero sul totale delle attività italiane in termini di valore aggiunto, le percentuali sono di 12,7% per l'industria e 7,5% per i servizi. Anche in questo caso, i valori sono inferiori a tutti i maggiori paesi europei che superano i valori italiani: Finlandia, Germania, Svezia, Francia, Austria e Portogallo.

Nel paragrafo successivo sarà affrontata un'altra analisi. Mentre fino ad ora abbiamo esaminato alcune forme di internazionalizzazione e i loro bassi livelli rispetto ad altri paesi europei, in seguito specificheremo più nel dettaglio le possibili tipologie di internazionalizzazione per verificare come, nel periodo della crisi, sono state le forme più evolute a crescere maggiormente.

1.1.4 Forme di internazionalizzazione

Un'analisi svolta dall'Istat nel Rapporto sulla competitività dei settori produttivi mostra come all'aumentare del grado di internazionalizzazione produttiva si registri un miglioramento della performance in termini di valore aggiunto ed aumento occupazionale, seppur con alcune differenze settoriali.

Il campione di riferimento è di oltre 90.000 imprese italiane che operano sui mercati esteri, queste sono state studiate nel periodo 2007-2010 e nell'ultimo anno registravano un ammontare di esportazioni pari a più dell'85% delle esportazioni italiane. Ogni impresa rientra in una classe all'interno di una tassonomia appositamente costruita delle diverse modalità operative sui mercati esteri. Le classi sono sette e vanno dalle forme più elementari a quelle più complesse di internazionalizzazione:

- 1) Solo esportatori: imprese che non importano ma svolgono un'attività di esportazione verso i paesi Ue e/o verso un massimo di quattro aree geografiche extra-Ue;
- 2) Importatori di beni intermedi;
- 3) Importatori di altri beni e servizi;
- 4) Esportatori-importatori;
- 5) Global: imprese che effettuano attività sia di esportazione che di importazione, come le precedenti, in almeno 5 aree extra-europee;
- 6) Multinazionali estere: (MNE) imprese che hanno controllate estere;
- 7) Controllo estero: imprese localizzate sul territorio italiano ma controllate dall'estero;

Ogni impresa è assegnata ad una classe per ciascun anno di riferimento, inoltre se un'impresa presenta caratteristiche comuni a più classi essa sarà attribuita alla classe più elevata.

Nella tabella seguente è possibile vedere alcuni dati nonché la divisione in classi del campione di imprese.

Tab. 1.1.4 – Caratteristiche strutturali delle imprese per classi di internazionalizzazione (2010)

FORME DI INTERNAZIONALIZZAZIONE	Numero di imprese	Numero di addetti	Fatturato medio (migliaia di euro)	Dimensione media (addetti)	Produttività media (valore aggiunto per addetto)	Profittabilità media (MOL/valore aggiunto)
Controllo estero	4.261	936.749	95.817	219,8	103,9	34,8
MNE	3.133	647.232	81.524	206,6	86,0	34,8
Global	10.467	933.482	29.853	89,2	65,5	35,4
Esportatori-importatori	28.176	992.827	12.375	35,2	62,7	40,3
Importatori di beni intermedi	13.608	412.095	10.758	30,3	60,9	43,6
Importatori di altri beni e servizi	7.605	143.983	5.183	18,9	54,3	50,0
Solo esportatori	24.168	323.776	3.520	13,4	46,6	41,4
Totale	91.418	4.390.145	17.455	48,0	60,5	40,9

Fonte: Istat, rapporto sulla competitività dei settori produttivi (2013)

La classe con maggior numero di imprese è quella “esportatori-importatori” seguita da “solo esportatori”. Le due forme di maggiore internazionalizzazione, controllate dall'estero e controllanti l'estero (MNE), risultano molto meno numerose ma sono caratterizzate da due particolarità: dimensioni e produttività sopra la media. Il numero medio di addetti delle multinazionali è di 219,8, controllo estero, e di 206,6, controllanti l'estero, a fronte dei 13,4 addetti medi di un'impresa solo esportatrice. Anche la produttività media apparente, calcolata come valore aggiunto per addetto, risulta maggiore. Per quanto riguarda la profittabilità media, calcolata come margine operativo lordo su valore aggiunto, essa risulta maggiore per le imprese solo importatrici, questo *gap* rilevato per le altre forme di internazionalizzazione sarà spiegato più avanti con l'analisi della competitività del sistema produttivo italiano.

Nell'analisi settoriale condotta dall'Istat si rileva sul campione di imprese, per ciascun settore della classificazione Ateco a due cifre, la variazione percentuale di addetti e di valore aggiunto nel periodo 2007-2010. Per entrambi gli indicatori si rivela per ogni settore qual è la classe di internazionalizzazione per la categoria *top performer*, ovvero la miglior variazione percentuale, e per la categoria *bottom performer*, la peggior variazione percentuale.

I risultati sono inequivocabili: in quasi tutti i settori le classi di internazionalizzazione *top performer* sono le quattro di maggior grado (esportatori-importatori, global, MNE e controllo estero) sia per variazioni di addetti (24 settori su 25) che per variazioni di valore aggiunto (23 settori su 25), mentre in tutti i settori le classi di internazionalizzazione *bottom performer* sono le tre di minor grado (solo esportatori, solo importatori di beni intermedi e di altri beni e servizi) sia per variazioni di addetti che di valore aggiunto.

Quest'analisi ci fa dunque capire come, negli anni della crisi, all'aumentare del livello di apertura ai mercati esteri sia aumentata anche la misura di crescita economica delle imprese.

1.2 Competitività del sistema produttivo italiano

Dopo aver dimostrato, tramite i dati di questi ultimi anni, come l'apertura ai mercati esteri rappresenti un potenziale sbocco di crescita economica nel medio periodo per l'Italia, acquisisce notevole rilevanza l'analisi della competitività del sistema produttivo italiano. Questo studio consente di definire il potenziale italiano di espansione economica e di trovare quindi quei fattori determinanti per la crescita come tutte quelle cause che invece potrebbero ostacolarla.

La base di questa analisi è un confronto internazionale a livello europeo per le principali variabili economiche di un sistema produttivo. Grazie a questo *benchmarking* sarà possibile capire sotto quali aspetti l'Italia presenta dei *gap*, delle lacune che ne impediscono la piena crescita produttiva.

L'analisi dell'Istat è stata effettuata per il periodo 2000-2011 distinguendo i due sottoperiodi dalla fase più acuta della crisi internazionale, ovvero 2000-2007 e 2008-2011. La performance dell'Italia è stata paragonata ai 20 paesi europei più rilevanti in termini di contributo al Pil dell'Unione.

Le variabili considerate per definire la competitività di ciascun paese sono divise in quattro macro-aree.

In *Crescita, struttura e performance* rientra la Produttività apparente del lavoro, il Pil, il Costo del lavoro per unità di prodotto, il Tasso di natalità delle imprese e il Valore aggiunto apportato dalle grandi imprese.

Nell'area *Internazionalizzazione* rientrano gli Scambi di merci e servizi con il mondo e gli Investimenti diretti esteri (Ide) in entrata e uscita.

In *Conoscenza* rientra la spesa in Ricerca e Sviluppo, il numero di domande di Brevetti per milione di abitanti e la percentuale di Laureati.

Nei *Fattori strutturali e di contesto*, infine, si considera un indice di performance Logistica, gli Investimenti pubblici e una serie di indicatori del sistema giuridico riguardo il mercato dei prodotti, del lavoro, il numero di giorni necessari per iniziare una nuova attività economica o per l'*enforcement* dei contratti.

Riportiamo in seguito l'insieme dei dati registrati per i 20 paesi, alcuni di essi sono standardizzati per facilitare il confronto e per ogni variabile è riportata la formula di calcolo nonché il suo significato.

La graduatoria dei paesi è sulla base della crescita della produttività apparente negli anni 2000-2011.

Tab. 1.2 – Crescita, internazionalizzazione e principali indicatori strutturali per alcuni paesi dell'Ue27 nel periodo 2000-2011

PAESI	Crescita struttura e performance								Internazionalizzazione						Conoscenza			Fattori strutturali di contesto								
	Produttività del lavoro (b)			Prodotto interno lordo (c)		Costo del lavoro per unità di prodotto (d)		Nuove imprese (e)	Valore aggiunto grandi imprese (f)	Scambi di merci (g)		Scambi di servizi (h)		Investimenti diretti esteri in uscita (i)	Investimenti diretti esteri in entrata (j)	Spesa in R&S (k)	Brevetti (l)	Laureati (m)	Logistica (n)	Regolamentazione mercati dei prodotti (o)	Regolamentazione mercato del lavoro (p)	Avvio attività economica (q)	Enforcement dei contratti (r)	Investimenti pubblici (s)		
	2000-2011	2007-2011	2010	2000-2011	2007-2011	2011	2000-2011	2009	2008	2000	2011	2000	2011	2010	2010	2010	2009	2010	2010	2010	2008	2008	2011	2011	2000	2011
Romania	4,7	0,9	43,0	2,6	0,4		11,4	9,6	47,4	60,7	71,4	10,3	10,4	0,9	42,6	0,47	1,8	18,1	2,84			0	-40	3,9	5,3	
Slovacchia	3,0	1,5	74,6	3,2	2,1	68,1	1,4		45,9	123,5	160,7	19,9	14,8	3,8	57,1	0,63	8,8	22,1	3,24		0,72	4	13	2,6	2,3	
Bulgaria	2,5	2,9	41,4	2,5	0,7	83,7	4,4	17,5	36,0	76,3	110,3	30,0	22,1	3,3	100,4	0,60	1,2	27,7	2,83			4	12	3,7	3,2	
Polonia	2,1	2,2	53,9	2,5	3,9		0,0		47,3	49,2	74,8	11,5	13,3	6,6	41,5	0,74	6,8	35,3	3,44	1,68	0,95	18	278	2,4	5,9	
Repubblica Ceca	1,9	0,5	67,7	1,8	0,8	71,9	1,1	11,0	41,6	103,3	123,2	20,7	20,1	7,3	63,8	1,56	22,6	20,4	3,51	1,19	0,98	6	59	3,4	3,7	
Ungheria	1,5	0,0	60,1	0,5	-0,7	64,8	3,2	9,1	45,9	128,9	149,1	23,7	28,1	15,1	69,0	1,16	21,5	25,7	2,99	0,94	0,83	-10	-157	3,3	2,9	
Irlanda	0,8	1,4	125,7	0,8	-2,3	56,4	0,8		42,0	126,1	87,3	55,2	102,9	167,3	118,5	1,79	77,4	49,9	3,89	0,66	0,56	-1	98	3,4	3,4	
Svezia	0,8	0,5	115,4	0,9	1,2	73,0	0,0	7,1	41,4	65,7	67,5	21,0	26,3	79,6	75,5	3,42	332,0	45,8	4,08	0,95	0,94	1	-44	2,8	3,4	
Finlandia	0,3	-0,5	109,6	0,5	-0,3	74,5	0,8	9,5	47,1	64,2	59,6	13,8	19,1	56,8	35,6	3,87	215,7	45,7	3,89	0,85	0,98	0	-177	2,4	2,5	
Regno Unito	0,2	-0,7	105,3	0,2	-0,6	65,6	1,3	9,9	50,4	41,9	46,2	15,2	20,2	71,4	49,8	1,77	83,4	43,0	3,95	0,60	0,38	-1	-153	1,2	2,2	
Spagna	0,0	2,3	107,8	0,6	-0,4	67,1	0,9	7,4	32,5	46,2	45,2	15,1	15,6	45,0	43,7	1,39	31,6	40,6	3,63	0,74	1,50	14	-37	3,2	2,8	
Grecia	0,0	-1,6	77,8	-0,1	-3,3	69,4	1,3			40,3	37,2	22,9	18,3	14,0	11,5		10,6	28,4	2,96	1,76	1,37	-4	267	3,6	1,6	
Austria	-0,1	-0,3	114,7	0,3	0,9	70,3	0,2	6,9	39,9	69,4	85,0	21,3	25,3	45,0	43,4	2,76	218,4	23,5	3,76	1,06	0,97	14	-155	1,6	1,0	
Paesi Bassi	-0,1	0,0	135,8	0,0	0,4	70,8	0,8	13,2	34,2	105,5	124,6	29,1	32,6	121,4	75,7	1,83	179,5	41,4	4,07	0,69	0,98	-6	-38	3,1	3,5	
Portogallo	-0,1	0,4	65,2	-1,0	-0,7	71,5	0,8	15,0	32,1	57,6	59,4	11,3	15,4	28,0	47,8	1,59	14,3	23,5	3,34	1,04	1,45	-9	-5	4,1	2,6	
Germania	-0,2	-0,3	123,9	-0,3	0,7	66,9	-0,7	8,3	46,3	54,8	80,1	11,7	15,1	42,5	27,5	2,82	294,5	29,8	4,11	0,97	1,07	1	-158	1,9	1,6	
Francia	-0,2	0,0	133,5	-0,2	0,2	70,9	0,8	15,6		45,6	46,5	11,1	10,8	60,4	38,9	2,26	134,3	43,5	3,84	1,06	1,53	-7	-221	3,1	3,1	
Danimarca	-0,3	-0,1	120,3	-0,8	-1,0	79,6	1,5		38,3	58,5	64,1	28,5	38,0	68,5	45,5	3,06	242,6	47,0	3,85	0,76	0,75	-8	-142	1,7	2,1	
Belgio	-0,4	-0,5	135,7	0,0	0,7	77,9	0,9	4,5	40,3	123,6	131,2	29,8	36,1	84,8	82,0	1,99	143,6	44,4	3,94	1,05	1,09	-10	-47	2,0	1,7	
Italia	-1,2	-0,8	102,0	-1,0	-1,0	67,5	1,2	7,2	28,9	42,2	48,7	10,4	10,3	23,5	16,0	1,26	82,0	19,8	3,64	1,01	0,95	-8	648	2,3	2,0	
UE27 (valori medi dell'area)	0,9	0,1	100,0	1,4	-0,1	70,0	1,2		41,4	55,7	66,5	15,8	19,1	33,9	24,2	2,00	115,8	33,6	3,49					2,3	2,5	

(b) Produttività apparente del lavoro. I livelli sono misurati per ore lavorate e corretti per standard di potere d'acquisto (Spa), indice Ue27=100. Le variazioni, misurate a partire dal rapporto tra valore aggiunto e occupati, sono espresse come differenze in punti percentuali tra il tasso di crescita medio annuo della produttività di un paese e l'analogo indicatore a livello Ue27. Indice 2005=100.

(c) Prodotto interno lordo a valori concatenati. Le variazioni sono espresse come differenze in punti percentuali tra il tasso di crescita medio annuo di un paese e l'analogo indicatore a livello Ue27 per i periodi considerati.

(d) Rapporto tra costo del lavoro per dipendente e produttività apparente (Clup). I livelli sono misurati in migliaia di euro. Le variazioni sono espresse come differenze in punti percentuali tra il tasso di crescita medio annuo del Clup di un paese e l'analogo indicatore a livello Ue27. Indice 2005=100.

(e) Tasso di natalità dato dal rapporto tra numero di nuove imprese e imprese attive per l'anno considerato.

(f) Quota percentuale delle grandi imprese (250 addetti e oltre) sul valore aggiunto delle imprese non finanziarie.

(g) Apertura agli scambi internazionali di merci misurata come incidenza percentuale degli scambi commerciali di merci rispetto al Pil (prezzi correnti).

(h) Apertura agli scambi internazionali di servizi misurata come incidenza percentuale degli scambi commerciali di servizi rispetto al Pil (prezzi correnti).

(i) Investimenti diretti esteri in uscita. Incidenza percentuale dello stock rispetto al Pil su valori a prezzi correnti (per la Polonia 2009).

(j) Investimenti diretti esteri in entrata. Incidenza percentuale dello stock rispetto al Pil su valori a prezzi correnti (per la Polonia 2009).

(k) Spesa in ricerca e sviluppo in percentuale del Pil.

(l) Domande di brevetti per milioni di abitanti.

(m) Percentuale di laureati nella classe di età 30-34.

(n) Indice di performance logistica (si veda il glossario). Intervallo di variazione dell'indicatore tra 1 (peggiore performance) e 5 (migliore performance).

(o) Grado di regolamentazione del mercato dei prodotti. Indicatore relativo rispetto alla media dei paesi considerati.

(p) Grado di protezione del mercato del lavoro. Indicatore relativo rispetto alla media dei paesi considerati.

(q) Numero di giorni necessari per iniziare una nuova attività economica. Differenze assolute rispetto alla media dei paesi considerati.

(r) Numero di giorni necessari per l'enforcement dei contratti. Differenze assolute rispetto alla media dei paesi considerati.

(s) Investimenti pubblici, in percentuale del Pil.

Fonte: Istat, rapporto annuale 2012

Dalla tavola si evince facilmente la correlazione tra produttività del lavoro e crescita: per alcuni paesi dell'est Europa quali Romania, Slovacchia, Bulgaria, Polonia e Repubblica Ceca il forte aumento della produttività del lavoro nell'ultimo decennio ha infatti determinato, anche in fase di crisi, una crescita del Pil ben sopra la media dei maggiori paesi dell'Ue²⁷. In questo contesto spiccano i bassi valori di crescita della produttività in Italia che la pongono alla fine della classifica, ne consegue poi che la crescita del Pil in tutto il decennio risulti inferiore alla media.

La spiegazione di questa ridotta crescita economica si può trovare nel basso profilo competitivo italiano.

Come si spiegava all'inizio di questo capitolo, la competitività si può definire attraverso una serie di variabili economiche e dunque si può spiegare attraverso una molteplicità di cause. Ne analizziamo in seguito alcune per poi approfondirne altre nei prossimi paragrafi.

Il tasso di natalità delle imprese è correlato positivamente con la dinamica della produttività, seppur in modo contenuto e in parte influenzato dall'affidabilità dei dati (risalenti al 2009, la fase più acuta della crisi). Ciò dimostra l'importanza del contributo che le nuove imprese possono apportare all'economia, in Italia la bassa natalità delle imprese influisce sulla crescita della produttività e si può ricollegare ai fattori strutturali di contesto che analizzeremo più avanti.

L'apertura internazionale agli scambi di merci e, in misura più contenuta, agli scambi di servizi influenza positivamente la crescita della produttività per quasi tutti i paesi dell'est Europa in cima alla classifica. In tutte le forme di internazionalizzazione l'Italia presenta valori al di sotto della media europea, questa condizione potrebbe aver rappresentato un ostacolo al dispiegamento del potenziale di crescita del Paese negli anni 2000-2011: una sua rimozione, insieme al superamento di altri limiti strutturali per la competitività, potrebbe riattivare un sostenuto tasso di crescita.

Riguardo la spesa in investimenti pubblici, esiste una correlazione positiva tra i tassi di crescita reale negli anni 2000-2011 e l'intensità della spesa pubblica nel periodo iniziale nonché una correlazione positiva con la crescita della spesa pubblica e la crescita della produttività del lavoro. L'Italia è partita con valori appena nella media per poi ridurre l'incidenza degli investimenti pubblici sul Pil nel corso del decennio, perdendo così i benefici dell'ammodernamento del sistema delle infrastrutture sia materiali che immateriali.

Gli indicatori di conoscenza mostrano per l'Italia, anche in questo caso, valori al di sotto della media. Sia nell'ammontare di spesa in Ricerca e Sviluppo, nel numero di domande di Brevetti e persino nell'istruzione avanzata persistono rilevanti *gap* rispetto ai principali paesi europei.

Nel paragrafo successivo saranno analizzati più approfonditamente gli aspetti della conoscenza e dell'innovazione intesi come fattori determinanti, e dunque essenziali, della crescita. Più avanti si discuteranno anche altri aspetti del profilo competitivo italiano che invece costituiscono fattori critici per la crescita in quanto tendono a frenare il sistema economico. Questi sono la logistica, ovvero il sistema dei trasporti, la regolamentazione del mercato del lavoro, l'*enforcement* dei contratti, cioè l'efficienza della giustizia civile, e infine l'economia sommersa.

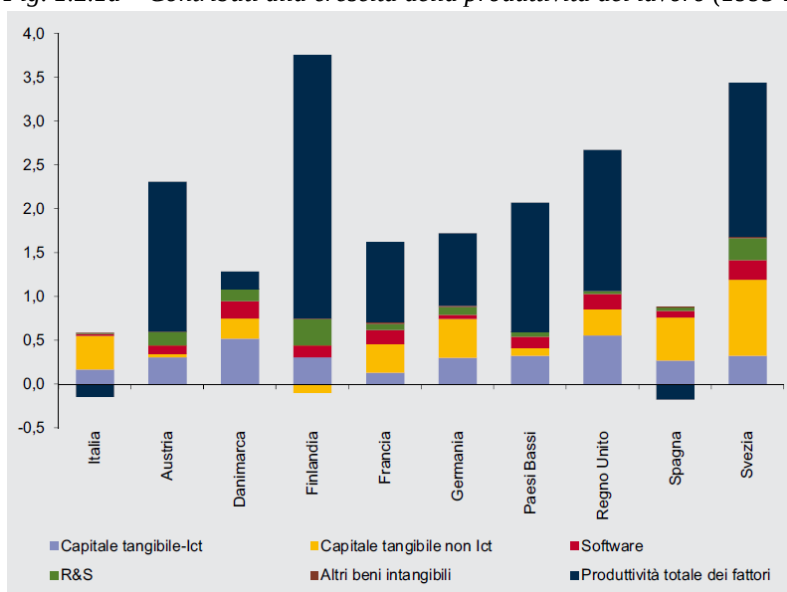
1.2.1 Conoscenza e Innovazione

Il fattore produttivo della conoscenza nonché la capacità di innovazione di un sistema produttivo (ovvero le *Dynamic Capabilities* definite dalla più recente letteratura) costituiscono elementi essenziali per la crescita economica nello scenario internazionale moderno. La globalizzazione e le moderne tecnologie di comunicazione hanno reso le rivoluzioni tecnologiche più frequenti e diffuse territorialmente, per questo motivo si definisce la conoscenza come un fattore determinante della crescita.

In questo paragrafo sarà definita più in dettaglio la posizione economica dell'Italia nell'ambito della conoscenza e dell'innovazione analizzando la composizione degli investimenti in beni immateriali, in *Information and Communication Technology* (Ict) nonché lo stock di capitale umano a confronto con i principali paesi europei.

Analizzando la variazione della produttività totale dei fattori produttivi negli anni 1995-2007 distinguendone i diversi contributi, i risultati confermano come i principali paesi europei abbiano beneficiato di aumenti rilevanti di produttività tutti correlati ad aumenti nella spesa in investimenti immateriali e materiali. Nei primi rientrano la spesa in Software, quella in Ricerca e Sviluppo (R&S) e in altri beni intangibili, nei secondi si distinguono le tecnologie Ict da quelle non-Ict. Il seguente grafico mostra le variazioni percentuali della produttività totale dei fattori produttivi nonché della spesa nei beni tangibili e intangibili della conoscenza.

Fig. 1.2.1a – Contributi alla crescita della produttività del lavoro (1995-2007)



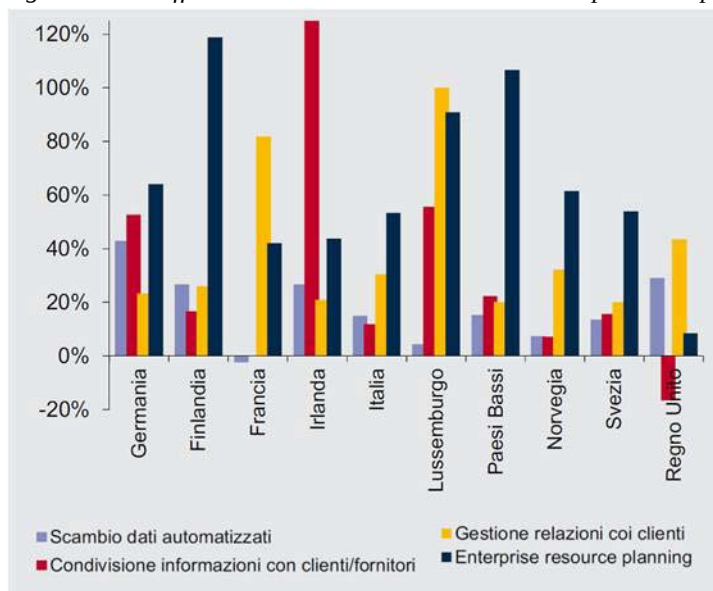
Fonte: Istat, rapporto annuale 2012

In questo intervallo di tempo la crescita media annua è stata del 2,2%, l'Italia registra tuttavia un tasso medio annuo dello 0,4% con contributi negativi della produttività totale dei fattori, come nel caso della Spagna, mentre i risultati migliori sono della Finlandia e della Svezia. Ma ciò che più spicca per l'Italia è un aumento quasi trascurabile della spesa in Software e R&S che la distingue dagli altri paesi europei che, invece, hanno saputo cogliere l'importanza di questa nuova fonte di crescita.

La stessa analisi è stata ripetuta per il periodo 2007-2010 mostrando tassi di crescita negativi per la produttività del lavoro in quasi tutti i paesi considerati.

Gli investimenti in Ict possono essere distinti, con un maggiore grado di dettaglio, in diverse aree tecnologiche. Un progetto europeo denominato Esslimit, al quale hanno partecipato gli Istituti nazionali statistici di 15 paesi europei, considera la diffusione di questi strumenti Ict nelle imprese a rapida crescita, rispetto alle altre, nel periodo 2007-2009. Le imprese a rapida crescita sono definite in ambito europeo secondo i seguenti parametri: almeno 10 addetti e un tasso di crescita del fatturato (o degli addetti) superiore al 20% annuo per almeno tre anni. Gli strumenti di Ict rientrano in quattro aree operative: lo scambio di dati automatizzati, la condivisione delle informazioni con clienti e fornitori, la gestione delle relazioni con i clienti (*Customer relationship management*) e l'*enterprise resource planning*, ovvero sistemi informatici che integrano tra loro la gestione delle *operations*, le vendite, gli acquisti, la gestione del magazzino e la contabilità. Il seguente grafico mostra per alcuni paesi europei la percentuale di questi quattro strumenti Ict nelle imprese ad alta crescita sul totale delle imprese.

Fig. 1.2.1b – Diffusione di alcuni strumenti Ict nelle imprese a rapida crescita rispetto alle altre (2007-2009)

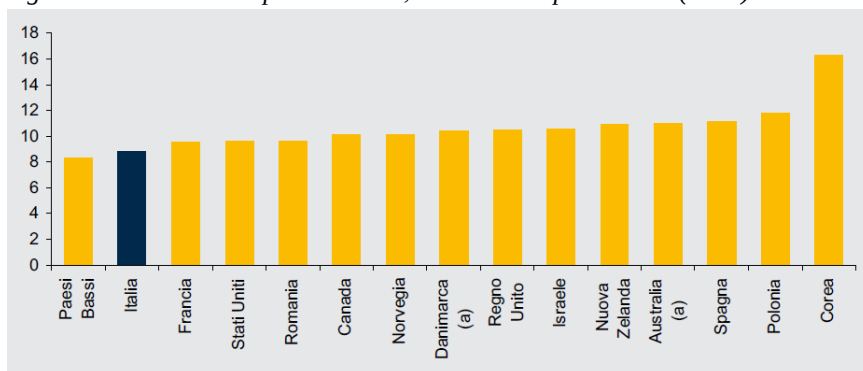


Fonte: Istat, rapporto annuale 2012

In Italia l'incidenza di tutti questi strumenti nelle imprese ad alta crescita è più bassa rispetto alle economie europee più avanzate: in quasi tutti i paesi europei si rileva almeno un ambito dell'Ict più radicato nelle imprese altamente innovative che contribuisce maggiormente alla loro capacità di sviluppare nuova conoscenza.

Per analizzare lo stock di capitale umano nel nostro sistema produttivo, a confronto con quello di altri paesi nel mondo, l'Ocse ha utilizzato la metodologia del *Lifetime labour income approach*: la popolazione viene suddivisa in gruppi sulla base di alcune variabili quali l'età, il sesso e il livello di istruzione per poter stimare per ogni gruppo il flusso di redditi futuri (e scontati al valore attuale) e ponderarlo per le probabilità di sopravvivenza. Si ottiene in questo modo una misura monetaria dello stock di capitale umano che supera di numerose volte l'ammontare del Pil. Questo valore è stato calcolato per il 2006 riferito a quattordici paesi nel mondo, nel seguente grafico è rappresentato in rapporto al Pil del paese.

Fig. 1.2.1c – Stock di capitale umano, incidenza rispetto al Pil (2006)



Fonte: Istat, rapporto annuale 2012

La posizione dell'Italia in questa classifica non è favorevole in quanto nella maggior parte dei casi il rapporto tra capitale umano e Pil è di 10 (oltre 16 per la Corea) mentre nel nostro caso è di appena 8,8.

Questo *gap* di capitale umano, come le lacune negli investimenti in beni tecnologici tangibili ed intangibili, condiziona negativamente le capacità del nostro Paese di crescita e di incremento della produttività nel medio periodo.

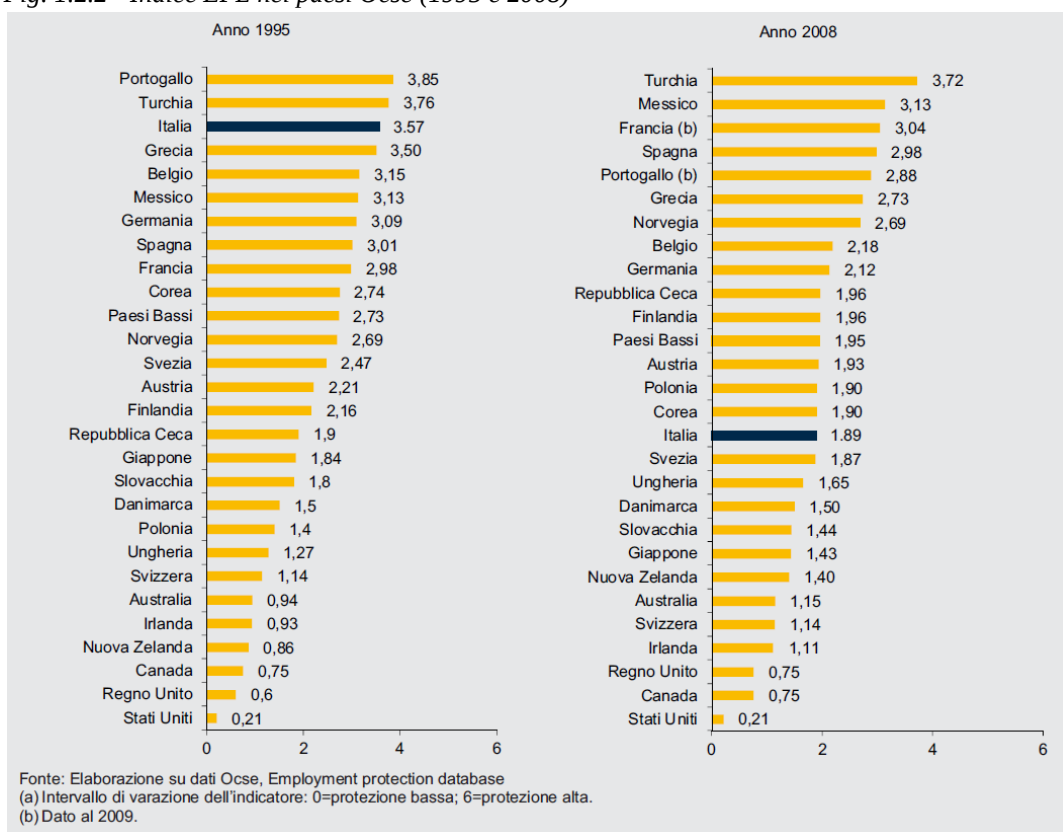
1.2.2 Regolamentazione del mercato del lavoro

La rigidità del complesso normativo del mercato del lavoro può ritenersi una delle cause che impediscono il pieno dispiegamento delle potenzialità economiche dell'economia italiana. Il grado di rigidità è misurato dall'Ocse con un indice sintetico chiamato *Employment Protection Legislation (EPL)*.

Questo indice è dato dal prodotto di tre indicatori, ognuno con una sua ponderazione e riferito ad un'area lavorativa: licenziamenti dei lavoratori a tempo indeterminato, licenziamenti collettivi e regolamentazione delle forme di lavoro temporaneo. Ogni indicatore può assumere un valore compreso tra 0 e 6, ad alti valori corrisponde una minore flessibilità per le imprese e quindi una maggiore protezione del lavoratore.

Il calcolo di questo indice è stato effettuato, per l'Italia come per molti altri paesi nel mondo, nel 1995 e nel 2008. All'inizio il nostro paese si collocava tra i paesi Ocse con la maggior regolamentazione, al terzo posto e preceduto da Portogallo e Turchia, per poi collocarsi dopo più di un decennio al sedicesimo posto.

Fig. 1.2.2 - Indice EPL nei paesi Ocse (1995 e 2008)



Fonte: Istat, rapporto annuale 2012

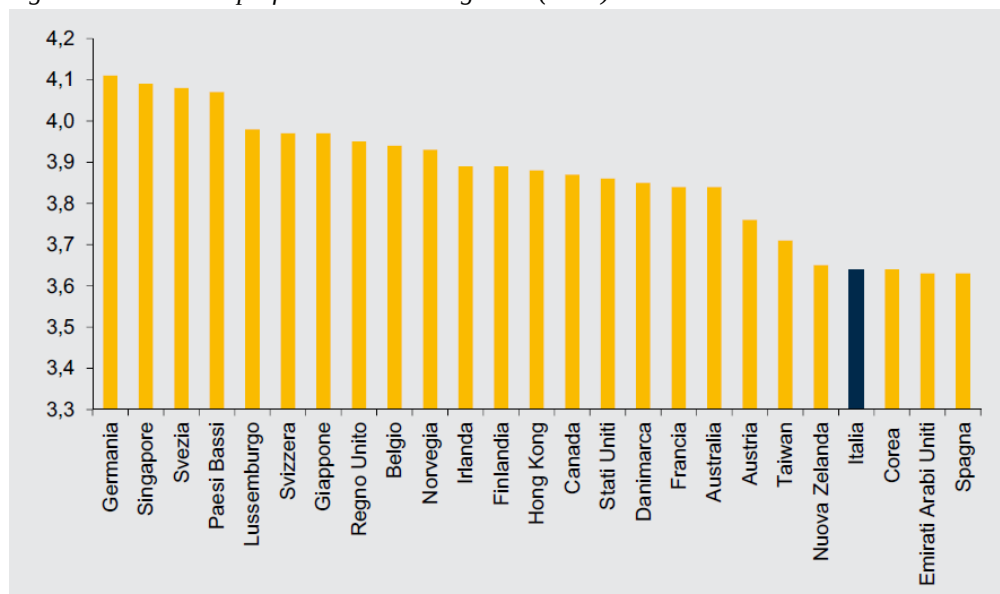
Questa riduzione dell'indice è dovuta tuttavia alla deregolamentazione della sola area relativa alle forme di lavoro temporaneo, gli indicatori delle altre due aree sono rimasti infatti costanti. Tra le riforme che hanno permesso una maggiore flessibilità ricordiamo il pacchetto Treu del 1997 e il decreto legislativo n. 368 del 2001 in attuazione di una direttiva comunitaria, infatti questo processo di flessibilizzazione è avvenuto nello stesso arco temporale anche in altri paesi europei. Per quanto riguarda l'Italia, la protezione dei lavoratori a tempo indeterminato non risulta eccessivamente rigida rispetto agli altri paesi considerati (sia nel 1995 che nel 2008), come invece risulta tale la regolamentazione dei licenziamenti collettivi che presenta il valore dell'indicatore più alto di tutti i paesi considerati (anch'esso osservato sia nel 1995 che nel 2008).

1.2.3 Sistema dei trasporti

L'efficienza del sistema logistico di un paese è stata misurata dalla Banca Mondiale attraverso il *Logistic Performance Index* (LPI) che considera sei diverse componenti essenziali per la rete logistica: certezza e rapidità delle procedure doganali, qualità delle infrastrutture materiali e immateriali legate al commercio e al trasporto, effettiva concorrenza nei prezzi di trasporto, efficienza dei servizi legati al trasporto, tracciabilità e puntualità delle operazioni di trasporto e di consegna (Rapporto annuale dell'Istat 2012).

Questo indice è stato calcolato per ben 155 paesi su dati del 2010, nella figura di seguito è riportata la classifica dei primi 25 stati dove l'indice assume valori tra 1 (performance peggiore) e 5 (performance migliore).

Fig. 1.2.3 – Indice di performance nella logistica (2010)



Fonte: Istat, rapporto annuale 2012

In questa classifica della Banca Mondiale l'Italia si pone al ventiduesimo posto. Poiché il sistema logistico considera tutte le possibili forme di trasporto, ne analizziamo brevemente le singole modalità.

Per quanto riguarda la modalità stradale, possiamo misurarne l'incidenza sul sistema logistico complessivo attraverso uno studio dell'Eurostat. Questo si basa su un indicatore sintetico che considera l'intensità di trasporto, intesa come l'incidenza sul Pil delle tonnellate per chilometro via terra ed acqua.

Osservando i valori registrati nel periodo 2000-2009 per l'Italia e l'Ue15, risulta come l'Italia rientri nella media europea dell'intensità media dei trasporti, tuttavia il suo sistema di trasporti è maggiormente vincolato alla modalità stradale: intorno al 90%, rispetto alla media Ue del 80% circa.

Confrontando poi le imprese di trasporto italiane con quelle europee, esse risultano più numerose ma di minori dimensioni: nel 2009 il numero di imprese di trasporto era il doppio di quelle della Germania, Francia o Regno Unito e il numero medio di addetti per impresa era circa la metà di quello europeo. Questa può essere una prima spiegazione dell'inefficienza del nostro sistema logistico, ovvero la perdita di economie di scala.

Analizzando il trasporto ferroviario di merci, la maggior quota di trasporto ferroviario nell'Ue15 nel 2010 è quella della Germania (42%) seguita da Spagna (11,7%), Svezia (9,2%), Austria (7,8%) e infine dall'Italia (7,3%). Negli anni 2002-2010, mentre l'Italia ha registrato una diminuzione del traffico ferroviario di circa l'8,3%, la Germania lo ha aumentato di circa il 37%.

Concludendo con il trasporto marittimo, consideriamo la classifica Eurostat dei paesi europei per maggior movimentazione di merci: dal 2002 l'Italia era al secondo posto preceduta dal Regno Unito, a partire dal 2008 è scesa al terzo posto superata sempre dal Regno Unito che ha visto passare in testa i Paesi Bassi. Nella classifica dei maggiori porti d'Europa, per quantità di merce imbarcata e sbarcata, risulta come in tutto il periodo 2001-2010 sia Rotterdam (Paesi Bassi) il porto più movimentato nonché il cuore degli scambi internazionali via mare. Nell'ultimo anno del periodo considerato i maggiori porti italiani,

Genova e Trieste, sono al 15° e 16° posto in quanto hanno registrato in questo intervallo un calo del traffico rispettivamente del 4% e 9,3% mentre i porti di Rotterdam, Anversa (Belgio) e Amburgo (Germania), i primi tre in classifica, hanno aumentato il traffico rispettivamente del 33,4%, 39,4% e 26%.

Questa classifica mostra come lo scenario dei porti d'Europa è composto da poche postazioni nei Paesi Bassi, Belgio e Germania che coprono buona parte del traffico via mare e molte più postazioni in paesi quali Italia, Francia e Spagna. Anche in questo caso si può ipotizzare una perdita di efficienza delle economie di scala.

1.2.4 Giustizia civile

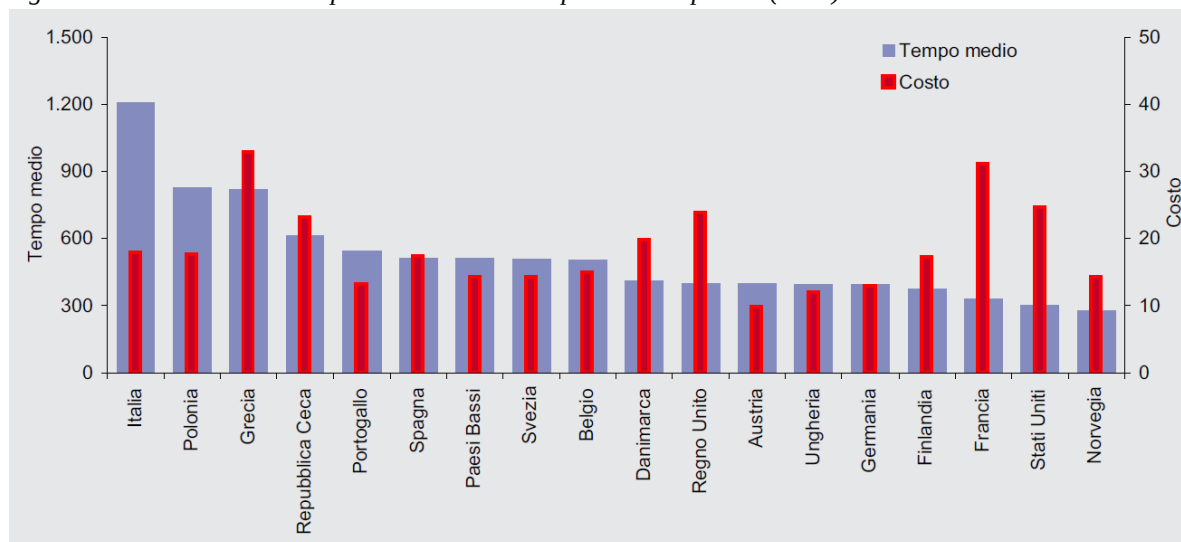
L'efficienza della giustizia civile è uno degli aspetti del sistema produttivo italiano che, secondo i dati dell'Istat, sembra necessitare maggiormente di un intervento al fine di migliorarne il funzionamento.

Un confronto con gli altri paesi europei risulta particolarmente delicato a causa della difficoltà di rendere omogenee basi di dati riferite a sistemi giuridici così differenti ed eterogenei in numerosi aspetti.

L'analisi che prenderemo in considerazione è quella svolta dalla Banca Mondiale con il progetto *Doing Business* che pone a confronto oltre 120 paesi nel tentativo di misurare le differenze organizzative riguardo i sistemi giudiziari. Ogni paese giuridicamente evoluto ricerca la correttezza del giudizio garantendo ad entrambe le parti del processo civile il diritto di presentare al giudice tutte le prove ed argomentazioni per far valere le proprie ragioni. Questo sistema si manifesta in tempi di ricerca delle prove nonché in tempi di analisi di queste, dunque all'aumentare della possibilità di difesa garantita alle parti aumentano i costi di funzionamento del sistema, se poi si vuole rendere più rapido questo funzionamento allora i costi dovranno essere ulteriormente incrementati. In questo inevitabile *trade-off*, ogni paese adotta un sistema procedurale che riflette la miglior combinazione di queste dimensioni per il suo contesto nazionale.

Confrontando l'Italia con gli altri paesi europei, confronto consentito dal fatto che i sistemi giuridici europei offrono tutti un medesimo valore di correttezza del giudizio, emergono grandi differenze, mostrate nel seguente grafico, sia in termini di costi sostenuti dai privati, misurato come percentuale del valore della causa, che in termini di tempo, in giorni, dei processi civili legati ad inadempienza contrattuale.

Fig. 1.2.4 – Durata media dei processi civili e costi processuali privati (2011)



Fonte: Istat, rapporto annuale 2012

In questa classifica per durata media dei processi, l'Italia è tristemente al primo posto, confrontandola con il paese con i costi processuali minori (Austria) i cittadini italiani aspettano un tempo medio tre volte maggiore sostenendo costi pari a quasi il doppio.

I dati raccolti forniscono tuttavia una risposta all'inefficienza del sistema giuridico italiano: nel 2008 l'Italia è al quinto posto tra i paesi europei con i maggiori livelli di spesa pubblica per la giustizia civile mentre il numero di magistrati è in linea con la media europea. Dunque il problema non è la quantità di risorse impiegate quanto la loro amministrazione.

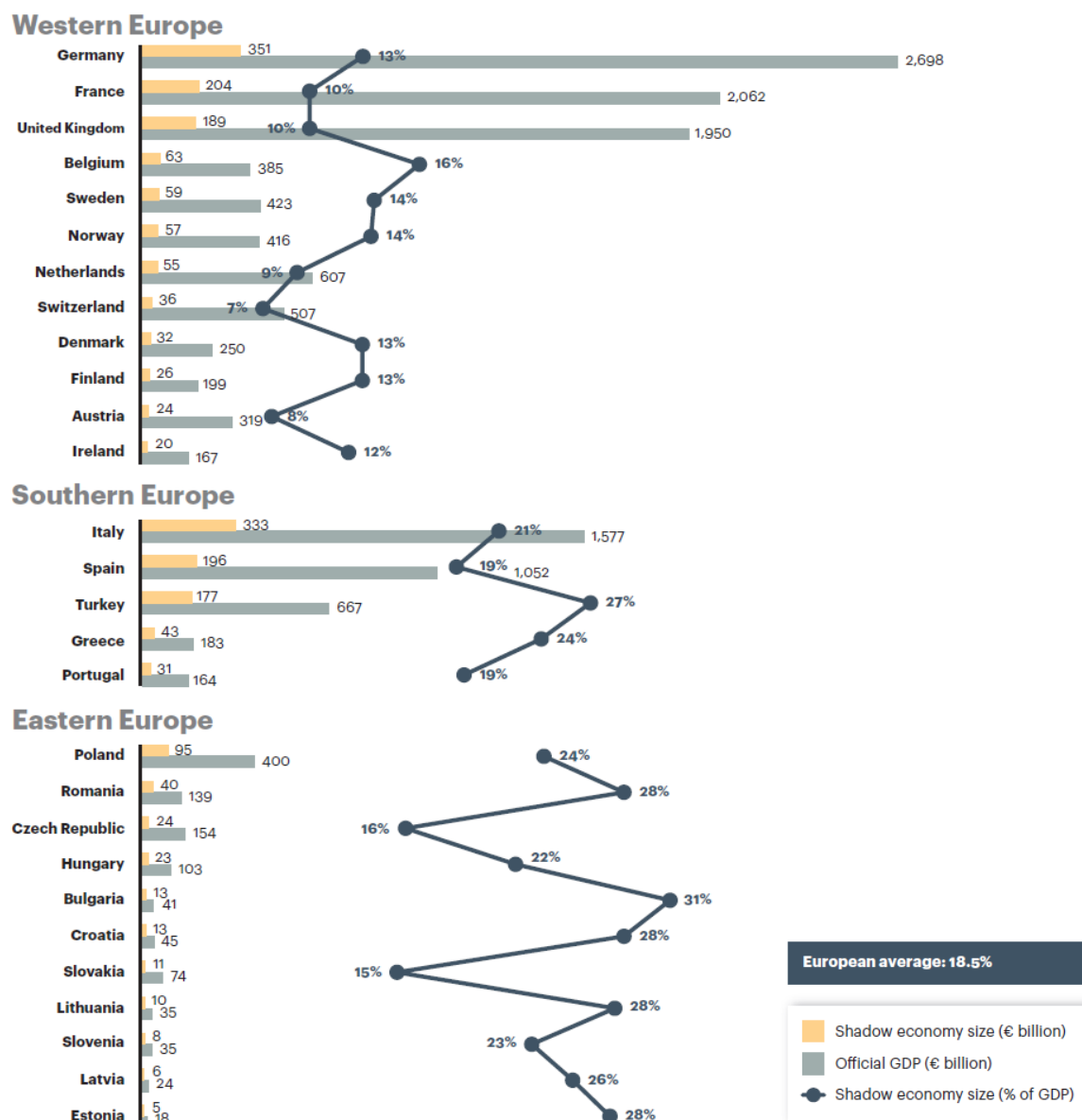
Un'analisi econometrica svolta su dati dell'Istat e del Ministero della giustizia porta a concludere che, anche in questo ambito, la perdita di efficienza è dovuta al mancato sfruttamento di economie di scala. L'eccessiva frammentazione dei tribunali impedisce infatti la specializzazione delle attività dei magistrati, i dati dimostrano inoltre come l'introduzione del giudice unico (1998) abbia determinato un recupero dell'efficienza e dal confronto europeo risulta che il numero medio di cittadini italiani serviti da una stessa corte di prima istanza sia circa la metà rispetto a Francia, Germania e Regno Unito.

1.2.5 Economia sommersa

L'Italia si caratterizza da molti anni per una elevata quota di economia sommersa rispetto ai principali paesi dell'Unione. Il rapporto *"The shadow economy in Europe"* (2013) redatto da Visa, A.T. Kearney e Friedrich Schneider riporta delle interessanti stime riguardo l'ordine di grandezza dell'economia sommersa (*shadow economy*) in relazione al Pil (*Gross Domestic Product*) di 28 paesi europei.

La percentuale media europea dell'economia sommersa sul Pil è del 18,5%, l'Italia la sorpassa con un valore stimato del 21%. Per quanto vi siano altri paesi, soprattutto in Europa dell'est, che superino vistosamente questa percentuale, sta di fatto che l'ammontare dell'economia sommersa in Italia sia di grande entità, superata di poco dal sommerso tedesco (ricordando tuttavia che la Germania registra il Pil più elevato in Europa) e addirittura più grande dell'intera economia austriaca.

Fig. 1.2.5 – Economia sommersa come percentuale del Pil in Europa (Marzo 2013)



Fonte: *The shadow economy in Europe, 2013*

Ciò che rende molto deleterio questo valore così alto è la distorsione del posizionamento competitivo dell'economia italiana sul mercato internazionale. Infatti, il gettito di imposte che viene perso porta lo Stato ad aumentare la pressione fiscale sulle imprese che operano in condizione di regolarità per raggiungere gli obiettivi di bilancio, manovra che in questo periodo di ripresa grava particolarmente sulle imprese.

Per quanto riguarda, invece, le imprese irregolari, esse presentano sul mercato condizioni di costo inferiori alla norma che alterano il grado di concorrenza. Inoltre, la possibilità di operare in nero rappresenta un disincentivo alla crescita dimensionale nonché alla crescita della produttività, in altre parole persistono deboli stimoli all'innovazione, al miglioramento dei fattori organizzativi e ai nuovi investimenti. Ciò risulta particolarmente problematico in quanto, come analizzato precedentemente, le ridotte dimensioni delle imprese italiane costituiscono spesso un limite per l'espansione nei mercati internazionali.

CAPITOLO 2

I DATI DELLE REGIONI ITALIANE: DELINEARE LA COMPETITIVITÀ

L'analisi della competitività delle regioni italiane prende spunto da quella, di carattere internazionale a livello europeo, dell'Italia condotta dall'Istat nel Rapporto annuale 2012 che funge da base logica per le argomentazioni e metodologica per la rilevazione dei dati.

Le variabili scelte per delineare i profili competitivi di quelli che saranno i gruppi di regioni appartengono alle stesse macro-aree già citate nella tabella 1.2, ovvero i fattori di crescita reale, struttura e performance, il grado di apertura internazionale, gli investimenti in conoscenza e innovazione e, infine, i fattori strutturali di contesto. Riguardo questi ultimi, poiché la maggior parte di essi sono stabiliti a livello nazionale, ci occuperemo esclusivamente dell'ammontare di investimenti pubblici. Per quanto riguarda le altre sotto variabili all'interno delle macro-aree, esse saranno il più possibile simili a quelle proposte dall'Istat.

Grazie a questi aspetti che andranno a definire i profili competitivi regionali, sarà possibile effettuare la *cluster analysis* per trovare una partizione delle regioni italiane in gruppi che condividano un medesimo livello di competitività, ovvero una particolare combinazione delle variabili considerate.

Definito lo scopo della ricerca, in questo capitolo saranno esposti gli enti dai quali sono stati rilevati i dati nonché l'insieme di informazioni raccolte suddivise per variabile.

2.1 Le Basi di dati

La nostra ricerca inizia, come ogni studio statistico, con la rilevazione dei dati e dunque il primo gradino da affrontare è la scelta delle fonti di dati.

Ciò che determina il valore di una fonte di dati è il suo grado di autorevolezza, questo si può misurare a sua volta in termini di veridicità delle informazioni prodotte nonché in termini di comparabilità rispetto ai dati di altre fonti e ai dati precedentemente prodotti dalla fonte stessa.

La valutazione di queste due componenti avviene in base a molti fattori riguardo l'ente di produzione dei dati: la disposizione di apparati che effettuano attività statistiche in modo "robusto e stabile", ovvero coinvolgendo un'ampia organizzazione di persone e di sedi residenti in tutti i luoghi oggetto della ricerca; il rispetto di regole e parametri di qualità concordati, di valenza sia formale (nomenclature, classificazioni, simboli) che sostanziale (definizioni, procedure, metodologie) e di portata sia nazionale che internazionale; la trasparenza delle metodologie e dei sistemi utilizzati; la regolarità e professionalità nella produzione dei dati.

Le fonti scelte per questa ricerca a livello regionale sono tre: l'Istat, la Unioncamere e l'ICE.

Per questi enti, riportiamo in seguito alcune informazioni generali a sostegno della loro autorevolezza e, per conferire maggiore realtà a queste istituzioni, anche la locazione della sede centrale. Presentiamo infine per ognuno di essi i documenti ottenuti dai siti internet, spiegando per questi quale ambito informativo della nostra ricerca hanno contribuito a delineare.

2.1.1 Istat

L'Istituto nazionale di statistica è un ente di ricerca pubblico che nasce in Italia nel 1926. Esso è il principale produttore di dati statistici a supporto di vari utilizzatori, dal 1989 assume la funzione di coordinatore all'interno del Sistema statistico nazionale (*Sistan*) che riunisce diversi enti e amministrazioni pubbliche a livello nazionale e locale.

Fig. 2.1.1 – Il logo e la sede centrale dell'Istat a Roma, via Cesare Balbo



Fonte: sito dell'Istat

I punti di accesso alle informazioni offerte dall'Istat sono gli uffici e gli sportelli informativi, il Laboratorio per l'Analisi dei Dati Elementari (Adele), la biblioteca e i Centri di Informazione Statistica.

Attraverso il sito web, è possibile accedere ad una immensa mole di dati, in modo assolutamente gratuito, riguardo sia le informazioni statistiche che le metodologie utilizzate.

L'ente mette a disposizione le “Banche dati” organizzate in: “Temi trasversali”, “Economia”, “Demografia e società” e “Censimenti”. Le “Tavole di dati” sono prodotte a cadenza irregolare al termine di una indagine, i “Microdati” sono collezioni di dati utilizzate per alcune indagini pubbliche, l'accesso ad essi richiede una registrazione o uno scopo specifico della ricerca che su di essi si vuole effettuare. Vi è poi una grande quantità di pubblicazioni scaricabili in formato pdf: “Comunicati stampa” in diversi argomenti e diffusi secondo un calendario, le “Edizioni chiave” che raccolgono e sintetizzano informazioni di carattere trasversale e le “Pubblicazioni scientifiche” come la Rivista di Statistica Ufficiale e gli Istat Working Papers.

Tra le “edizioni chiave” ricordiamo le pubblicazioni annuali quali l'Annuario statistico italiano, che espone l'intero patrimonio statistico pubblico, il Rapporto annuale, un'analisi sulle trasformazioni e le problematiche emergenti del paese, e Noi Italia, che offre indicatori e confronti europei o regionali in tutti gli aspetti del paese. Tra le altre pubblicazioni editoriali ricordiamo il Rapporto sulla competitività dei settori produttivi dove è stata svolta l'analisi sulle forme di internazionalizzazione delle imprese italiane i cui risultati sono stati esposti nel paragrafo 1.1.4.

Nella sezione “Strumenti” del sito web è possibile consultare le varie definizioni e classificazioni, a livello sia nazionale che internazionale, utilizzate dall'Istat nelle sue indagini, i principi guida che conferiscono il carattere dell'affidabilità alle sue ricerche e

infine le metodologie statistiche, nonché i software utilizzati, presenti anche in formato didattico e informativo nella sezione “Informazioni per studenti e docenti”.

I dati delle regioni italiane che provengono dall’Istat sono stati rilevati dalla sezione “Conti economici regionali” (Database conti regionali) e dalla sezione “Banca dati di Indicatori territoriali per le politiche di sviluppo” (Indicatori territoriali per tema: Demografia d’impresa). Questi hanno delineato il profilo competitivo regionale a livello di fattori di crescita, struttura e performance (ad esclusione del valore aggiunto delle grandi imprese) e a livello di fattori strutturali di contesto, ovvero l’ammontare di investimenti pubblici in quanto unico elemento considerato in questa area competitiva.

2.1.2 Unioncamere

L’Unione italiana delle Camere di commercio, industria, artigianato e agricoltura è l’ente pubblico fondato nel 1901 con lo scopo di rappresentare, coordinare e supportare l’intero Sistema camerale italiano, ovvero le Camere di commercio di tutte le provincie.

Fig. 2.1.2 – Il logo e la sede centrale dell’Unioncamere a Roma, piazza Sallustio



Fonte: sito di Unioncamere

La funzione del Sistema camerale è la promozione economica delle imprese nell’ambito locale in cui si trovano, le attività che svolge riguardano l’orientamento al lavoro e alle professioni, progetti di avvicinamento scuola-lavoro, la formazione manageriale, la diffusione di informazioni economiche, la semplificazione delle procedure per avviare e svolgere attività economiche, la tutela dei diritti di proprietà industriale, il supporto dell’internazionalizzazione, dell’innovazione e tante altre attività. In tutte queste, l’Unioncamere supporta le Camere di commercio, fornisce servizi di consulenza e assistenza in ambiti quali l’interpretazione della normativa e la contabilità, gestisce un fondo di perequazione (affinché tutte le Camere dispongano di uguali risorse) e cura il raccordo con le organizzazioni imprenditoriali, dei lavoratori e dei consumatori. Attraverso il sito di Unioncamere è possibile conoscere nei dettagli tutte le attività svolte dall’intero Sistema camerale e scaricare tutta la documentazione, i comunicati e le indagini statistiche condotte dall’organizzazione.

La rete degli uffici “studi e statistica” delle Camere di commercio è Starnet, questo sistema informativo è stato fondato nel 2000 all’interno del progetto più ampio di organizzare e rendere fruibile l’intero patrimonio informativo del Sistema camerale a qualsiasi utilizzatore ed anche favorire il confronto e la condivisione all’interno del sistema.

Sul sito di Starnet è disponibile l’appendice statistica al Rapporto Unioncamere 2012, essa si compone di ben 150 tavole di dati aggregati a livello nazionale e suddivise in 18 tematiche. Da questo documento sono stati ricavati i dati regionali riguardo l’economia della conoscenza e il valore aggiunto per fascia dimensionale delle imprese.

2.1.3 ICE

L’ICE-Agenzia per la promozione all’estero e l’internazionalizzazione delle imprese italiane è stata istituita nel 2011 come ente di diritto pubblico sottoposto ai poteri di indirizzo e vigilanza del Ministero dello sviluppo economico, del Ministero degli affari esteri e del Ministero dell’economia e delle finanze.

Fig. X – Il logo e la sede centrale dell’ICE a Roma, via Liszt



Fonte: sito di ICE

La sua funzione principale è quella di favorire l’apertura all’estero da parte delle imprese italiane, soprattutto le piccole e medie imprese, attraverso attività di consulenza, la promozione di attività economiche, la produzione di informazioni sul Paese, sulle opportunità commerciali, sulle gare internazionali e sugli investimenti esteri in entrata o in uscita dall’Italia e, infine, la rappresentanza diplomatica all’estero.

Nello svolgimento di tali attività da parte della sede di Roma, l’Ufficio di Milano e della rete nel mondo, l’ICE-Agenzia opera in sinergia con le regioni, le Camere di commercio, le organizzazioni imprenditoriali e tutti i soggetti pubblici o privati interessati.

Attraverso il sito web è possibile scegliere i servizi di consulenza per espandere il proprio *business* all’estero (solo alcuni sono soggetti a tariffazioni), accedere alle pubblicazioni di studi di mercato, settoriali e tematici ed inoltre sono disponibili dati statistici suddivisi in cinque aree.

I servizi di consulenza vanno dalle informazioni di guida generali sul mercato estero, in base al prodotto in questione, alle indagini di mercato, la ricerca di clienti e partner esteri con l’organizzazione di incontri d’affari, la ricerca di informazioni riservate su imprese estere nonché la promozione di eventi promozionali o le inserzioni pubblicitarie sui media esteri.

Le cinque aree della sezione “Ricerca e Statistiche” sono: Dati sugli scambi di merci, Dati sugli Investimenti Diretti Esteri in Italia, Tariffe Doganali, Dati Macroeconomici e Schede Prodotto. Grazie alle prime due aree è stato possibile ottenere le informazioni e i dati statistici riguardanti il grado di apertura internazionale delle regioni italiane.

2.2 Le variabili considerate

L’analisi della competitività delle regioni italiane si può ora delineare scendendo ad un maggiore livello di dettaglio. Nel paragrafo 2.1 si spiegava come le macro-aree economiche da considerare per ogni regione fossero la crescita reale, l’apertura internazionale, la conoscenza e la spesa pubblica. Queste aree possono essere definite attraverso vari indicatori che andremo a riprendere quasi totalmente dall’analisi dell’Istat nel Rapporto annuale 2012, ricordando tuttavia che quest’ultima si sviluppa a livello nazionale mentre, in questa sede, ci occuperemo di un livello regionale e quindi adotteremo alcuni cambiamenti.

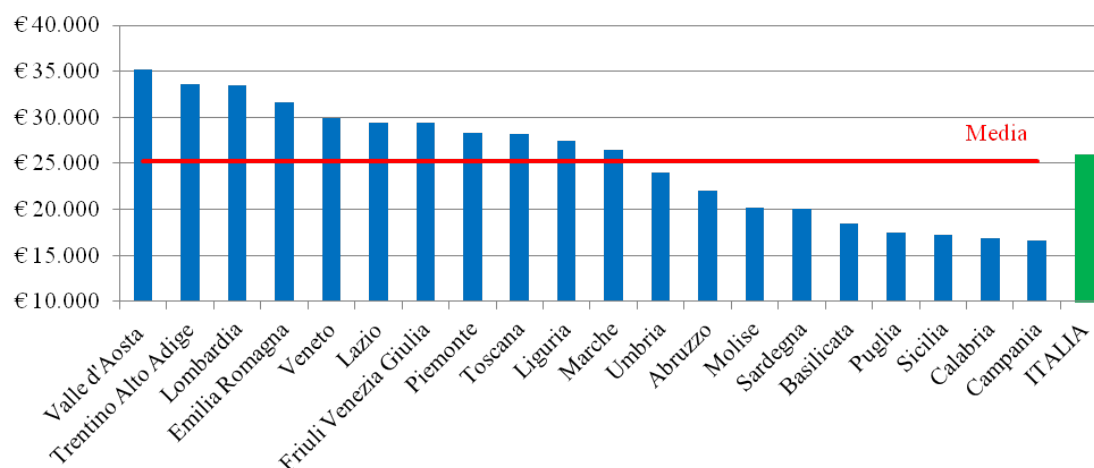
Riportiamo in seguito l’elenco di tutti gli indicatori, ovvero le variabili considerate come la base dell’analisi dei gruppi che condividano uno stesso profilo competitivo. Per ogni variabile andremo a riportare la distribuzione unitaria dei valori registrati per tutte le regioni italiane. Questi non saranno esposti in tabelle ma attraverso grafici a barre, di più intuitiva interpretazione, e saranno ordinati in modo decrescente, formando così una classifica delle regioni. I valori registrati per l’Italia non sempre coincidono con la somma o la media di quelli registrati per le regioni poiché nel calcolo dell’indicatore nazionale confluiscono altri valori (spesso di carattere amministrativo o politico) di portata nazionale e non attribuibili ad alcuna regione specifica.

2.2.1 Pil pro capite

Il Prodotto Interno Lordo delle regioni italiane è ovviamente la prima variabile considerata. Il modo più intuitivo per depurare questo valore dalla grandezza economica della regione è dividerlo per la popolazione media annua, ottenendo così il Pil pro capite annuo ai prezzi correnti.

Nella seguente classifica al primo posto spicca la Valle d’Aosta con circa €35.264, un valore più che doppio di quello all’ultimo posto della Campania di €16.601.

Fig. 2.2.1 – Pil pro capite annuo nelle regioni italiane (2011)



Fonte: elaborazione su dati Istat, Conti economici e regionali

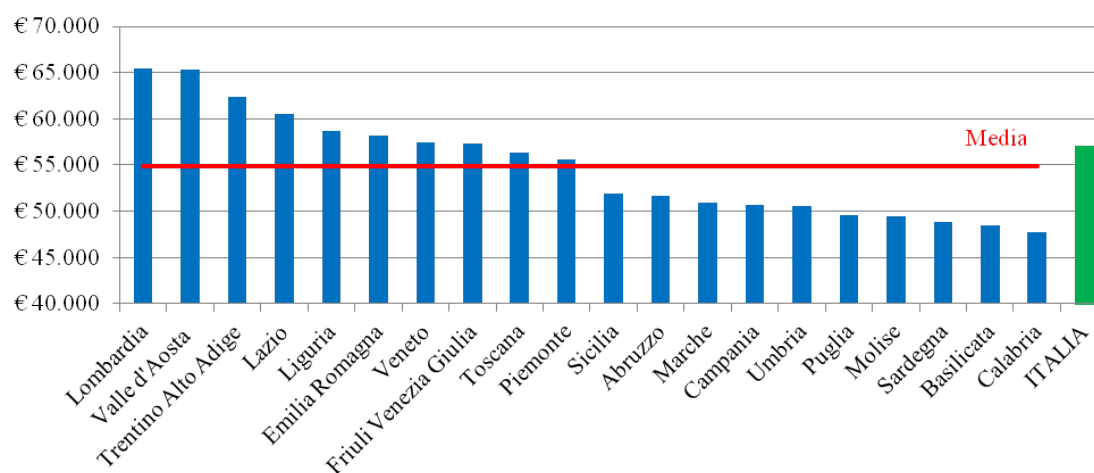
Il valore medio è di circa €25.297 (abbastanza vicino a quello dell'Italia di €26.002) che si può intendere come il reddito medio annuo di un cittadino italiano in un intervallo medio di più o meno €6.234 (la deviazione standard). Un'altra considerazione sulla distribuzione del Pil è il suo carattere fortemente territoriale: il grafico mostra la netta distinzione tra regioni del Nord, Centro e Sud d'Italia con un livello di ricchezza decrescente, una configurazione che persiste in molte delle variabili che andremo ad analizzare in seguito.

2.2.2 Produttività del lavoro

La Produttività del lavoro si calcola come il rapporto tra valore aggiunto ai prezzi base (con il 2005 come anno di riferimento) e la media annua di occupati. Il valore aggiunto per occupato si definisce produttività del lavoro "apparente" in quanto non considera il valore aggiunto apportato dalla forza lavoro complessiva ma quello prodotto dall'insieme delle imprese.

Questo indicatore rappresenta perciò un aspetto essenziale della competitività di un sistema economico, considerando quanto ogni singolo occupato apporta in termini di crescita reale (e senza considerare l'effetto dell'inflazione) al sistema produttivo complessivo, attraverso l'impresa per cui lavora.

Fig. 2.2.2 – Produttività del lavoro nelle regioni italiane (2011)



Fonte: elaborazione su dati Istat, Conti economici e regionali

Nella figura vediamo in vetta la Lombardia (€65.404) e all'ultimo posto la Calabria (€47.674), il valore medio è di circa €54.842 (più basso del valore nazionale di €57.129) con una deviazione standard di €5.644. Anche in questo caso persiste una configurazione fratturata tra Nord e Sud, le regioni del Centro non hanno una postazione centrale in questa classifica.

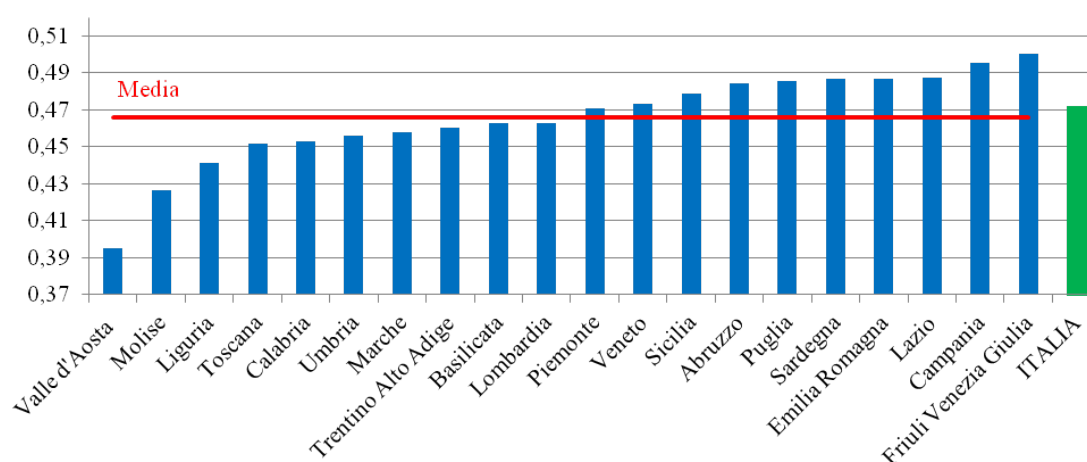
2.2.3 Costo del lavoro

Il Costo del lavoro per unità di prodotto (Clup) si ottiene dividendo i redditi da lavoro dipendente per il valore aggiunto ai prezzi base (con il 2005 come anno di riferimento). Di conseguenza è un valore che si può interpretare come il costo della forza lavoro per aver prodotto una singola unità di valore aggiunto, sempre tramite le imprese in cui i dipendenti lavorano.

Questo è l'unico degli aspetti che consideriamo in questa analisi della competitività ad operare in senso inverso: un sistema produttivo competitivo ha bassi costi del lavoro.

Come si può infatti notare nella figura seguente, la classifica delle regioni è in ordine crescente:

Fig. 2.2.3 – Costo del lavoro nelle regioni italiane (2011)



Fonte: elaborazione su dati Istat, Conti economici e regionali

Confrontando il costo del lavoro con la produttività del lavoro, mentre per quest'ultima era il Centro d'Italia che assumeva valori sia in alto che in basso della classifica, nel caso del costo del lavoro non vi è alcuna frattura territoriale e sia regioni del Nord che regioni del Sud occupano posizioni in testa come in coda.

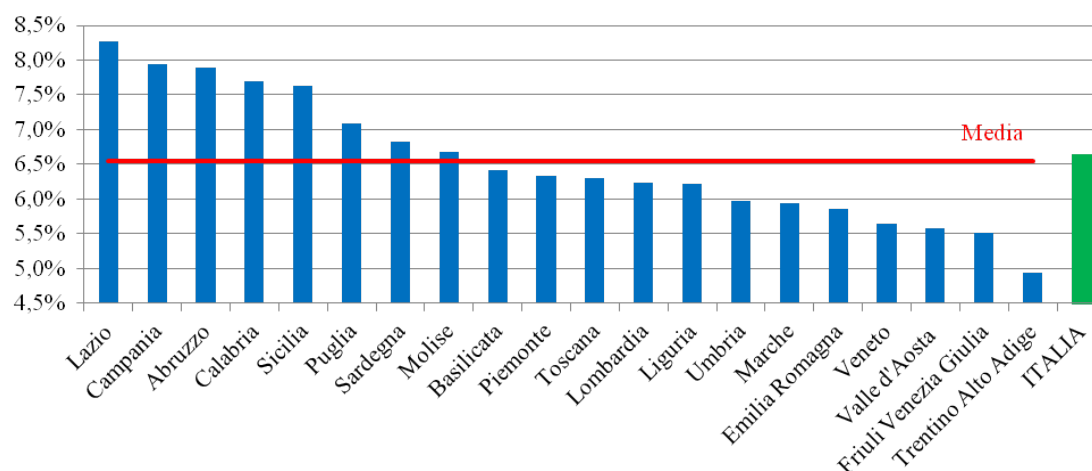
Infatti la prima in classifica è una regione del Nord (la Valle d'Aosta con 0,395) come l'ultima in classifica (il Friuli Venezia Giulia con 0,487).

Il costo del lavoro medio è di 0,46 (quasi coincidente con il valore dell'Italia di 0,47) con una deviazione standard di 0,025.

2.2.4 Tasso di natalità delle imprese

Il tasso di natalità delle imprese viene calcolato come percentuale delle imprese nate in un dato anno sul totale delle imprese attive nello stesso anno. Questo indicatore nell'analisi dell'Istat era correlato positivamente con la crescita della produttività del lavoro, ciò dimostra l'importanza del contributo che le nuove imprese possono apportare ai metodi organizzativi e innovativi. Tuttavia bisogna considerare anche l'aspetto negativo che può essere rivelato da questo indicatore: se ad alti tassi di natalità non è correlata una crescita del valore aggiunto, allora vuol dire che le nuove imprese non stanno aggiungendo valore ma lo stanno sostituendo a quelle precedenti. Ciò può avvenire perché l'aggiunta di una nuova impresa riduce la produttività di quelle già esistenti oppure perché le nuove imprese subentrano al fallimento delle precedenti. Una situazione del genere, ovvero un ricambio improduttivo, risulta certamente poco competitiva in quanto nel tempo di sostituzione di un'impresa con un'altra si ha un'interruzione della produzione che grava sul sistema produttivo.

Fig. 2.2.4 – Tasso di natalità delle imprese nelle regioni italiane (2010)



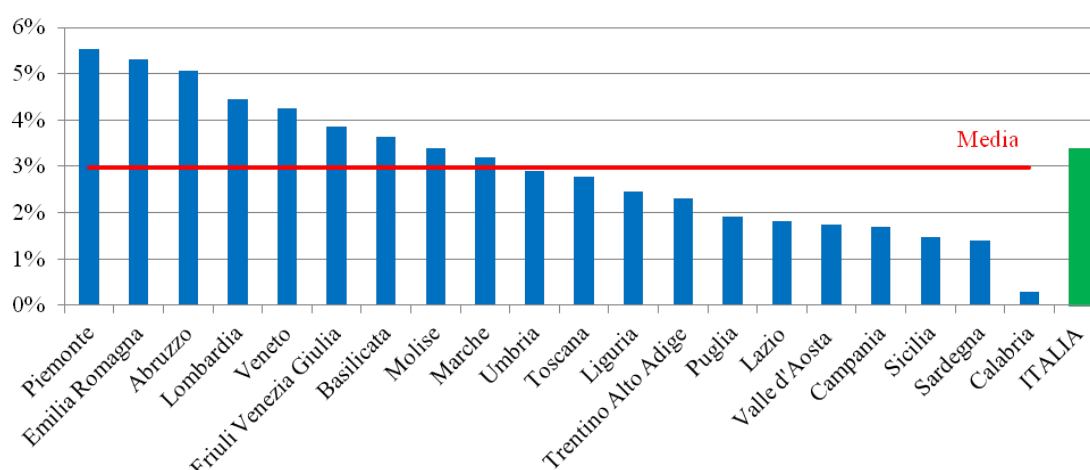
Fonte: elaborazione su dati Istat, Indicatori per le politiche di sviluppo: Demografia d'impresa

Al vertice della classifica si trova il Lazio con un tasso dell'8,26%, alla fine il Trentino Alto Adige con il 4,93%. Il valore medio delle regioni è del 6,54%, quasi coincidente con il 6,65% dell'Italia. Nella distribuzione territoriale del tasso di natalità delle imprese italiane sembra che a Sud vi siano tassi molto più elevati che a Nord. Questo indicatore deve quindi essere correlato ad altre variabili per verificare se ciò che rivela è un aumento della competitività o una riduzione dovuta ad un ricambio delle imprese improduttivo.

2.2.5 Valore aggiunto delle grandi imprese

Il valore aggiunto delle sole grandi imprese è un aspetto che assume una maggiore rilevanza nell'analisi della competitività delle regioni italiane alla luce dei risultati del Rapporto annuale dell'Istat, ovvero che le imprese italiane di dimensioni maggiori sono quelle che riescono più delle altre a penetrare i mercati mondiali. Le imprese considerate nel grafico seguente sono quelle del settore manifatturiero con 250 addetti e oltre, il valore aggiunto regionale di queste imprese è poi depurato dalla grandezza economica della regione dividendolo per il Pil di questa.

Fig. 2.2.5 – Valore aggiunto delle grandi imprese nelle regioni italiane come percentuale del Pil (2009)



Fonte: elaborazione su dati Unioncamere, Appendice statistica al Rapporto 2012

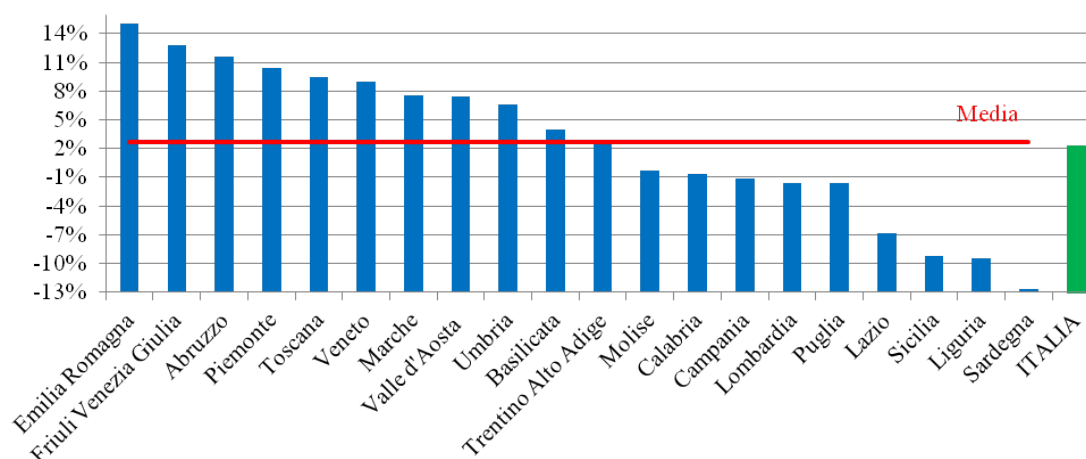
In questa classifica la divisione territoriale italiana con il Nord ai primi posti e il Sud agli ultimi si manifesta debolmente. Con un valore medio del 3% (poco più basso di quello dell'Italia del 3,4%), al primo posto il Piemonte registra una percentuale del 5,5%, all'ultimo posto la Calabria registra lo 0,3% che si distacca nettamente da tutti gli altri valori.

2.2.6 Scambi di merci con il mondo

Il grado di apertura internazionale delle regioni italiane può essere misurato innanzitutto considerando il saldo commerciale con il mondo, ovvero l'ammontare di esportazioni nette di merci e servizi come percentuale del Pil. Questo si può manifestare sotto forma di maggiori esportazioni rispetto alle importazioni (percentuale positiva) oppure il caso contrario (percentuale negativa).

Nel seguente grafico possiamo notare come il valore medio dell'incidenza di esportazioni nette sul Pil sia molto basso, di appena 2,66%, ma con un'alta deviazione standard dell'8%. Possiamo infatti notare un ampio intervallo di variazione: l'Emilia Romagna registra il 15% di saldo commerciale sul Pil (primo posto), mentre la Sardegna il -12,6% (ultimo posto).

Fig. 2.2.6 – Scambi di merci con il mondo nelle regioni italiane come percentuale del Pil (2011)



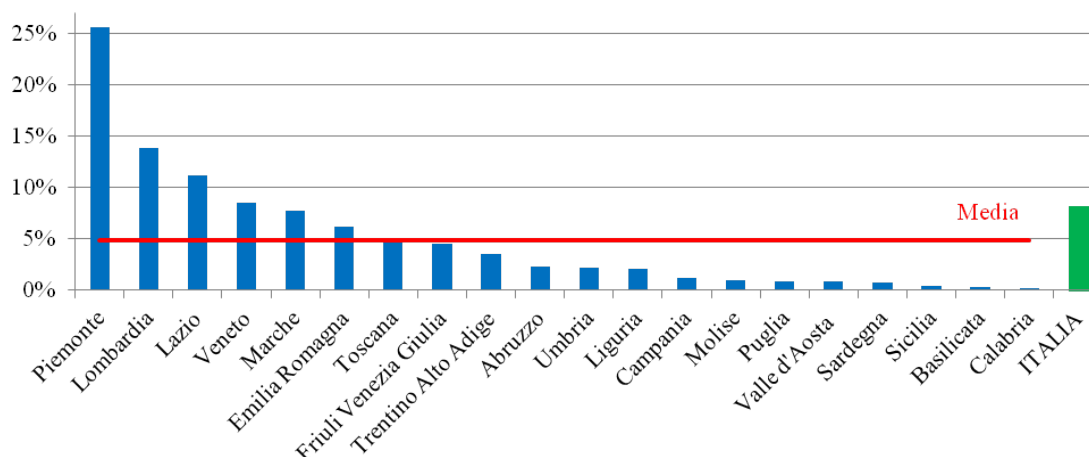
Fonte: elaborazione su dati ICE

Nel 2011 più della metà delle regioni italiane esportano all'estero più di quanto importano (11 su 20), mentre le altre 9 registrano maggiori importazioni. Anche in questo contesto persiste una divisione tra le regioni del Nord, che tendenzialmente registrano esportazioni sopra la media, e regioni del Sud, che registrano invece maggiori importazioni.

2.2.7 Addetti di imprese estere partecipate da imprese italiane

Un'altra modalità di misurazione del grado di internazionalizzazione considera le relazioni con l'estero di lungo periodo, ovvero gli Investimenti diretti esteri in uscita che si possono rilevare attraverso l'incidenza delle attività delle multinazionali italiane all'estero sul totale delle attività italiane. Questa è stata calcolata in termini di addetti, ovvero la percentuale di addetti esteri in media all'anno controllati da imprese italiane sul totale degli occupati dipendenti medi annui. Ogni regione manifesta una sua percentuale in un ampio intervallo di variazione: il valore medio è 4,9%, molto più basso della percentuale nazionale dell'8,2%, con un'alta deviazione standard del 6,2%.

Fig. 2.2.7 – Addetti di imprese estere partecipate da imprese italiane come percentuale del totale degli occupati dipendenti (2011)



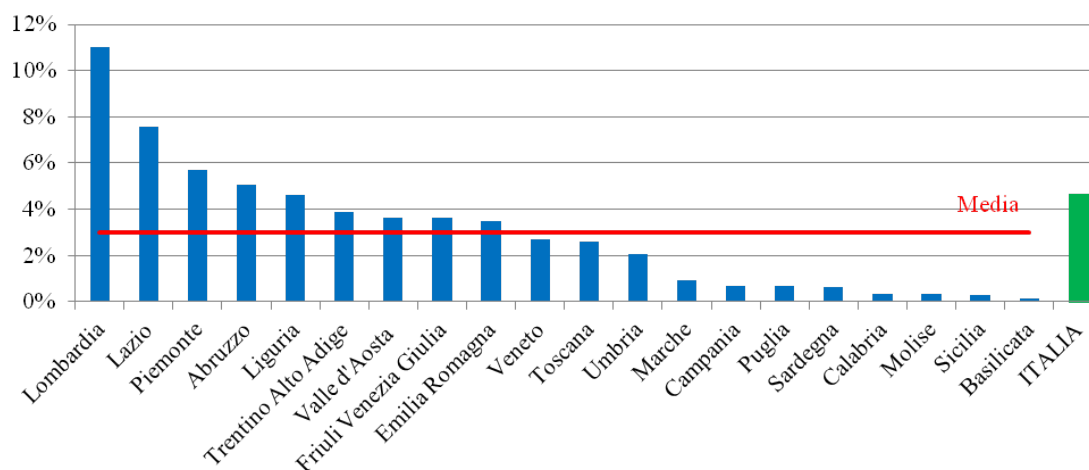
Fonte: elaborazione su dati ICE

In cima alla classifica spicca il Piemonte con un'incidenza di addetti esteri di ben 25,5% che si distacca da tutti gli altri valori, in coda la Calabria registra lo 0,2%. La frammentazione territoriale tra Nord e Sud rileva anche in questo aspetto competitivo, seppur in modo non decisivo: delle sei regioni con internazionalizzazione sopra la media, quattro sono dell'Italia Settentrionale.

2.2.8 Addetti di imprese italiane a partecipazione estera

Passiamo dagli Investimenti diretti esteri in uscita a quelli in entrata, ovvero consideriamo insieme agli addetti esteri controllati anche quelli italiani controllati dall'estero calcolati come la percentuale degli addetti di imprese italiane a partecipazione estera (media annua) sul totale degli occupati dipendenti, in media all'anno. Il profilo competitivo delle regioni è piuttosto simile al precedente, tuttavia l'intervallo di variazione è molto minore: l'incidenza media delle attività delle multinazionali estere in Italia, in termini di addetti, è del 3% (con un valore poco più alto per l'Italia del 4,7%) e la deviazione standard è del 2,8%.

Fig. 2.2.8 – Addetti di imprese italiane a partecipazione estera come percentuale del totale degli occupati dipendenti (2011)



Fonte: elaborazione su dati ICE

In questo caso sono ben nove le regioni che registrano valori sopra la media, sette delle quali dell'Italia Settentrionale. Anche in questo caso si ha una prevalenza del Nord in cima alla classifica, la Lombardia con 11%, e del Sud in fondo a questa, la Basilicata con lo 0,16%.

2.2.9 Spesa in Ricerca e Sviluppo

Analizziamo il livello di conoscenza ed innovazione delle regioni italiane.

Le rilevazioni sulla Ricerca e Sviluppo (R&S) sperimentale in Italia sono condotte dall'Istat seguendo la metodologia del Manuale dell'Ocse-Eurostat sulla rilevazione statistica di questo genere di attività (Manuale di Frascati). Questo documento è stato pubblicato nel 1964 e aggiornato nel 2002, la sua funzione è quella di assicurare la comparabilità dei risultati a livello internazionale.

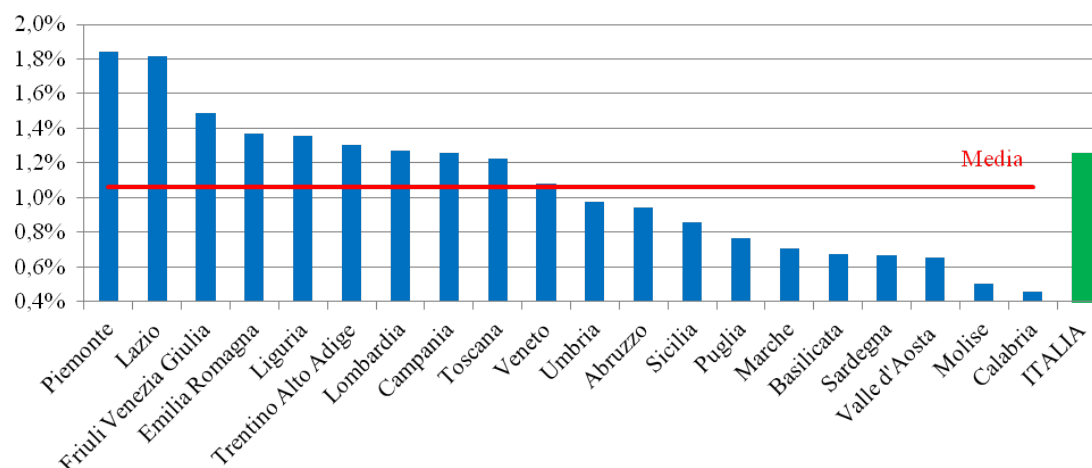
Innanzitutto l'attività di R&S viene definita come quel "complesso di lavori creativi intrapresi in modo sistematico sia per accrescere l'insieme delle conoscenze (compresa la conoscenza dell'uomo, della cultura e della società), sia per utilizzare dette conoscenze per nuove applicazioni".

In questa sede considereremo la spesa per ricerca *intra-muros*, ovvero svolta con proprio personale e con proprie attrezzature senza ricorrere a mezzi di terzi. Gli enti che sono stati sottoposti alla rilevazione delle spese in R&S sono le Amministrazioni pubbliche, le Istituzioni private non profit, le imprese e le Università.

Secondo i dati del 2009, in Italia il totale della spesa in R&S si divide tra questi settori rispettivamente per il 13,1%, 3,3%, 53,3% e 30,3%, dove le imprese hanno la quota maggiore.

Analizziamo ora la percentuale di spesa in R&S come percentuale del Pil delle regioni italiane.

Fig. 2.2.9 – Spesa in ricerca e sviluppo intra-muros nelle regioni italiane in percentuale del Pil (2009)



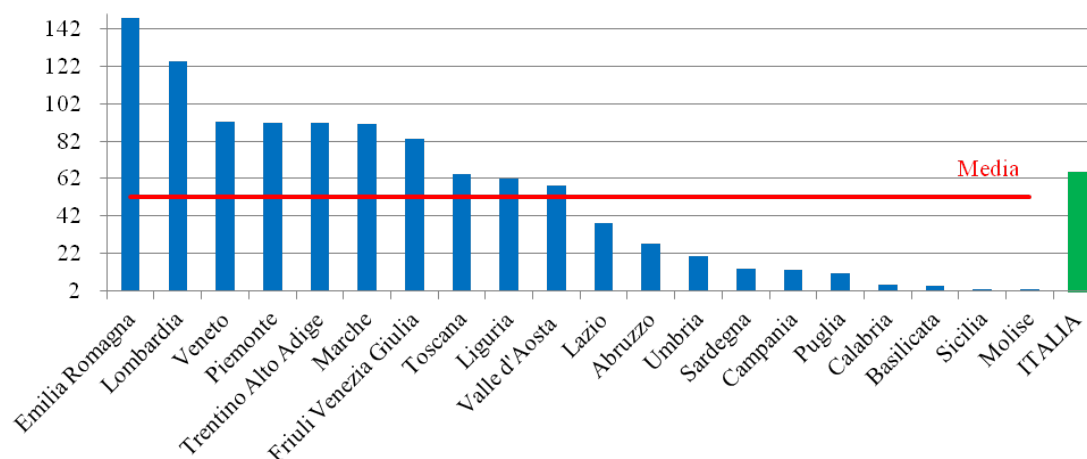
Fonte: elaborazione su dati Unioncamere, Appendice statistica al Rapporto 2012

Dalla classifica emerge al primo posto il Piemonte (1,84%) con un valore molto vicino a quello del Lazio (1,82%), mentre all'ultimo posto la Calabria registra lo 0,46%. Il valore medio è di circa l'1% (più basso del valore nazionale di 1,2%) con una deviazione standard di 0,4%. In questo caso la distinzione Settentrione e Meridione è più marcata: delle dieci regioni che superano il valore medio, sette appartengono all'Italia Settentrionale, due alla Centrale e solo una alla Meridionale.

2.2.10 Numero di Brevetti

Il secondo aspetto che consideriamo per rilevare il grado di conoscenza di ogni regione è il numero annuo di domande di brevetti europei, per milione di abitanti, pubblicate dall'Organizzazione europea dei brevetti (in inglese, European Patent Organisation, EPO). Questa è un'organizzazione pubblica internazionale che nasce nel 1977 dalla Convenzione europea dei brevetti, la sua sede attuale è a Monaco di Baviera (Germania).

Fig. 2.2.10 – Numero di brevetti nelle regioni italiane per milione di abitanti (2010)



Fonte: elaborazione su dati Unioncamere, Appendice statistica al Rapporto 2012

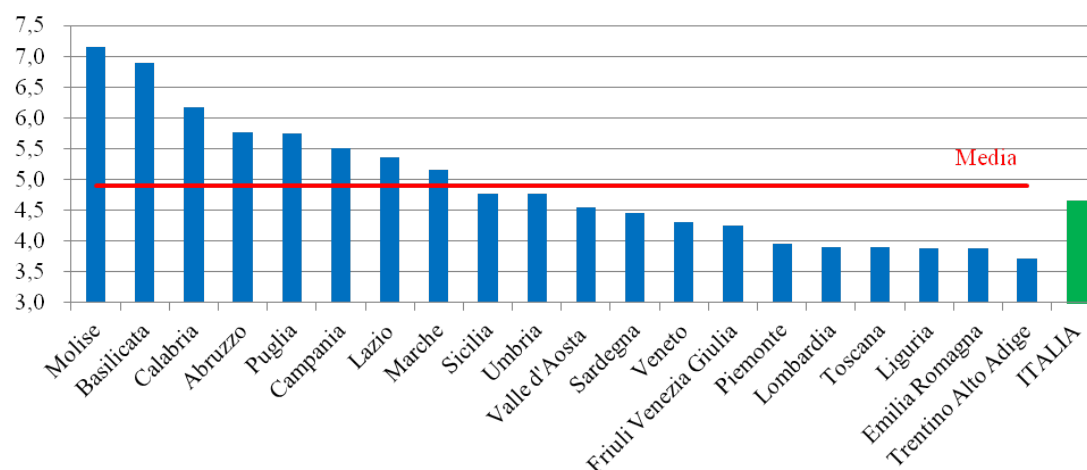
Analizzando la classifica delle regioni, al primo posto spicca l'Emilia Romagna con 147,5 domande di brevetti, mentre all'ultimo il Molise registra solo 3,1 domande. Il valore medio è di 52,5 domande (poco più basso del valore per l'Italia di 65,3), la deviazione standard è di 44,4 domande.

Anche in questo indicatore della conoscenza il Nord d'Italia risulta molto più competitivo, infatti tutte le sue regioni registrano valori oltre la media.

2.2.11 Numero di Laureati

L'ultimo indicatore della conoscenza delle regioni è il numero di laureati per migliaia di abitanti. Questo indicatore dell'istruzione avanzata assume notevole importanza alla luce dei risultati dell'Istat, secondo i quali l'Italia registra un basso numero di laureati a confronto con gli altri paesi europei.

Fig. 2.2.11 – Numero di laureati nelle regioni italiane per migliaia di abitanti (2010)



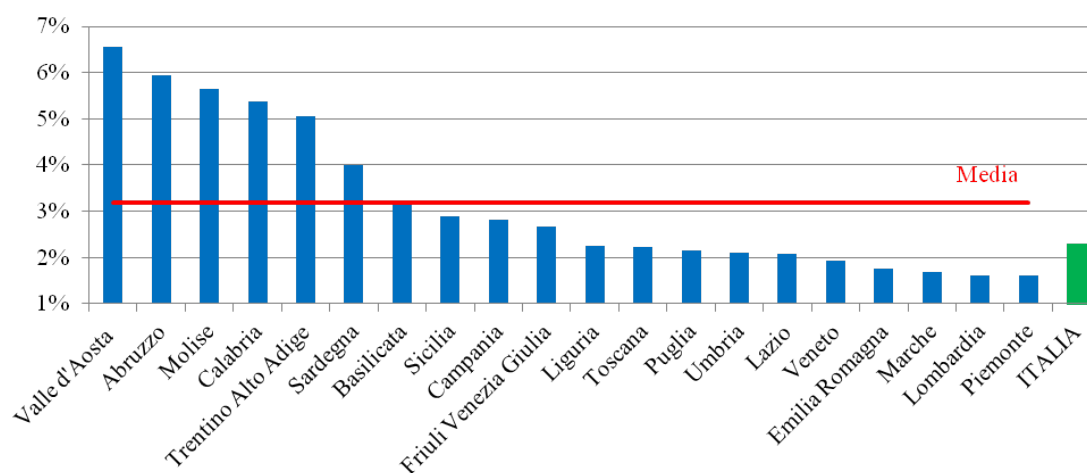
Fonte: elaborazione su dati Unioncamere, Appendice statistica al Rapporto 2012

Nella classifica delle regioni italiane i risultati si distinguono dai precedenti: solo regioni del Sud e del Centro Italia registrano valori oltre la media di 4,9 laureati per migliaia di abitanti, mentre tutte le regioni del Nord ne sono al di sotto. In testa alla classifica, il Molise registra 7,15 laureati, rivalendosi dei bassi valori in R&S e domande di brevetti, mentre all'ultimo posto il Trentino Alto Adige registra un valore di 3,7, pari quasi alla metà di quello del Molise.

2.2.12 Investimenti pubblici

Consideriamo infine l'ultima variabile che compone il profilo competitivo delle regioni italiane, ovvero l'ammontare di spesa pubblica come percentuale del Pil. In questo stock rientra: amministrazione pubblica e difesa, assicurazione sociale obbligatoria, istruzione, sanità e assistenza sociale; attività artistiche, di intrattenimento e divertimento; riparazione di beni per la casa e altri servizi.

Fig. 2.2.12 – Spesa pubblica nelle regioni italiane in percentuale del Pil (2010)



Fonte: elaborazione su dati Istat, Conti economici e regionali

Il grafico mostra un'incidenza media degli investimenti pubblici sul Pil molto bassa (3,2% con una deviazione standard dell'1,6%), nella distribuzione spicca un gruppo di regioni molto al di sopra della media e un altro gruppo, più numeroso, di regioni molto al di sotto. La situazione territoriale è molto variegata, infatti non è possibile conformare il livello di spesa pubblica tra le regioni del Nord, Centro o Sud.

Al primo posto la Valle d'Aosta registra un'alta percentuale del 6,5%, all'ultimo posto vi è ancora una regione del Nord, il Piemonte, con solo l'1,6%.

CAPITOLO 3

LA METODOLOGIA STATISTICA DELLA CLUSTER ANALYSIS

Nel precedente capitolo sono stati esposti tutti gli aspetti economici che delineano il profilo competitivo delle regioni italiane. In questo capitolo sarà spiegata la metodologia statistica della *cluster analysis* con cui si considereranno contemporaneamente tutte le variabili di tutte le regioni per poterle raggruppare in gruppi distinti, all'interno dei quali le regioni condividano la medesima combinazione di variabili, ovvero il medesimo livello competitivo, e si distinguano nettamente dagli altri gruppi.

La metodologia della *cluster analysis*, ovvero l'analisi dei gruppi, vuole classificare in diversi gruppi (detti *cluster*) le varie unità statistiche osservate in base al criterio di similarità delle unità, molto intuitivamente: unità simili tra loro saranno poste nello stesso *cluster*.

L'*output* di questa analisi è dunque un raggruppamento, una partizione la cui efficacia ed utilità può essere valutata in termini di sintesi e in termini di affidabilità. Il primo concetto riguarda il numero di gruppi da stabilire, minore è questo numero e maggiore è la sintesi. Il concetto di affidabilità riguarda invece il rispetto della *coesione interna*, ovvero quanto le unità all'interno di un gruppo sono simili tra loro, e della *separazione esterna*, ovvero quanto i gruppi sono diversi tra loro. È facile notare il *trade-off* che esiste tra sintesi ed affidabilità: raggruppare tutte le unità in pochi gruppi fornisce una partizione estremamente sintetica ma meno affidabile di una partizione con più gruppi.

La miglior partizione è quella che massimizza l'eterogeneità tra gruppi nonché l'omogeneità tra unità di uno stesso gruppo e che minimizza il numero di gruppi, ovvero garantisce un buon livello di sintesi che si può giudicare solo in luce dello scopo, dell'obiettivo della ricerca statistica.

Nella nostra analisi della competitività regionale applicheremo la metodologia di questo capitolo per trovare il miglior raggruppamento possibile per vari livelli di sintesi, cioè per vari numeri di gruppi. Infine si sceglierà il numero di gruppi che permette maggiormente di definire il profilo competitivo di ognuno di essi, ma questo argomento sarà trattato più avanti.

Analizziamo ora le varie fasi di una *cluster analysis* che possiamo riassumere in quattro passi:

- 1) *Unità da raggruppare*: costruzione della matrice dei dati;
- 2) *Similarità delle unità*: misura di "distanza" tra unità statistiche;
- 3) *Metodi di formazione dei gruppi*: i diversi sistemi di utilizzo della distanza tra le unità in modo da raggrupparle e formare una partizione;
- 4) *Valutazione della partizione*: attraverso un indice di bontà R^2 .

Una volta trovato il miglior raggruppamento delle unità sarà possibile calcolare le statistiche descrittive, quali la media e la varianza delle unità, per ogni gruppo. In questo modo si potrà capire cosa caratterizza un insieme di unità rispetto agli altri considerato la sua peculiare combinazione di variabili.

3.1 Le unità da raggruppare

Nell'analisi che condurremo le unità da raggruppare sono le venti regioni italiane, ognuna delle quali registra valori nelle variabili esposte nel precedente capitolo. Tali variabili saranno le caratteristiche competitive di ogni regione che determineranno il grado di similarità o diversità tra queste.

La rilevanza delle variabili è stata ricercata sia a livello qualitativo, scegliendo quelle con maggior potere discriminatorio scartando quelle che confonderebbero la classificazione, sia a livello quantitativo, ovvero un numero di variabili tale che sia l'aggiunta che l'eliminazione di una variabile lasci pressoché inalterata la partizione individuata.

L'insieme dei dati raccolti deve essere a questo punto disposto in una forma matematica sulla quale sia possibile applicare la *cluster analysis*. Questa è, intuitivamente, una matrice dei dati che elenca le n unità statistiche e per ognuna riporta i valori registrati nelle p variabili. La matrice così costruita, che chiamiamo X , avrà dimensioni $n \times p$ dovute a n righe (ovvero n vettori-unità) e p colonne (ovvero p vettori-variabile):

$$X = \begin{pmatrix} x_{11} & \dots & x_{1s} & \dots & x_{1p} \\ \vdots & & \vdots & & \vdots \\ x_{i1} & \dots & x_{is} & \dots & x_{ip} \\ \vdots & & \vdots & & \vdots \\ x_{n1} & \dots & x_{ns} & \dots & x_{np} \end{pmatrix}$$

vettore-unità

vettore-variabile

Riportiamo ora un esempio concreto di matrice dei dati riferita a 5 studenti universitari immaginari, dello stesso anno e facoltà, ai quali sono stati chieste 4 informazioni di carattere quantitativo:

Studente	Età	Reddito familiare mensile	Media (in trentesimi)	Esami svolti
A	21	€ 2.300	21,5	20
B	22	€ 2.000	23	18
C	22	€ 2.100	29,8	18
D	23	€ 1.800	25	17
E	24	€ 1.500	26,7	19

I valori rilevati in ogni ambito informativo costituiscono l'essenza statistica di ogni unità.

Nella nostra analisi considereremo ogni variabile equivalente, l'importanza di ogni aspetto rilevato per ogni regione sarà quindi la stessa ai fini del raggruppamento delle unità.

Per considerare ogni variabile come equivalente, la matrice dei dati deve essere standardizzata, ovvero si andrà a calcolare la media ($\bar{x}_s$) e la deviazione standard (σ_s) di ogni vettore-variabile per poi standardizzare ogni valore x_{is} della matrice (ovvero della riga i -esima e della colonna s -esima) sottraendo la media e dividendo per la deviazione standard entrambe della colonna s -esima:

$$z_{is} = \frac{x_{is} - \bar{x}_s}{\sigma_s} \quad \text{per } i = 1, \dots, n$$

Applicando questa procedura alla precedente matrice esemplificativa si otterrà la seguente matrice degli scarti standardizzati:

Studente	Età	Reddito familiare mensile	Media (in trentesimi)	Esami svolti
A	-1,23	1,18	-1,14	1,40
B	-0,35	0,20	-0,68	-0,35
C	-0,35	0,52	1,42	-0,35
D	0,53	-0,46	-0,06	-1,23
E	1,40	-1,44	0,46	0,53

In questo modo le variabili sono perfettamente comparabili poiché non ci sono più diverse unità di misura e hanno tutte lo stesso ordine di grandezza: la media di ogni vettore-variabile è 0 e la varianza è 1.

D'ora in avanti, quando ci riferiremo ad una matrice dei dati generica questa sarà, nell'analisi che condurremo, la matrice degli scarti standardizzati.

3.2 Similarità delle unità

Il concetto di similarità o rassomiglianza tra unità si traduce a livello algebrico molto semplicemente con il concetto di differenza: dato un insieme di numeri, i due che hanno la minore differenza sono i più simili tra loro. Dunque se ogni unità avesse una sola variabile sarebbe molto facile raggruppare l'insieme delle unità, prendendo ad esempio in considerazione la sola variabile "Reddito familiare", le classi potrebbero essere semplicemente "Alto", "Medio" e "Basso". In ambito statistico, tuttavia, il concetto di differenza non è sufficiente in quanto si considerano unità che presentano valori per diverse variabili, ovvero non si parla di numeri ma di vettori.

Introduciamo dunque il concetto di *distanza*, definita come una funzione tra due vettori che indica il loro grado di similarità. Questa funzione gode almeno delle proprietà di non negatività, di identità (se la distanza è uguale a zero i due vettori sono uguali) e di simmetria (la distanza da un vettore ad un secondo vettore è uguale alla distanza dal secondo vettore al primo).

Esistono varie tipologie di distanze (euclidea, di Lagrange, della città a blocchi, di Canberra, di Jaccard), quella che useremo per la nostra analisi è la distanza euclidea, la più intuitiva nonché la più usata per la *cluster analysis*. Date due unità statistiche (vettori i e j) a p variabili ciascuna, la distanza euclidea (indicata con ${}_2d_{ij}$) è la norma della differenza tra i vettori:

$${}_2d_{ij} = ||x_i - x_j|| = \sqrt{\sum_{s=1}^p (x_{is} - x_{js})^2}$$

Dunque per ogni variabile si calcola la differenza tra i due valori delle due unità e si eleva alla seconda così da rispettare la proprietà di non negatività (poiché una distanza negativa non ha alcun significato). In seguito si sommano tutti i quadrati delle differenze per avere un unico valore la cui radice quadrata rappresenta la distanza tra le due unità, considerando contemporaneamente tutte le variabili.

L'insieme delle distanze calcolate per ogni coppia di unità, secondo una data tipologia di distanza (euclidea nel nostro caso), è rappresentato dalla *matrice delle distanze*. Partendo da una matrice dei dati di n unità e p variabili, la matrice delle distanze D avrà dimensioni

$n \times n$ poiché va a relazionare le n unità con le stesse n unità. La distanza di un'unità da se stessa sarà zero (proprietà di identità), quindi la matrice D avrà valori nulli sulla diagonale principale. Inoltre la distanza tra un'unità ed una seconda unità è uguale alla distanza della seconda unità dalla prima (proprietà di simmetria), quindi la matrice D sarà simmetrica. Generalmente la matrice D si rappresenta nel seguente modo:

$$D = \begin{pmatrix} 0 & \dots & d_{12} & \dots & d_{1n} \\ & & \vdots & & \vdots \\ & & 0 & \dots & d_{in} \\ & & & & \vdots \\ & & & & 0 \end{pmatrix}$$

Tuttavia i valori possono occupare il triangolo superiore come quello inferiore.

A titolo esemplificativo, riportiamo in seguito la matrice delle distanze euclidee calcolata a partire dalla matrice degli scarti standardizzati dell'esempio precedente:

	A	B	C	D	E	
A	0	2,24	3,29	3,72	4,13	→ Max distanza
B		0	2,12	1,54	2,80	→ Min distanza
C			0	2,17	2,93	
D				0	2,25	
E					0	

La casella A-B è stata calcolata considerando i seguenti valori:

Studente	Età	Reddito familiare	Media	Esami svolti
A	-1,23	1,18	-1,14	1,40
B	-0,35	0,20	-0,68	-0,35

Il calcolo della distanza euclidea è il seguente:

$${}_2d_{AB} = \sqrt{(-1,23 + 0,35)^2 + (1,18 - 0,20)^2 + (-1,14 + 0,68)^2 + (1,40 - 0,35)^2}$$

L'interpretazione della matrice delle distanze può avvenire su scala ordinale, ovvero i valori numerici assunti dalle distanze possono essere utilizzati per costruire una graduatoria di similarità, cioè dalla minore distanza alla maggiore distanza. Al vertice della classifica si avrebbe la coppia di unità più simili, nel nostro esempio lo studente B e lo studente D (la distanza di 1,54 è la più piccola), mentre in fondo alla classifica si avrebbero gli studenti più diversi, lo studente A e lo studente E (per la distanza di 4,13).

A questo punto può essere spiegata più chiaramente l'importanza della comparabilità delle variabili e dunque quella di usare una matrice di scarti standardizzati: se le distanze fossero state calcolate sulla matrice dei dati originale, la variabile con il maggior ordine di grandezza (nel nostro esempio il Reddito Familiare) e con la maggiore variabilità avrebbe prevalso sulle altre e i valori delle distanze sarebbero stati determinati quasi totalmente da questa variabile.

3.3 Metodi di formazione dei gruppi

I metodi di partizione delle unità statistiche si distinguono in gerarchici e non gerarchici.

I *metodi gerarchici* si basano su un rigido criterio di aggregazione graduale in cui ogni raggruppamento eseguito ad un dato livello gerarchico non può essere sciolto nel livello successivo. Il criterio di raggruppamento è dato dal livello di distanza: le unità più simili tra loro si aggregano e il procedimento si ripete per tutti i livelli di distanza (senza poter separare le unità ormai aggregate ai livelli inferiori) fin quando tutte le unità sono riunite in un unico gruppo (metodi aggregativi) oppure si procede al contrario disaggregando le unità che tra loro hanno alti livelli di distanza fin quando il gruppo formato dalle n unità non viene diviso in n gruppi di una unità (metodi divisivi).

I *metodi non gerarchici* si basano invece su un continuo spostamento delle unità finché non viene massimizzata una funzione obiettivo. Questo procedimento iterativo si può applicare su un insieme di unità stabilendo già dall'inizio il numero di gruppi voluti nella partizione. Come si può notare questi metodi risultano molto più flessibili, ma necessitano di una impostazione iniziale dei parametri ben ponderata.

Generalmente questa si ottiene integrando il metodo non gerarchico con uno o più metodi gerarchici, in seguito spiegheremo una procedura di formazione dei gruppi che considera contemporaneamente i risultati di vari metodi così da sfruttare più vantaggi operativi ed arrivare ad un risultato stabile ed autorevole.

3.3.1 Metodi gerarchici

I metodi gerarchici di cui ci occuperemo hanno come input una matrice delle distanze, calcolata su una matrice dei dati, sulla quale si applica la seguente procedura:

- 1) Si individuano le due unità con la minore distanza e si riuniscono a formare il primo gruppo, il risultato sarà una partizione di $n-1$ gruppi di cui uno soltanto è formato da due unità mentre tutti gli altri sono formati da una singola unità;
- 2) Si ricalcola la distanza del gruppo ottenuto dagli altri elementi (che all'inizio della procedura sono singole unità, più avanti saranno raggruppamenti di unità), ricavando una nuova matrice delle distanze che avrà dimensioni diminuite di uno. Il metodo utilizzato per ricalcolare la distanza tra due gruppi di unità (contenenti molte ma anche una unità) definisce la tipologia di metodo gerarchico utilizzato;
- 3) Si individua nella nuova matrice delle distanze la coppia di unità (o gruppi) con minore distanza, si andranno quindi a riunire in un unico gruppo;
- 4) Si ripetono le fasi 2 e 3 finché tutte le unità sono riunite in un solo gruppo;

Terminata questa procedura si disporrà di una sequenza di raggruppamenti ognuno per un livello di distanza, si dovrà quindi decidere fino a quale livello di distanza è più opportuno raggruppare le unità.

I metodi gerarchici di partizione che prenderemo in analisi sono caratterizzati da diverse definizioni di distanza tra gruppi di unità che permettono di ricostruire la matrice delle distanze per ogni raggruppamento. Nella nostra analisi considereremo i più usati:

- **Metodo del legame singolo:** (*single linkage* o *nearest neighbour*) la distanza tra due gruppi di unità è data dalla minima delle distanze tra ciascuna delle unità d'un gruppo e ciascuna delle unità dell'altro gruppo;

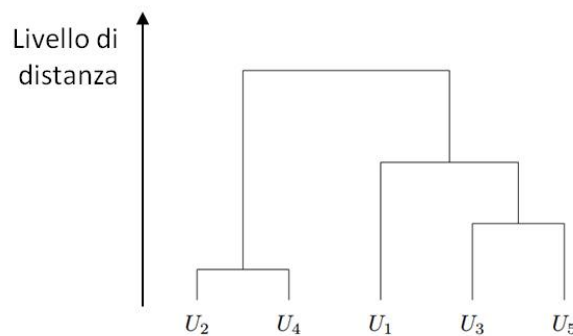
- **Metodo del legame completo:** (*complete linkage* o *furthest neighbour*) la distanza tra due gruppi di unità è data dalla massima delle distanze tra ciascuna delle unità d'un gruppo e ciascuna delle unità dell'altro gruppo;
- **Metodo del legame medio:** (*average linkage* o UPGMA) la distanza tra due gruppi di unità è data dalla media aritmetica di tutte le distanze tra ciascuna delle unità d'un gruppo e ciascuna delle unità dell'altro gruppo;

Ovviamente tutti questi metodi utilizzano la stessa tipologia di distanza utilizzata per calcolare la matrice delle distanze su cui essi si applicano.

Vi è poi un altro metodo che prenderemo in considerazione che non si basa tuttavia sul concetto di distanza ma su una funzione obiettivo, lo spieghiamo brevemente in quanto i concetti alla sua base saranno ripresi nel paragrafo Valutazione della partizione.

- **Metodo di Ward:** (metodo della minima devianza) considera la matrice dei dati per aggregare ad ogni passo della procedura gerarchica i gruppi che comportano il minor incremento della Devianza nei gruppi;

La procedura dei metodi gerarchici aggregativi può essere rappresentata attraverso un grafico molto efficace chiamato *dendrogramma*, da δένδρον che in greco vuol dire albero. Questo grafico mostra quali unità o gruppi di unità vengono uniti insieme per ogni livello di distanza, alla massima distanza essi si troveranno riuniti in un unico gruppo che forma il “tronco” dell'albero mentre i primi raggruppamenti formano le “radici”.



Ogni metodo di raggruppamento su una stessa matrice delle distanze produrrà un dendrogramma diverso, nella figura ne è proposto un esempio su cinque unità statistiche.

3.3.2 Metodi non gerarchici

Il metodo non gerarchico di cui ci occuperemo è il *metodo delle k-medie*, dove k indica il numero di gruppi che si richiedono a questo metodo (k si può ricavare dall'applicazione dei metodi gerarchici) mentre *medie* indica il cuore della procedura: l'allocazione delle unità nei k gruppi avverrà considerando la minor distanza dai valori medi di ogni gruppo. Tali valori si definiscono *centroidi*: un centroide di un insieme di vettori-unità in p variabili è quel vettore composto dalle medie aritmetiche di tutti i vettori-unità per ogni variabile.

L'*input* del metodo delle k -medie, oltre il numero k , è una matrice dei dati, che ipotizziamo di dimensioni $n \times p$, sulla quale applica il seguente algoritmo:

- 1) Dopo aver stabilito un numero k di gruppi in cui classificare le n unità, si stabiliscono i k poli iniziali, ovvero i vettori in p variabili che fungeranno da centri di allocazione dei k gruppi. Questi possono essere indicati dal ricercatore o stabiliti casualmente dal metodo;

- 2) Ogni unità nella matrice dei dati viene allocata nel gruppo il cui polo risulta il più vicino secondo una tipologia di distanza, si costruisce così la partizione iniziale;
- 3) Per ogni unità nella matrice dei dati si calcola la distanza (della tipologia scelta) non più dai poli iniziali ma dai centroidi dei k gruppi. Ogni unità sarà quindi assegnata al gruppo con il centroide dal quale risulta più vicina. Se l'unità non rimane nel suo gruppo di appartenenza allora si ricalcola il centroide del gruppo che ha lasciato e quello del gruppo in cui è entrata;
- 4) Si ripete il passo 3 finché ogni unità è alla minima distanza dal centroide del suo gruppo di appartenenza e dunque non avvengono più spostamenti.

La metrica utilizzata nel calcolo delle distanze con il metodo delle k -medie è generalmente quella euclidea poiché, come si può dimostrare matematicamente, essa garantisce la convergenza della procedura iterativa nonché la massimizzazione dell'obiettivo di coesione interna. Come ogni metodo non-gerarchico anche il metodo delle k -medie ha una funzione obiettivo, questa è la devianza nei gruppi (che sarà spiegata nel paragrafo Valutazione della partizione) che deve essere minimizzata, la metrica euclidea consente questa operazione.

Il primo punto critico di questa procedura è la scelta del numero k , questo deve essere sufficientemente elevato da consentire alle unità anomale (o *outliers*) di formare gruppi poco numerosi (o anche formati da una sola unità) in modo da evitare una distorsione dei dati.

Il secondo punto critico è poi la scelta dei k poli iniziali. Questi vettori possono essere scelti in modo casuale tra le unità della matrice dei dati (magari facendo in modo che tra essi vi sia una distanza minima), ma con questo sistema la procedura delle k -medie produrrà un raggruppamento finale ogni volta diverso. L'alternativa è stabilire i poli iniziali utilizzando i risultati ottenuti con un metodo gerarchico, questa soluzione consente di fondere il sistema stabile e strutturato di un metodo gerarchico con la flessibilità del metodo delle k -medie (ovvero la possibilità di effettuare numerosi aggiustamenti fino a trovare la miglior allocazione delle unità).

Nel paragrafo Procedura di formazione dei gruppi sarà spiegata un'integrazione di tutti i metodi di *cluster analysis* fin qui spiegati.

3.4 Valutazione della partizione

Data una partizione delle n unità statistiche in k cluster, che provenga da un metodo gerarchico come da uno non gerarchico, si possono facilmente valutare i gradi di coesione interna e di separazione esterna che rappresentano i parametri più intuitivi per misurare la bontà di una partizione.

La *coesione interna* di una partizione, ovvero l'omogeneità tra le unità all'interno dei gruppi, si può misurare attraverso la Devianza nei gruppi (detta *Within*). La devianza in un gruppo è la somma dei quadrati delle differenze di ogni unità del gruppo dalla media di tutte le unità del gruppo, per ogni p variabile. *Within* è data dalla somma di tutte le devianze nei gruppi, in formula è il quadrato dello scarto tra vettore unità, per tutte le unità del gruppo, e centroide del gruppo per tutti i gruppi:

$$W = \sum_{l=1}^k \sum_{s=1}^p \sum_{i=1}^{n_l} (x_{is} - \bar{x}_{s,l})^2$$

Minore è il valore di W e più le unità in ogni gruppo sono omogenee tra loro.

La *separazione esterna* di una partizione, ovvero l'eterogeneità tra i gruppi di unità, si può misurare attraverso la Devianza fra gruppi (detta *Between*). Per ogni gruppo si calcola la somma dei quadrati delle differenze tra valori medi delle unità nel gruppo, per ogni variabile, e valori medi di tutte le n unità, tale somma è moltiplicata per il numero di unità all'interno del gruppo. *Between* è data dalla somma di tutte le devianze fra gruppi, in formula è il quadrato dello scarto tra centroide del gruppo, per tutti i gruppi e moltiplicato per il numero di unità in esso, e centroide di tutte le unità:

$$B = \sum_{s=1}^p \sum_{l=1}^k n_l (\bar{x}_{s,l} - \bar{x}_s)^2$$

Maggiore è il valore di B e più le unità in gruppi diversi sono eterogenee tra loro.

Si può dimostrare matematicamente che la somma di Devianza nei gruppi (W) e Devianza fra gruppi (B) equivale alla Devianza totale (*Total*, T) calcolata come la somma dei quadrati delle differenze di ogni unità, dalla media di tutte le unità, per ogni p variabile.

Dunque $T = W + B$ dove:

$$T = \sum_{s=1}^p \sum_{i=1}^n (x_{is} - \bar{x}_s)^2$$

L'indicatore che utilizzeremo per valutare la bontà di una partizione è l'indice R^2 che considera la Devianza nei gruppi, o la Devianza fra gruppi, rispetto alla Devianza Totale.

R^2 si calcola a partire da questi valori:

$$R^2 = B/T \quad \text{oppure} \quad R^2 = 1 - W/T$$

L'indice è una percentuale che misura la bontà di una partizione, ovvero assume valori nell'intervallo $[0;1]$ dove la prossimità all'unità simboleggia una partizione con alta coesione interna (ovvero una ridotta Devianza nei gruppi) e separazione esterna (ovvero un'elevata Devianza fra gruppi corrispondente).

A titolo esemplificativo, riportiamo il calcolo di R^2 per una partizione casuale del precedente esempio:

Gruppo 1:

B	-0,35	0,20	-0,68	-0,35
D	0,53	-0,46	-0,06	-1,23
medie	0,09	-0,13	-0,37	-0,79

Gruppo 2:

A	-1,23	1,18	-1,14	1,40
C	-0,35	0,52	1,42	-0,35
E	1,40	-1,44	0,46	0,53
medie	-0,06	0,09	0,25	0,53

Il calcolo di T è molto semplice in quanto la media di ogni vettore variabile è 0, essendo la matrice standardizzata, dunque basterà elevare al quadrato tutti i valori delle unità e sommarli.

$$T = \sum_{s=1}^p \sum_{i=1}^n x_{is}^2 = 15,98$$

La devianza nei gruppi si calcola separatamente, per il primo gruppo mostriamo il calcolo:

$$\text{Devianza}_1 = \begin{array}{ccccccc} (-0,35-0,09)^2 & + & (0,2-0,13)^2 & + & (-0,68-0,37)^2 & + & (-0,35-0,79)^2 \\ + & & + & & + & & + \\ (0,53-0,09)^2 & & (-0,46-0,13)^2 & & (-0,06-0,37)^2 & & (-1,23-0,79)^2 \end{array} = 1,18$$

Il valore di Devianza_2 è 12,17, dunque la devianza nei gruppi (W) sarà $\text{Devianza}_1 + \text{Devianza}_2$, ovvero 13,35.

A questo punto possiamo calcolare R^2 come $1 - (13,35/15,98) = 0,16$, un valore decisamente basso poiché una partizione appena sufficiente ha almeno 0,6.

Il valore di R^2 andrà inevitabilmente a crescere con l'aumento del numero di gruppi di una partizione, questa correlazione ricorda il *trade-off* che esiste tra sintesi, dovuta a pochi gruppi, e variabilità dei dati, che si riduce all'aumentare dei gruppi.

3.5 Procedura di formazione dei gruppi

È chiaro a questo punto come ogni metodologia di *cluster analysis* produca il suo risultato sebbene applicata sulla medesima matrice dei dati. Nella ricerca di un raggruppamento consolidato e veritiero è dunque saggio non affidarsi ad una sola metodologia ma trovare la giusta integrazione di più strumenti possibili.

La procedura di formazione dei gruppi che proponiamo in questa sede prende in considerazione tutte le metodologie spiegate in questo capitolo. La base di partenza è costituita dai metodi gerarchici, l'insieme dei dendrogrammi prodotti da ogni metodo può dare una prima idea del numero ideale di gruppi. La ricerca non deve vertere tuttavia al numero esatto, ma almeno ad un breve intervallo di questi (un *range* di valori ragionevoli). Per ogni numero fattibile di gruppi, ogni metodo gerarchico proporrà la sua partizione ideale. Questa si può scoprire “tagliando” il grafico ad albero, basta porre una linea in alto e farla scendere gradualmente finché da essa non partiranno un numero di linee pari al numero di gruppi voluto nella partizione. Ogni linea porterà con sé una serie di ramificazioni, tutte le unità ad esse attaccate apparterranno ad un gruppo distinto.

A questo punto si può valutare ogni partizione tramite l'indice R^2 : calcolando questo indice per le partizioni ottenute con ogni metodo e per ogni numero fattibile di gruppi, è possibile verificare qual è il metodo gerarchico più “autorevole” per numerosità dei gruppi (l'indice R^2 crescerà con la numerosità dei gruppi).

Il passo successivo prevede l'entrata del metodo delle k-medie, che sarà guidato dal miglior metodo gerarchico (secondo la valutazione dell'indice R^2) per ogni numero fattibile di gruppi.

Dunque per ogni numero di gruppi ragionevole sarà svolta una procedura k-medie (con k così determinato) che sarà guidata in questo modo: utilizzerà come poli iniziali i centroidi dei k gruppi ottenuti con il miglior metodo gerarchico tagliandone il dendrogramma per formare k gruppi.

In seguito si potrà valutare con l'indice R^2 ogni partizione ottenuta con k-medie, il risultato sarà migliore (o al peggio uguale) di quello ottenuto con il metodo gerarchico che ha guidato k-medie.

La conclusione di tutta questa procedura è dunque un raggruppamento “ideale”, poiché testato da numerose metodologie, per ogni numero ragionevole di gruppi. La scelta da compiere sarà solamente tale numero, considerando non solo una partizione che isoli gli *outliers* ma anche gli obiettivi della ricerca, ovvero dove posizionarsi nel *trade-off* tra sintesi (o utilità) e variabilità (o affidabilità) dei risultati.

3.6 Il software R

Il software scelto per l'implementazione delle metodologie statistiche sulla base di dati raccolta è il pacchetto gratuito "R". Dati i vantaggi che esso può offrire rispetto ad altri software commerciali di calcolo statistico (quali SAS, SPSS, STATA, S PLUS) ne riportiamo in seguito le principali caratteristiche.

Fig. X – Logo del software R



Fonte: sito di R

R viene definito come un linguaggio e un ambiente per l'analisi statistica e la rappresentazione dei grafici. È un software libero distribuito con la *General Public License* che assicura all'utente la libertà di utilizzo, copia, modifica (in quanto sono resi disponibili anche i codici sorgente) e distribuzione.

Venne scritto inizialmente da Robert Gentleman e Ross Ihaka del dipartimento di Statistica dell'Università di Auckland in nuova Zelanda nel 1996, a questo progetto contribuirono poi altri programmatori e studiosi per far uscire nel 1999 la prima versione. Dopo l'affluenza di altri collaboratori che hanno reso il programma dinamico e al passo con i tempi, fino ad oggi ne sono uscite circa trenta versioni.

Attraverso il sito www.r-project.org è possibile non solo scaricare il programma ma anche ottenere tante altre informazioni quali i manuali per l'uso, le discussioni e le *Frequent Asked Questions* riguardo il funzionamento del programma, l'elenco dei libri e l'accesso agli articoli della rivista "The R Journal".

Grazie alla composizione modulare del software, si possono infine scaricare tutti i pacchetti aggiuntivi dall'area CRAN del sito internet, ovvero *Comprehensive R Archive Network*. Tramite questi moduli, anch'essi distribuiti con licenza GPL, è possibile estendere le capacità del programma, collegarlo con database e sistemi informativi territoriali nonché adattarlo alle proprie esigenze operative e computazionali.

L'interfaccia di R è a riga di comando, ogni scrittura produce un *output* che può essere semplicemente visualizzato oppure legato ad un oggetto a cui si attribuisce un nome, così da poterlo facilmente richiamare per potervi apportare calcoli ed elaborazioni. Per quanto siano disponibili sul sito web (nella sezione *R GUI projects*) diverse interfacce grafiche, il vantaggio delle stringhe di codice è proprio quello della rapidità: dopo aver personalizzato una sequenza di operazioni su un *editor* di testo, questa si può riscrivere all'interno del programma, anche ripetutamente, cambiando di volta in volta solo alcuni parametri.

In questo modo è possibile eseguire elaborazioni lunghe e complesse, ripetute con diverse impostazioni, senza dover cliccare una miriade di comandi (come avverrebbe con un'interfaccia a finestre di comando) ma semplicemente cambiando qualche lettera nella sequenza dei comandi preparata sull'*editor* di testo e riproponendola per intero.

Una volta raggiunta questa confidenza con il programma, tutte le difficoltà nell'approccio iniziale saranno dimenticate e sarà possibile sfruttare tutta la rapidità tecnica che può offrire.

Nell'appendice di questo elaborato è disponibile l'elenco dei comandi utilizzati per effettuare una *cluster analysis* su una base di dati. Oltre alle stringhe di codice vi è poi una spiegazione del loro significato all'interno della procedura di raggruppamento per consentirne un utilizzo più consapevole.

CAPITOLO 4

PARTIZIONE DELLE REGIONI ITALIANE IN GRUPPI

4.1 La matrice dei dati

Nel precedente capitolo è stata spiegata la procedura statistica che sarà applicata, tramite il software R, sulla base di dati raccolti riguardo le regioni italiane. Il risultato finale sarà una partizione delle regioni in un determinato numero di *cluster*, ognuno dei quali sarà commentato e spiegato nelle sue caratteristiche competitive.

Le variabili scelte per costruire la matrice dei dati, nonché i relativi valori per ogni regione, sono tutte quelle esposte nel Paragrafo 2.2, ad eccezione della variabile Pil pro capite. Questa scelta deriva dal fatto che questo indicatore, così rilevante e generale, si può meglio definire come espressione della competitività di un sistema economico, piuttosto che uno degli aspetti che ne determina il profilo competitivo.

Dunque non sarà considerato al pari delle altre variabili che spiegano il raggruppamento finale, ma sarà spiegato come manifestazione della competitività del gruppo. In altre parole, sarà la combinazione di caratteristiche peculiari di ogni gruppo a dare peso al livello di Pil pro capite medio delle regioni che comprende.

Questa chiave di lettura del Pil pro capite ci fornisce un possibile obiettivo nella nostra ricerca statistica. È dunque questo lo scopo che ci aiuterà a posizionarci nel *trade-off* tra sintesi e affidabilità dei dati, in parole più concrete, ci permetterà di scegliere quel numero ottimo di gruppi che rispetta maggiormente la distribuzione della variabile Pil tra le regioni italiane.

Nella pagina seguente è riportata la matrice dei dati che costituisce la base della nostra ricerca. I valori indicati sono gli stessi che compaiono nei grafici del Paragrafo 2.2, le regioni sono state tuttavia ordinate secondo una graduatoria in base alla Produttività del lavoro, seguendo l'esempio dell'Istat.

A seguire è stata riportata la standardizzazione della matrice dei dati, su questa è stata effettuata la procedura della *cluster analysis*. Da questa matrice è possibile vedere più facilmente la graduatoria delle regioni per una data variabile nonché lo stacco tra la posizione di una regione e quella di un'altra.

Fig. 4.1a – Crescita reale, apertura internazionale, conoscenza e spesa pubblica nelle regioni italiane

	Produttività del lavoro	Costo del lavoro	Tasso di natalità delle imprese	Valore agg. grandi imprese	Scambi di merci col mondo	Addetti imprese estere partecipate	Addetti imprese a part.ne estera	Spesa R&S	Brevetti	Laureati	Investimenti pubblici
Lombardia	€ 65.404,9	0,4630	6,24%	4,45%	-1,56%	13,82%	11,03%	1,27%	124,63	3,91	1,61%
Valle d'Aosta	€ 65.266,3	0,3953	5,57%	1,75%	7,39%	0,84%	3,63%	0,66%	58,59	4,55	6,55%
Trentino Alto Adige	€ 62.332,0	0,4600	4,93%	2,30%	2,63%	3,54%	3,90%	1,31%	91,52	3,71	5,05%
Lazio	€ 60.568,3	0,4872	8,27%	1,82%	-6,87%	11,17%	7,57%	1,82%	38,09	5,35	2,09%
Liguria	€ 58.630,4	0,4415	6,21%	2,47%	-9,48%	2,08%	4,63%	1,36%	62,16	3,88	2,24%
Emilia Romagna	€ 58.242,8	0,4870	5,86%	5,31%	15,00%	6,13%	3,46%	1,37%	147,50	3,88	1,75%
Veneto	€ 57.396,3	0,4730	5,65%	4,25%	9,01%	8,55%	2,68%	1,08%	92,59	4,31	1,94%
Friuli Venezia Giulia	€ 57.291,4	0,5005	5,52%	3,86%	12,78%	4,55%	3,63%	1,49%	83,21	4,25	2,67%
Toscana	€ 56.350,6	0,4514	6,30%	2,78%	9,50%	4,62%	2,60%	1,23%	64,22	3,89	2,23%
Piemonte	€ 55.633,4	0,4710	6,33%	5,53%	10,41%	25,55%	5,68%	1,84%	91,80	3,96	1,60%
Sicilia	€ 51.890,6	0,4789	7,63%	1,47%	-9,16%	0,45%	0,29%	0,86%	3,20	4,78	2,90%
Abruzzo	€ 51.711,8	0,4842	7,89%	5,06%	11,61%	2,35%	5,08%	0,94%	27,60	5,77	5,93%
Marche	€ 50.943,7	0,4577	5,94%	3,20%	7,60%	7,70%	0,93%	0,71%	91,21	5,16	1,69%
Campania	€ 50.687,4	0,4954	7,95%	1,70%	-1,08%	1,21%	0,70%	1,26%	13,69	5,50	2,83%
Umbria	€ 50.543,5	0,4558	5,98%	2,89%	6,55%	2,23%	2,06%	0,98%	20,47	4,77	2,11%
Puglia	€ 49.525,2	0,4855	7,08%	1,91%	-1,64%	0,90%	0,68%	0,77%	11,46	5,75	2,15%
Molise	€ 49.452,7	0,4266	6,67%	3,39%	-0,26%	1,02%	0,34%	0,50%	3,13	7,15	5,64%
Sardegna	€ 48.861,7	0,4866	6,82%	1,39%	-12,63%	0,72%	0,63%	0,67%	13,94	4,45	4,00%
Basilicata	€ 48.445,5	0,4628	6,42%	3,63%	3,97%	0,33%	0,16%	0,67%	5,10	6,90	3,19%
Calabria	€ 47.674,8	0,4530	7,70%	0,30%	-0,61%	0,22%	0,35%	0,46%	5,52	6,16	5,38%

Fonte: elaborazione su dati Istat, Unioncamere, ICE

Fig. 4.1b - Crescita reale, apertura internazionale, conoscenza e spesa pubblica nelle regioni italiane a valori standardizzati

	Produttività del lavoro	Costo del lavoro	Tasso di natalità delle imprese	Valore agg. grandi imprese	Scambi di merci col mondo	Addetti imprese estere partecipate	Addetti imprese a part.ne estera	Spesa R&S	Brevetti	Laureati	Investimenti pubblici
Lombardia	1,87	-0,11	-0,33	1,02	-0,52	1,44	2,83	0,52	1,62	-0,97	-0,96
Valle d'Aosta	1,85	-2,80	-1,04	-0,85	0,59	-0,65	0,22	-1,00	0,14	-0,35	2,07
Trentino Alto Adige	1,33	-0,23	-1,73	-0,46	0,00	-0,22	0,32	0,60	0,88	-1,15	1,15
Lazio	1,01	0,85	1,84	-0,79	-1,19	1,01	1,61	1,86	-0,32	0,43	-0,67
Liguria	0,67	-0,96	-0,36	-0,35	-1,51	-0,45	0,58	0,73	0,22	-0,99	-0,57
Emilia Romagna	0,60	0,84	-0,73	1,61	1,54	0,20	0,16	0,76	2,14	-0,99	-0,88
Veneto	0,45	0,29	-0,97	0,88	0,79	0,59	-0,11	0,04	0,90	-0,58	-0,76
Friuli Venezia Giulia	0,43	1,38	-1,10	0,62	1,26	-0,06	0,22	1,04	0,69	-0,63	-0,31
Toscana	0,27	-0,57	-0,26	-0,13	0,85	-0,04	-0,14	0,41	0,26	-0,98	-0,58
Piemonte	0,14	0,20	-0,24	1,76	0,97	3,33	0,95	1,92	0,89	-0,91	-0,97
Sicilia	-0,52	0,52	1,16	-1,04	-1,47	-0,72	-0,96	-0,51	-1,11	-0,12	-0,17
Abruzzo	-0,55	0,73	1,44	1,44	1,11	-0,41	0,73	-0,29	-0,56	0,84	1,69
Marche	-0,69	-0,32	-0,66	0,16	0,62	0,45	-0,73	-0,87	0,87	0,25	-0,92
Campania	-0,74	1,17	1,50	-0,88	-0,47	-0,59	-0,81	0,49	-0,87	0,58	-0,22
Umbria	-0,76	-0,40	-0,61	-0,05	0,48	-0,43	-0,33	-0,20	-0,72	-0,13	-0,66
Puglia	-0,94	0,78	0,57	-0,73	-0,53	-0,64	-0,82	-0,72	-0,92	0,82	-0,63
Molise	-0,95	-1,55	0,13	0,29	-0,36	-0,63	-0,94	-1,37	-1,11	2,17	1,51
Sardegna	-1,06	0,82	0,29	-1,10	-1,90	-0,67	-0,84	-0,97	-0,87	-0,44	0,51
Basilicata	-1,13	-0,12	-0,14	0,46	0,16	-0,74	-1,00	-0,96	-1,07	1,93	0,01
Calabria	-1,27	-0,51	1,24	-1,85	-0,41	-0,75	-0,94	-1,48	-1,06	1,22	1,35

Fonte: elaborazione su dati Istat, Unioncamere, ICE

4.2 Cluster analysis: metodi gerarchici

I metodi gerarchici applicati sulla matrice degli scarti standardizzati sono tutti quelli esposti nel capitolo 3. Riportiamo in seguito i dendrogrammi di ogni metodologia, essendo questi i grafici più intuitivi e rappresentativi della sequenza delle operazioni di raggruppamento.

Fig. 4.2a - Metodo di Ward - Dendrogramma

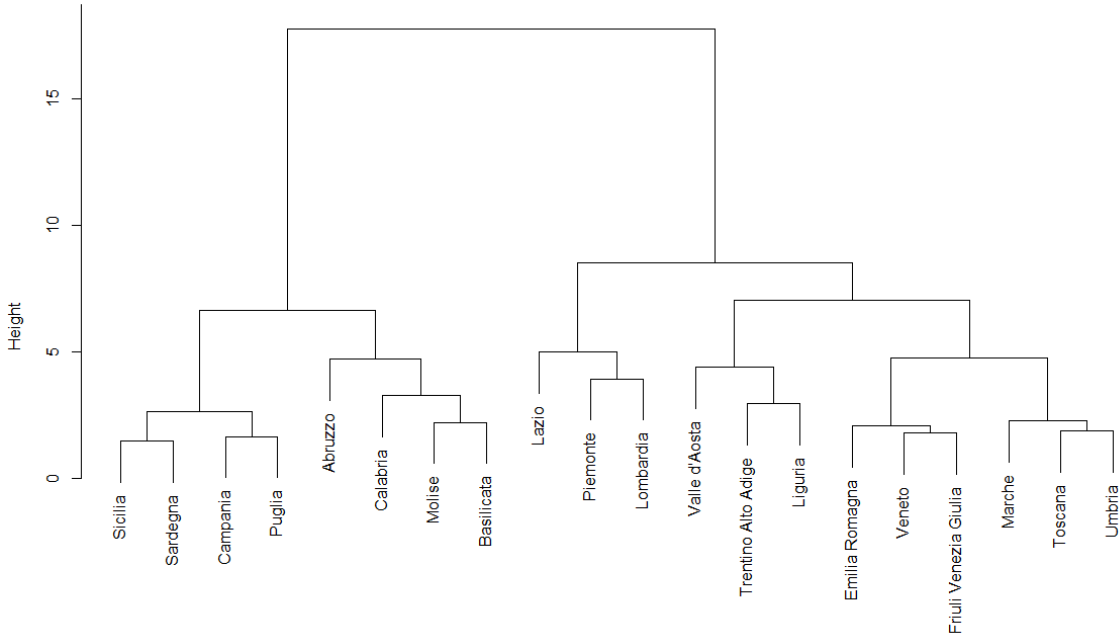


Fig. 4.2b - Metodo del legame medio - Dendrogramma

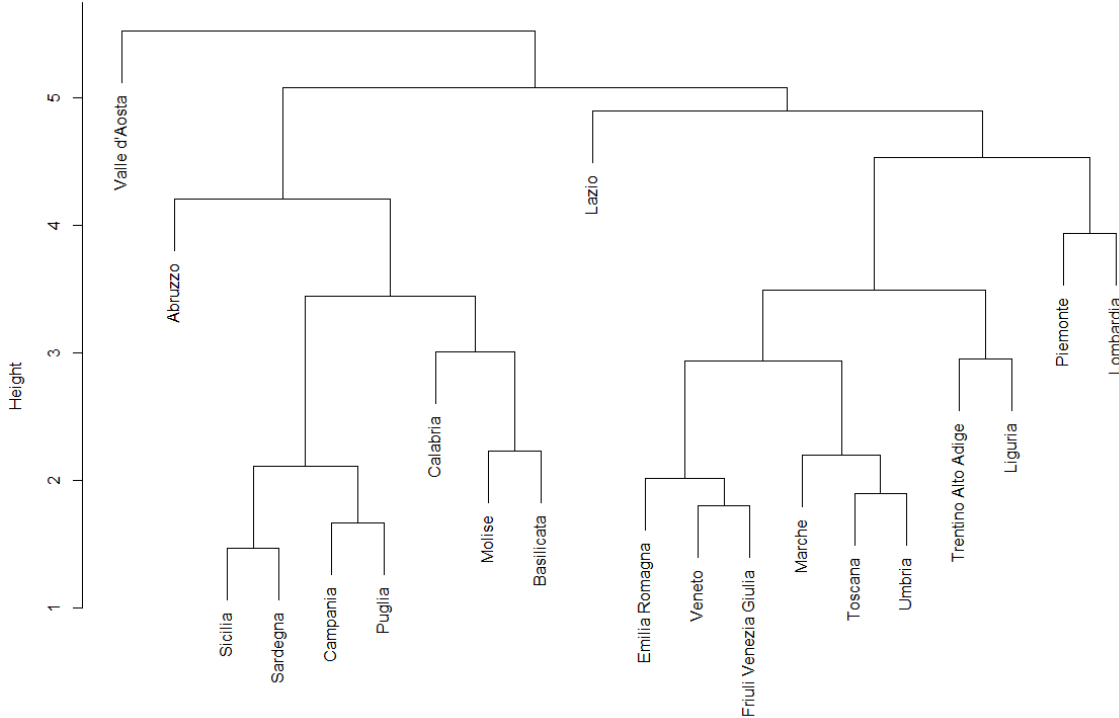


Fig. 4.2c - Metodo del legame completo - Dendrogramma

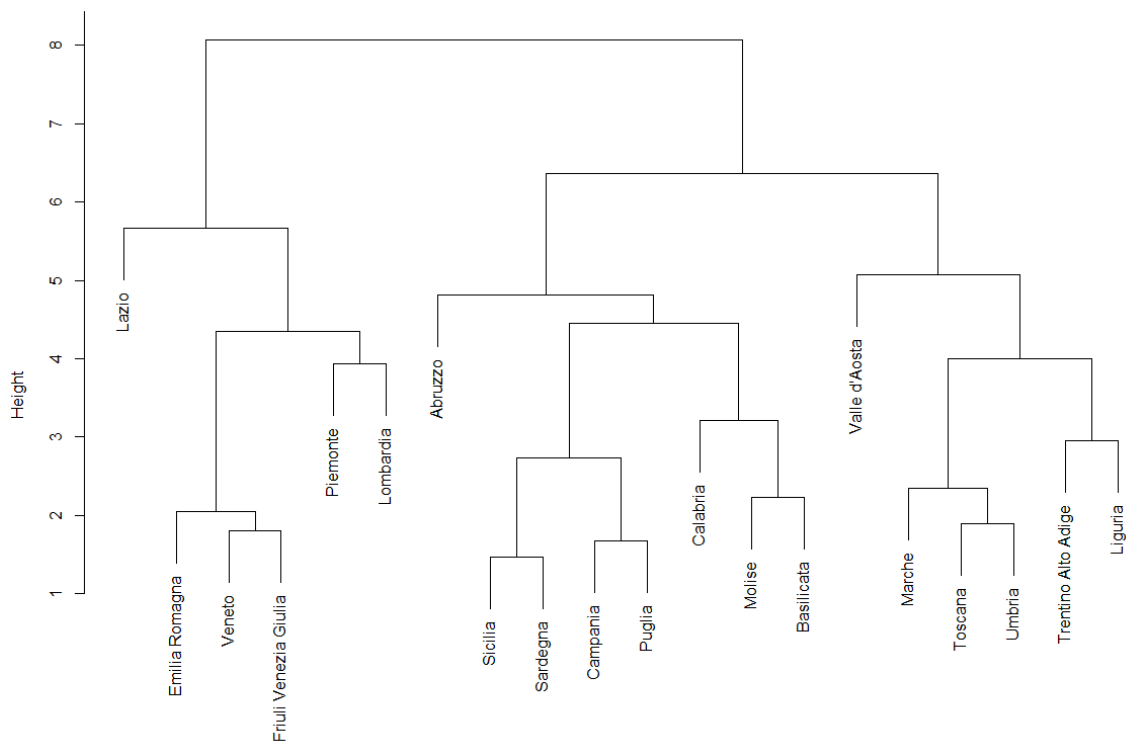
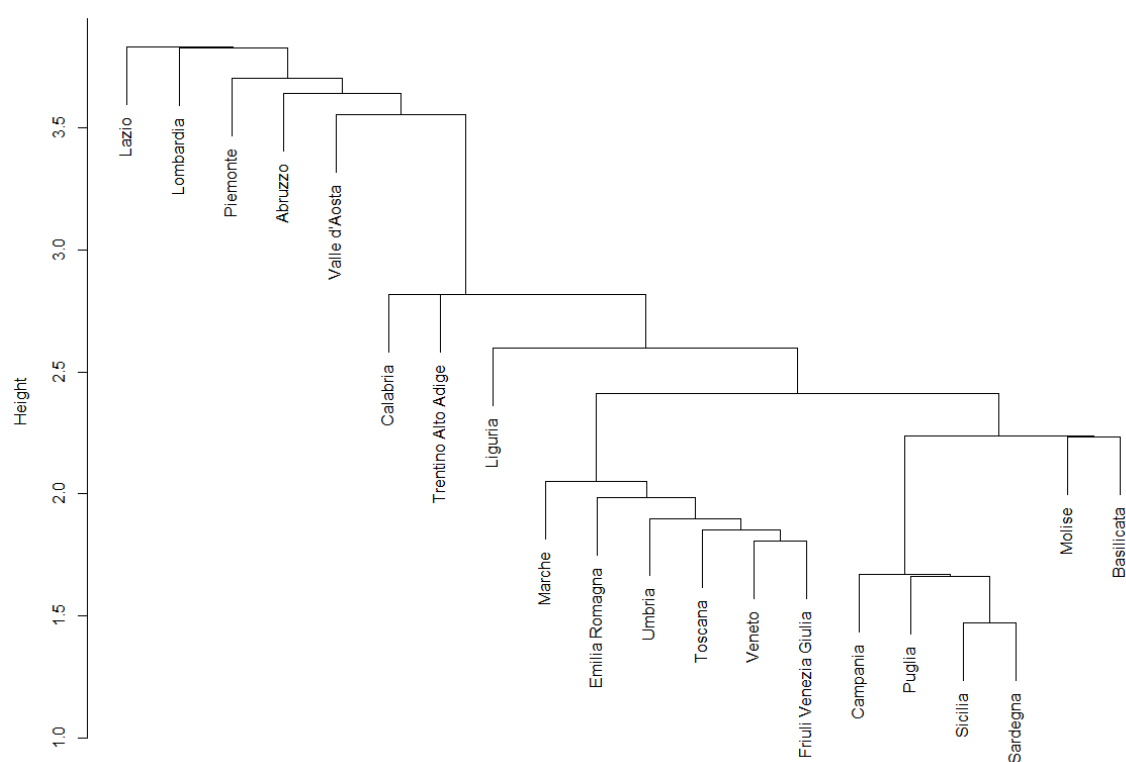


Fig. 4.2d - Metodo del legame singolo - Dendrogramma



Il primo aspetto da discutere sono i raggruppamenti che si ritrovano in tutti questi metodi gerarchici. Vi sono, cioè, alcune coppie di regioni che, a prescindere dalla metodologia, presentano una notevole similarità tanto da ritrovarsi sempre raggruppate a prescindere dalla metodologia gerarchica di *cluster analysis*. Tra queste evidenziamo:

(Sicilia – Sardegna);
(Molise – Basilicata);
(Veneto – Friuli Venezia Giulia);

Queste coppie si ritrovano in tutti e quattro i metodi aggregativi, se poi escludiamo il metodo del legame singolo vi sono altre coppie che persistono negli altri tre metodi:

(Piemonte – Lombardia);
(Toscana – Umbria);
(Trentino Alto Adige – Liguria);
(Campania – Puglia);

Si può notare come queste coppie di regioni presentino una relativa vicinanza territoriale, infatti nella maggior parte dei casi le regioni hanno un confine in comune. La coppia Sicilia – Sardegna presenta la stessa conformazione insulare mentre in altri casi, quali Trentino Alto Adige – Liguria e Molise – Basilicata, le regioni condividono solo la stessa fascia territoriale della penisola italiana.

Le regioni rimanenti sono quelle che presentano valori più distinti e che per questo risultano più difficili da collocare. Nel dendrogramma di Ward e in quello del legame medio sono l'Emilia Romagna e le Marche che si aggregano più facilmente, ovvero con una distanza minore rispetto alle rimanenti Calabria, Lazio, Abruzzo e Valle d'Aosta. Nel metodo del legame completo, invece, sono Abruzzo e Marche che registrano uno stacco minore, sempre rispetto alle rimanenti.

Il metodo del legame singolo è un caso a parte. Poiché esso si basa sul concetto che la distanza tra due gruppi è data dalla minore delle distanze tra ciascuna delle unità in un gruppo e ciascuna unità nell'altro, si manifesta spesso il cosiddetto “effetto catena”, ovvero non si distinguono gruppi omogenei sin dall'inizio della procedura ma un unico gruppo a cui vengono gradualmente assegnate tutte le rimanenti unità, come fossero gli anelli di una catena. In questo modo si vanno a riunire elementi anche molto distanti tra loro grazie ad una successione di punti intermedi, si ottiene poi un dendrogramma che presenta la minima lunghezza (*minimum spanning tree*) come si può infatti notare dalla figura precedente.

L'ultima nota la merita il dendrogramma di Ward. Questo si distingue per la capacità di individuare *cluster* che hanno una netta separazione esterna: all'altezza 5,5 sono ben evidenti cinque gruppi di simile numerosità che saranno riuniti ad alti livelli di distanza; all'altezza 10 si distinguono due grandi gruppi, che mostrano una netta frattura fra regioni del Sud in uno e regioni del Nord nell'altro con il Centro mescolato tra i due, che saranno riuniti dopo un lungo stacco, una configurazione di così netto contrasto che nessun altro metodo gerarchico era riuscito ad individuare.

Possiamo a questo punto utilizzare i risultati della *cluster analysis* gerarchica per definire un intervallo ragionevole di possibili numerosità del raggruppamento.

Il metodo di Ward, per un'altezza di circa 5,5, suggerisce cinque gruppi, il metodo del legame medio, per un'altezza di circa 4,5, ne suggerisce quattro, due dei quali composti da

una sola regione, infine il metodo del legame completo va a formare quattro gruppi, di cui uno formato solo dal Lazio, per un'altezza di 5,5. Il legame singolo è decisamente poco utilizzabile in questo contesto, in seguito ne confermeremo la bassa affidabilità calcolando i valori di R^2 per l'intervallo di numerosità.

Un numero minimo di quattro gruppi si può ipotizzare per un raggruppamento ragionevolmente sintetico, nello stabilire il numero massimo di gruppi occorre tener conto del numero ridotto di unità statistiche.

Con 4 gruppi la numerosità media di ognuno è di cinque unità, se arriviamo ad ipotizzare un raggruppamento con il doppio di gruppi, ovvero 8, la numerosità media di ognuno sarebbe 2,5 secondo un ragionevole livello di variabilità dei dati. Questi due valori, 4 e 8, come estremi dell'intervallo "numero di gruppi" possono costituire il *trade-off* tra utilità e affidabilità dei dati ad un buon livello di dettaglio.

Nel prossimo paragrafo analizzeremo la valutazione con l'indice R^2 di tutti i metodi gerarchici per ogni valore di questo intervallo.

4.2.1 Valutazione della bontà delle partizioni

L'indice R^2 rappresenta il miglior parametro per valutare la bontà di una partizione, misurandone la separazione esterna e la coesione interna attraverso il calcolo della Devianza fra gruppi e Devianza nei gruppi.

Nella seguente tabella sono elencati tutti i metodi gerarchici discussi nel precedente paragrafo e per ognuno si presenta un possibile taglio del dendrogramma, ovvero a formare 4, 5, 6, 7 oppure 8 gruppi, che sono i valori dell'intervallo "numero di gruppi" precedentemente stabilito. La partizione delle regioni che si ottiene per ogni caso si può leggere dai precedenti grafici a dendrogramma ponendo una linea in alto e facendola scendere gradualmente finché da essa non partiranno un numero di linee pari al dato numero di gruppi (le regioni che si ramificano da ogni linea appartengono ad un gruppo distinto).

Ogni cella della tabella rappresenta un possibile raggruppamento ragionevole, la tabella riporta per ogni raggruppamento il valore dell'indice R^2 che permetterà di fare una classifica dei migliori raggruppamenti.

Tab. 4.2.1 – Valori dell'indice R^2 per metodo gerarchico e numero di gruppi

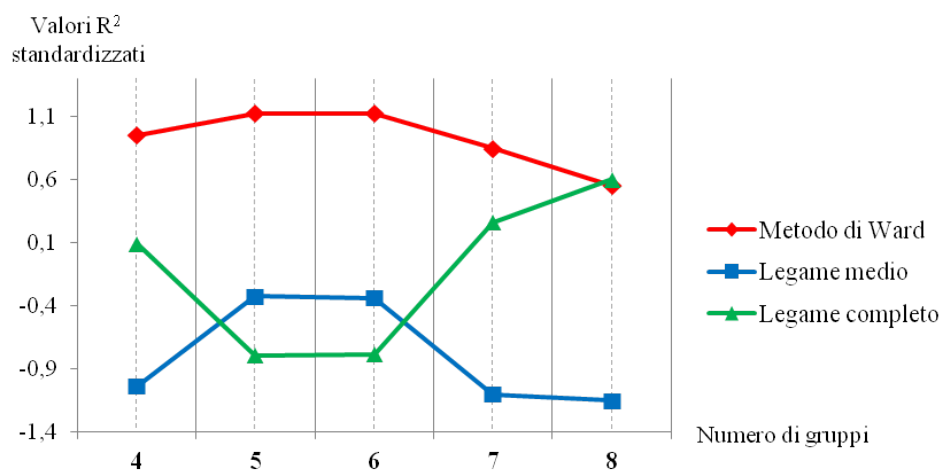
	Numero Gruppi				
	4	5	6	7	8
Metodo di Ward	56,0%	63,5%	69,5%	74,1%	79,2%
Legame medio	51,4%	61,3%	67,1%	70,8%	76,1%
Legame completo	54,0%	60,6%	66,4%	73,1%	79,3%
Legame singolo	30,5%	36,0%	45,2%	51,6%	56,7%

Possiamo innanzitutto notare come il valore dell'indice, quale che sia il metodo gerarchico di raggruppamento, tende a crescere all'aumentare del numero di gruppi, una considerazione piuttosto intuitiva in quanto formando più gruppi e meno numerosi aumenterà la variabilità tra gruppi nonché si ridurrà la variabilità tra le unità interne ai gruppi.

La tabella evidenzia come il metodo di Ward proponga raggruppamenti che risultano, per ogni dato numero di *cluster*, i più efficaci rispetto agli altri metodi, seppur alla fine viene leggermente superato dal metodo del legame medio. Il legame singolo, come già ci aspettavamo, presenta valori di bontà della partizione notevolmente inferiori rispetto agli altri metodi gerarchici.

Per visualizzare meglio il confronto tra i diversi metodi per ogni numero di *cluster*, dunque senza considerare la naturale crescita di tutti i valori di R^2 all'aumentare di questo numero, andiamo a standardizzare la precedente tabella (per colonna) senza tuttavia considerare il metodo del legame singolo, poco rilevante. I risultati sono esposti nel seguente grafico che consente di vedere non solo quale dei metodi risulta più efficace per ogni numero di *cluster* ma anche lo stacco che c'è tra le loro posizioni.

Fig. 4.2.1 – Valori standardizzati dell'indice R^2 per metodo gerarchico e numero di gruppi



Dalla figura risalta la prevalenza del metodo di Ward per quasi tutte le postazioni, gli altri due metodi gerarchici oscillano in efficacia della partizione e si sorpassano a vicenda.

La conclusione è che la miglior *cluster analysis* gerarchica è quella svolta con il metodo di Ward, di conseguenza sarà questo che andrà a guidare il metodo delle k-medie per ogni possibile numero di *cluster*.

4.3 Cluster analysis: metodo delle k-medie

Il metodo non gerarchico delle k-medie permette al termine di un'analisi dei gruppi gerarchica di renderne più flessibili i risultati, ovvero adottare alcuni aggiustamenti partendo da alcune impostazioni di base.

In questa sede applicheremo il metodo delle k-medie per ogni numero di gruppi dell'intervallo stabilito, in tutti i casi sarà il metodo di Ward a fornire l'impostazione di base, ovvero i poli iniziali. Questi sono calcolati come i centroidi dei gruppi ottenuti con il metodo di Ward tagliandone il dendrogramma in modo da avere da 4 a 8 gruppi. Definita questa procedura, riportiamo in seguito i risultati del metodo delle k-medie "guidato" dal metodo gerarchico prescelto. Questi sono i vettori di allocazione delle unità in un numero di gruppi prestabilito, il numero accanto ad ogni regione indica dunque il *cluster* di appartenenza.

Tab. 4.3 – Raggruppamenti secondo k-medie (basato sulla cluster analysis con il metodo di Ward) per numero di gruppi

4 Gruppi		5 Gruppi		6 Gruppi		7 Gruppi		8 Gruppi	
Piemonte	1	Piemonte	1	Piemonte	1	Piemonte	1	Piemonte	1
Lombardia	1	Lombardia	1	Lombardia	1	Lombardia	1	Lombardia	1
Lazio	1	Lazio	1	Valle d'Aosta	2	Valle d'Aosta	2	Valle d'Aosta	2
Valle d'Aosta	2	Valle d'Aosta	2	Trentino Alto Adige	2	Trentino Alto Adige	2	Trentino Alto Adige	2
Trentino Alto Adige	2	Trentino Alto Adige	2	Liguria	2	Veneto	3	Veneto	3
Liguria	2	Liguria	2	Veneto	3	Friuli Venezia Giulia	3	Friuli Venezia Giulia	3
Veneto	3	Veneto	3	Friuli Venezia Giulia	3	Emilia Romagna	3	Emilia Romagna	3
Friuli Venezia Giulia	3	Friuli Venezia Giulia	3	Emilia Romagna	3	Liguria	4	Liguria	4
Emilia Romagna	3	Emilia Romagna	3	Toscana	3	Toscana	4	Toscana	4
Toscana	3	Toscana	3	Umbria	3	Umbria	4	Umbria	4
Umbria	3	Umbria	3	Marche	3	Marche	4	Marche	4
Marche	3	Marche	3	Lazio	4	Lazio	5	Lazio	5
Abruzzo	4	Abruzzo	4	Abruzzo	5	Abruzzo	6	Abruzzo	6
Molise	4	Molise	4	Molise	5	Molise	6	Molise	7
Campania	4	Basilicata	4	Basilicata	5	Basilicata	6	Basilicata	7
Puglia	4	Campania	5	Campania	6	Campania	7	Calabria	7
Basilicata	4	Puglia	5	Puglia	6	Puglia	7	Campania	8
Calabria	4	Calabria	5	Calabria	6	Calabria	7	Puglia	8
Sicilia	4	Sicilia	5	Sicilia	6	Sicilia	7	Sicilia	8
Sardegna	4	Sardegna	5	Sardegna	6	Sardegna	7	Sardegna	8

Dalla precedente tabella si può facilmente notare come le coppie di regioni che persistevano in tutti i metodi gerarchici sono sempre rispettate. L'unica che viene scissa è la coppia (Trentino Alto Adige – Liguria) a partire dal raggruppamento a 7 gruppi, con questa divisione si guadagna tuttavia una maggior coerenza nella vicinanza territoriale. Un'altra osservazione riguarda le unità più anomale e distinte da quelle più facilmente raggruppabili, quali il Lazio e l'Abruzzo che al crescere del numero dei gruppi tendono a isolarsi come regioni a sé stanti (da 6 gruppi per il Lazio, e da 8 gruppi per l'Abruzzo). Infine i risultati riportano come il raggruppamento k-medie per 4 gruppi sia lo stesso proposto dal metodo di Ward (tagliando il dendrogramma alla minima altezza per formare 4 gruppi).

Definiti i risultati del metodo delle k-medie possiamo a questo punto misurarne l'efficacia di partizione attraverso l'indice R^2 .

4.3.1 Valutazione della bontà delle partizioni

I raggruppamenti offerti dal metodo delle k-medie rappresentano potenzialmente un netto miglioramento rispetto a quelli ottenuti con il prescelto tra i metodi gerarchici. Per verificare quanto di questa potenza sia diventata atto ci serviremo della valutazione dell'indice R^2 .

Nella seguente tabella sono esposti i valori dell'indice per i raggruppamenti proposti dal metodo delle k-medie guidato nonché per le partizioni dei metodi gerarchici più rilevanti.

Tab. 4.3.1a – Valori dell'indice R^2 per metodo (gerarchico e non) e numero di gruppi

	Numero Gruppi				
	4	5	6	7	8
Metodo delle k-medie	56,0%	63,7%	69,7%	74,8%	79,6%
Metodo di Ward	56,0%	63,5%	69,5%	74,1%	79,2%
Legame medio	51,4%	61,3%	67,1%	70,8%	76,1%
Legame completo	54,0%	60,6%	66,4%	73,1%	79,3%

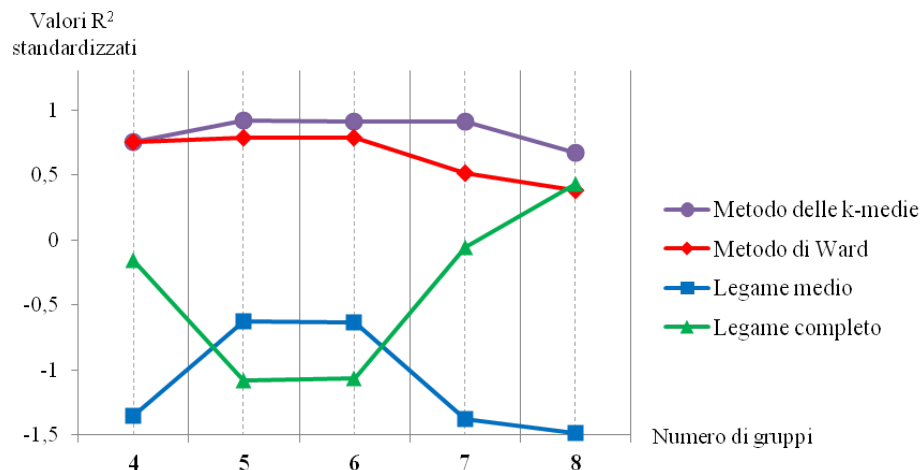
I risultati confermano le aspettative. Per quasi tutti i numeri dei gruppi, il metodo delle k-medie guidato dal prescelto tra i metodi gerarchici propone un raggruppamento migliore della sua guida, ad eccezione del caso con 4 gruppi in cui, essendo le partizioni identiche, viene valutato allo stesso modo.

Per rendere più stabili e comprovati i risultati ottenuti è stata ripetuta la *cluster analysis* con il metodo delle k-medie stabilendo tuttavia i poli iniziali della procedura in modo casuale. Ripetendo questa operazione per numerose volte (poiché k-medie casuale non presenta sempre gli stessi risultati) e per ogni numero di gruppi, i raggruppamenti ottenuti con la guida del metodo di Ward sono risultati i più efficienti secondo la valutazione dell'indice R^2 .

Possiamo a questo punto affermare che le partizioni ottenute con il metodo delle k-medie guidato dal Metodo di Ward, espone nella tabella 4.3, sono quelle con la massima valutazione di R^2 tra tutti i metodi di *cluster analysis* considerati.

Anche stavolta illustriamo graficamente la classifica dei vari metodi di partizione per ogni numero di gruppi attraverso la standardizzazione della precedente tabella.

Fig. 4.3.1b – Valori standardizzati dell'indice R^2 per metodo (gerarchico e non) e numero di gruppi



La partizione di k-medie che presenta il maggior stacco nella valutazione dell'indice R^2 è quella per 7 gruppi.

4.4 Scelta della partizione finale

La scelta della miglior partizione, nella rosa di possibilità offerta dal metodo delle k-medie basato sui risultati del metodo di Ward, coincide a questo punto con la scelta del numero ottimo di gruppi.

Consideriamo allora l'obiettivo della nostra analisi, ovvero definire un profilo competitivo per ogni *cluster* di regioni italiane che possa spiegare il livello di Pil medio delle regioni in esso comprese.

Sotto questa ottica, la miglior partizione avrebbe un riscontro pratico con la classifica delle regioni in base al Pil (paragrafo 2.2.1): collegando tra loro le regioni che appartengono al medesimo gruppo, in mezzo a queste nella classifica non dovrebbero esserci altre regioni. In altre parole il raggruppamento scelto si dovrebbe poter rivedere nella graduatoria del Pil divisa per fasce, in cui ogni fascia corrisponde ad un gruppo.

Per cercare il raggruppamento che più rispetta questo principio, ovvero che minimizza la variabilità del valore Pil in tutti i gruppi di regioni, possiamo servirci ancora una volta dell'indice R^2 . Sarà sufficiente considerare per ogni regione la sola variabile Pil per poi valutare i precedenti raggruppamenti (ottenuti considerando tutte le variabili ad esclusione del Pil) in termini di coesione interna e separazione esterna.

I risultati dell'indice sono di 84,66%, 85,30%, 85,49%, 94,27% e 94,53% calcolati considerando solo il valore del Pil delle regioni raggruppate secondo le precedenti partizioni ottenute con k-medie e composte da, rispettivamente, 4, 5, 6, 7, e 8 gruppi.

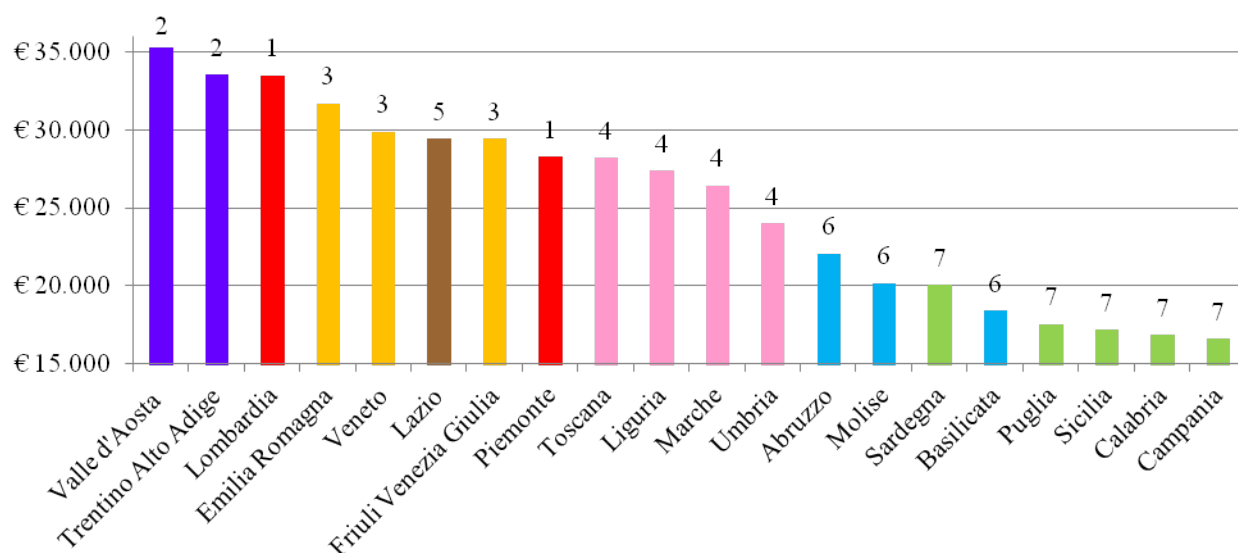
Possiamo notare, innanzitutto, come i valori dell'indice siano crescenti, poiché anche considerando solo una variabile, all'aumentare dei gruppi diminuisce la variabilità tra le unità. Inoltre la grandezza di queste percentuali sta a dimostrare come i raggruppamenti ottenuti, considerando le variabili competitive delle regioni, trovino il loro riscontro nella variabile Pil. In altre parole, dopo aver raggruppato le regioni considerando tutte le altre variabili, i gruppi ottenuti registrano valori del Prodotto interno lordo omogenei tra loro ed eterogenei tra gruppi diversi.

Infine i valori percentuali dimostrano piccoli incrementi della bontà della partizione da 4 a 6 gruppi come da 7 a 8 gruppi, tuttavia da 6 a 7 gruppi lo stacco non è di qualche decimale ma di quasi il 9%.

La partizione che sarà scelta come conclusione della *cluster analysis* sarà dunque quella ottenuta con il metodo delle k-medie, calcolato con i poli iniziali definiti con il metodo di Ward, a 7 gruppi. Scegliere quella da otto gruppi è poco sensato in quanto per l'aggiunta di un gruppo (formato da una sola regione) si ha un ridottissimo aumento dell'indice R^2 considerando solo il Pil.

Riportiamo in seguito la graduatoria delle regioni in base ai valori del Pil pro capite, questa volta ogni barra sarà colorata e numerata in modo da evidenziare i sette *cluster* della partizione e dunque quanto sia ridotta la variabilità dei valori del Pil.

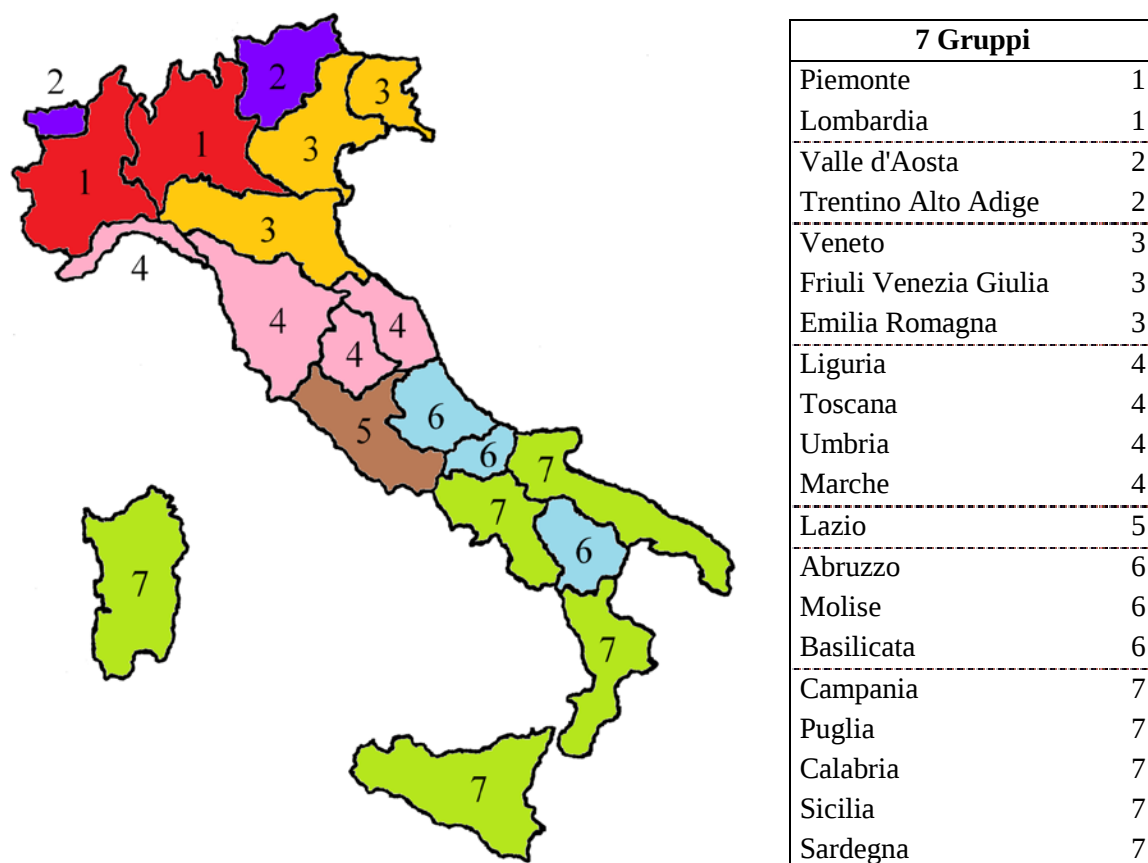
Fig. 4.4a – Pil pro capite annuo nelle regioni italiane (2011), colori e numeri in base al cluster di appartenenza



Da questo grafico possiamo notare come i valori più distanti che ricadono nel medesimo *cluster* sono quelli della Lombardia con il Piemonte nonché della Sardegna con le altre regioni in fondo alla classifica. Riguardo il Lazio, per quanto si trovi in una posizione intermedia senza appartenere allo stesso gruppo delle regioni che lo circondano, la variabilità delle unità non viene danneggiata in quanto va a formare un gruppo con una sola unità.

Per rappresentare in modo efficace la coerenza territoriale della partizione scelta, riportiamo in seguito l'insieme delle regioni italiane ognuna con un colore e un numero a seconda del *cluster* di appartenenza. I gruppi ottenuti comprendono quasi sempre regioni confinanti tra loro, inoltre si trovano distribuiti per fasce territoriali (Nord, Centro, Sud) in cui solo le regioni centrali condividono gruppi con regioni del Nord o del Sud.

Fig. 4.4b – Regioni d'Italia suddivise secondo la partizione scelta



4.5 I cluster regionali

L'analisi dei gruppi applicata sulle regioni italiane ha determinato una partizione composta da sette cluster regionali. Ogni cluster comprende un diverso insieme di regioni sul quale è possibile calcolare alcune statistiche descrittive, quali la media e la deviazione standard delle unità, che possano rappresentare il gruppo intero.

L'unico gruppo composto da una sola regione è il numero 5, il Lazio. Di questo gruppo i valori rappresentativi saranno dunque quelli registrati per questa regione.

Nella seguente tabella riportiamo la media e la deviazione standard, calcolate per ogni variabile precedentemente considerata, per ogni gruppo della partizione ed evidenziando le deviazioni standard più alte per ogni variabile, l'insieme dei gruppi è ordinato in base ai valori medi registrati per la variabile Pil pro capite.

Tab. 4.5a – Media e deviazione standard per ogni variabile delle regioni raggruppate in sette gruppi

	Pil pro capite	Produttività del lavoro	Costo del lavoro	Tasso di natalità delle imprese	Valore agg. grandi imprese	Scambi di merci col mondo	Addetti imprese estere partecipate	Addetti imprese a part.ne estera	Spesa R&S	Brevetti	Laureati	Investimenti pubblici
Gruppo 2	€ 34.410,6 ±(€ 1.208,1)	€ 63.799,2 ±(€ 2.074,9)	0,4277 ±(0,0458)	5,25% ±(0,45%)	2,02% ±(0,39%)	5,01% ±(3,36%)	2,19% ±(1,91%)	3,76% ±(0,19%)	0,98% ±(0,46%)	75,06 ±(23,28)	4,13 ±(0,59)	5,80% ±(1,06%)
Gruppo 1	€ 30.880,3 ±(€ 3.681,9)	€ 60.519,1 ±(€ 6.909,5)	0,4670 ±(0,0057)	6,28% ±(0,07%)	4,99% ±(0,76%)	4,43% ±(8,47%)	19,68% ±(8,30%)	8,35% ±(3,78%)	1,56% ±(0,40%)	108,22 ±(23,22)	3,94 ±(0,04)	1,61% ±(0,01%)
Gruppo 3	€ 30.324,2 ±(€ 1.206,0)	€ 57.643,5 ±(€ 521,6)	0,4869 ±(0,0137)	5,68% ±(0,17%)	4,47% ±(0,75%)	12,26% ±(3,03%)	6,41% ±(2,01%)	3,26% ±(0,50%)	1,31% ±(0,21%)	107,77 ±(34,73)	4,15 ±(0,23)	2,12% ±(0,49%)
Gruppo 5	€ 29.430,0	€ 60.568,3	0,4872	8,27%	1,82%	-6,87%	11,17%	7,57%	1,82%	38,09	5,35	2,09%
Gruppo 4	€ 26.501,7 ±(€ 1.829,3)	€ 54.117,0 ±(€ 4.008,3)	0,4516 ±(0,0072)	6,11% ±(0,18%)	2,84% ±(0,30%)	3,54% ±(8,77%)	4,16% ±(2,63%)	2,55% ±(1,55%)	1,07% ±(0,29%)	59,51 ±(29,20)	4,43 ±(0,64)	2,07% ±(0,26%)
Gruppo 6	€ 20.224,2 ±(€ 1.812,8)	€ 49.870,0 ±(€ 1.672,6)	0,4579 ±(0,0291)	6,99% ±(0,79%)	4,03% ±(0,90%)	5,10% ±(6,02%)	1,23% ±(1,03%)	1,86% ±(2,79%)	0,71% ±(0,22%)	11,94 ±(13,60)	6,61 ±(0,73)	4,92% ±(1,50%)
Gruppo 7	€ 17.656,7 ±(€ 1.395,0)	€ 49.728,0 ±(€ 1.628,29)	0,4799 ±(0,0161)	7,44% ±(0,47%)	1,35% ±(0,62%)	-5,02% ±(5,51%)	0,70% ±(0,39%)	0,53% ±(0,19%)	0,80% ±(0,30%)	9,56 ±(4,91)	5,33 ±(0,70)	3,45% ±(1,27%)

Gruppo 1: Lombardia e Piemonte; Gruppo 2: Valle d'Aosta e Trentino Alto Adige; Gruppo 3: Veneto, Friuli Venezia Giulia e Emilia Romagna; Gruppo 4: Liguria, Toscana, Umbria e Marche; Gruppo 5: Lazio; Gruppo 6: Abruzzo, Molise e Basilicata; Gruppo 7: Campania, Puglia, Calabria, Sicilia e Sardegna.

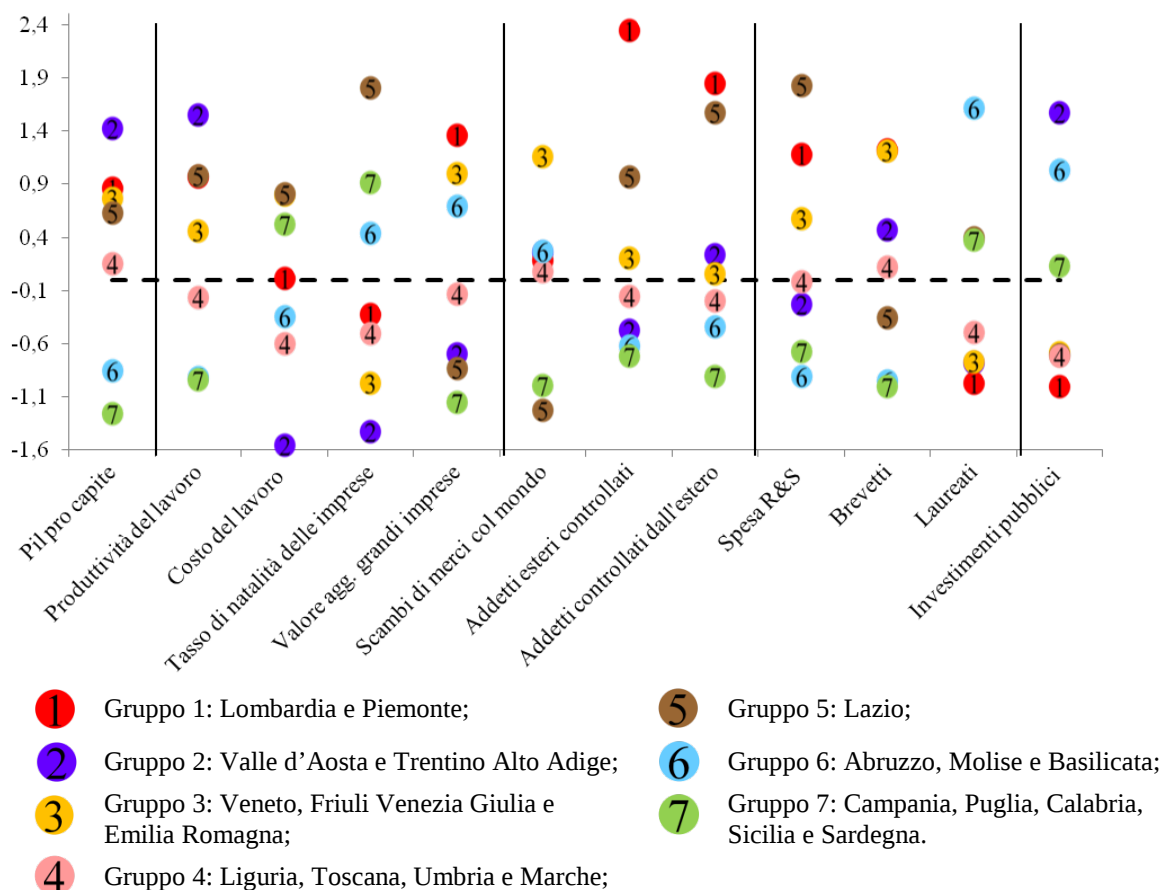
I valori esposti per ogni gruppo rappresentano il profilo competitivo delle regioni che comprende.

Continuiamo quindi a definire la competitività in modo poliedrico, ovvero come la combinazione di un insieme di aspetti che appartengono a quattro macro-aree: crescita reale, apertura internazionale, conoscenza e spesa pubblica. Inoltre la variabile Pil rimane l'espressione più generale della competitività del raggruppamento, in base ad essa possiamo quantificare il livello di competitività dei cluster mentre grazie alle altre variabili lo possiamo descrivere a livello qualitativo.

Riportiamo in seguito un altro grafico per facilitare il confronto tra gruppi per tutte le variabili considerate. Questo grafico ci servirà in seguito nella definizione del profilo competitivo dei singoli cluster in quanto ciò che si determina come “elevato” o “ridotto” dipende proprio dal confronto con gli altri valori.

Per poter confrontare tutte le variabili contemporaneamente occorre considerare i valori medi dei gruppi calcolati a partire non dalla matrice dei dati ma dalla matrice degli scarti standardizzati, in questo modo tutti i valori avranno la stessa unità di misura ed ordine di grandezza senza perdere la variazione che risulta tra loro. Il grafico che proponiamo riporta dunque i valori medi (per ogni variabile considerata) delle regioni raggruppate in cluster. Queste sono divise nelle quattro macro-aree più la variabile Pil, la retta tratteggiata indica lo zero e dunque il valore medio di tutte le regioni per ogni variabile (poiché i valori sono standardizzati, per definizione la loro media è 0 e la varianza è 1).

Fig. 4.5b – Valori medi per ogni cluster, calcolati su valori standardizzati



Attraverso questo grafico si perdono i valori assoluti registrati per ogni cluster, esposti nella Tab. 4.5a, ma si facilita il confronto: per ogni variabile è possibile tracciare non solo la graduatoria dei gruppi ma anche la distanza tra una posizione e l'altra.

I due grafici esposti in questo paragrafo rappresentano la base per delineare il profilo competitivo di ogni cluster identificato attraverso l'analisi dei gruppi. Nei prossimi paragrafi descriveremo le caratteristiche che accomunano i vari insiemi di regioni individuati, per ognuno si propone un grafico simile al precedente ma con soli due elementi in evidenza: i valori medi del cluster (calcolati su dati standardizzati) e la media dei valori di tutte le regioni (che per definizione è zero). In questo modo sarà possibile visualizzare allo stesso tempo tutte le caratteristiche peculiari di ogni gruppo, analizzate all'interno di ogni macro-area e discusse nell'ottica del valore registrato per la variabile Pil pro capite.

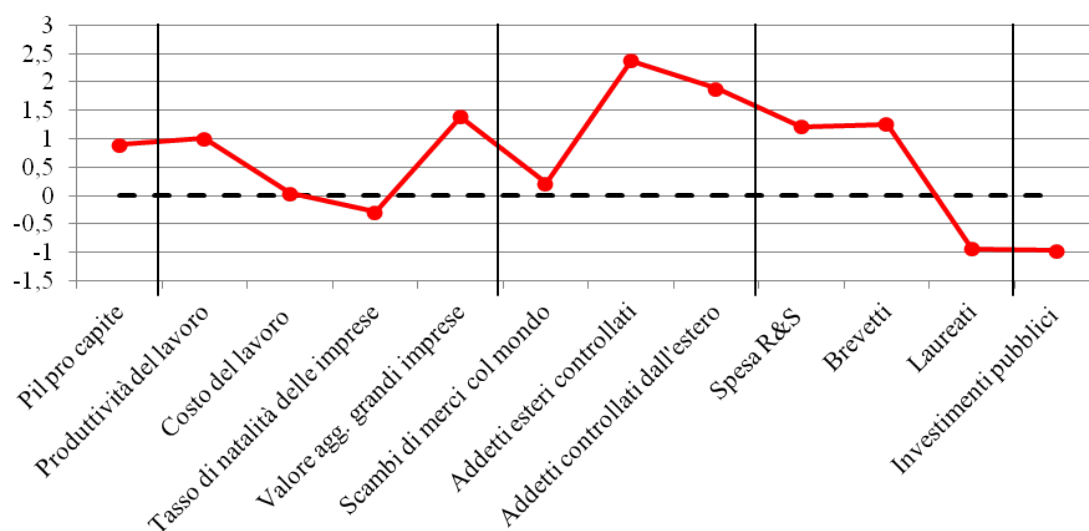
Ogni gruppo presenta, oltre al numero identificativo, anche una descrizione qualitativa delle sue caratteristiche principali. La prima è la fascia di appartenenza della variabile Pil pro capite, seguono poi quelli che sono i punti di forza del cluster in questione, ovvero quegli ambiti in cui registra i valori massimi rispetto a tutti gli altri gruppi.

4.5.1 Gruppo 1: reddito alto, grandi imprese, investimenti all'estero e brevetti



Il cluster 1 è composto dalle regioni della Lombardia e del Piemonte. Il Pil di questo gruppo (€ 30.880,3) rientra nella classe “Alto” insieme ai gruppi 3 e 5, tuttavia in questa classe la variabile Pil è quella che presenta la maggior deviazione standard (€3.681,9) poiché riunisce due regioni che, seppur molto simili per aspetti competitivi, non condividono il medesimo livello di manifestazione della competitività (la Lombardia ha un Pil di più di €33.000 e il Piemonte di €28.000). Dunque per quanto la combinazione competitiva sia la stessa, sembra che la Lombardia sia molto più in grado di collegarla ad una maggiore produzione interna.

Fig. 4.5.1 – Profilo competitivo di Lombardia e Piemonte



I dati riguardanti la crescita reale di questo gruppo sono caratterizzati da un elevato contributo delle grandi imprese in termini di valore aggiunto (circa 5%), il più alto di tutti i gruppi, e da un'alta produttività (€ 60.519), al secondo posto insieme al gruppo 5, a fronte

di un costo del lavoro e di un tasso di natalità delle imprese nella media (rispettivamente 0,4670 e 6,28%). Il valore della produttività apparente del lavoro risulta tuttavia molto variabile (circa €10.000 di differenza tra Lombardia e Piemonte), comunque la posizione in classifica della produttività e del costo del lavoro suggeriscono una buona efficienza del lavoro (ma non massima).

L'apertura internazionale risulta molto sviluppata in termini di Investimenti diretti esteri sia in entrata che in uscita, seppure in modo altalenante tra le due regioni (alte varianze). A fronte di esportazioni nette nella media (4,43% del Pil), il gruppo presenta infatti la maggior incidenza di addetti esteri controllati e di addetti italiani controllati dall'estero sul totale degli addetti rispetto a tutti gli altri cluster (rispettivamente 19,68% e 8,35%).

Il fattore conoscenza risulta anch'esso molto sviluppato ma in modo particolare: prevale l'incidenza dell'innovazione rispetto all'istruzione avanzata. I dati mostrano infatti come la percentuale della spesa in Ricerca e Sviluppo sul totale del Pil sia al secondo posto (1,56%, superata dal Lazio) mentre la domanda di brevetti per milione di abitanti sia addirittura al primo (108,22). In contrapposizione a queste brillanti classifiche, il numero di laureati per migliaia di abitanti è all'ultimo posto rispetto a tutti gli altri gruppi (3,94).

Anche l'incidenza degli investimenti pubblici sul totale del Pil registra il valore più basso di tutte le regioni (1,61%). Tale valore, relazionato con la classe alta della variabile Pil, può far ipotizzare una buona efficienza della spesa pubblica poiché a contenute spese (relative) ritorna un buon livello di reddito.

Questo è il profilo competitivo del cluster 1, buona efficienza del lavoro, grandi imprese, rapporti di investimento stabili con l'estero e una stabile ricerca nell'innovazione sono le sue chiavi per il successo che si manifesta maggiormente per la Lombardia.

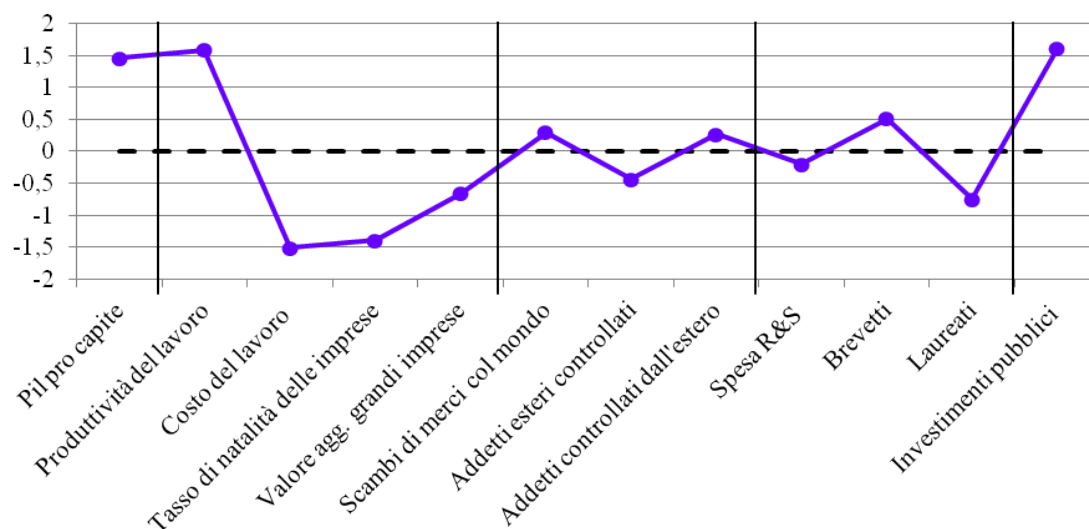
4.5.2 Gruppo 2: reddito altissimo, efficienza del lavoro e spesa pubblica



Il cluster 2 si compone delle regioni della Valle d'Aosta e del Trentino Alto Adige, esso presenta un profilo competitivo molto curioso.

Questo gruppo appartiene all'élite, registra il più alto valore di Pil pro capite con un buono stacco dagli altri gruppi. Tale valore si aggira intorno ai € 34.410 e rientra nella classe "Altissimo", tuttavia la causa sembra ritrovarsi in pochi aspetti del profilo competitivo del cluster 2 che risultano i più alti nelle classifiche con gli altri gruppi.

Fig. 4.5.2 – Profilo competitivo di Valle d'Aosta e Trentino Alto Adige



La prima spiegazione dell'altissima produzione interna è sicuramente un'ottima combinazione di produttività e costo del lavoro: in entrambe le variabili questo gruppo presenta la migliore postazione in classifica, ovvero la più alta produttività apparente (€ 63.799) e il più basso costo del lavoro (0,4277). A fronte di questo ottimo posizionamento nell'efficienza del lavoro, il gruppo registra la più bassa natalità delle imprese (5,25%) e un basso contributo sul valore aggiunto delle grandi imprese (2,02%).

Sia il livello di internazionalizzazione che quello della conoscenza si presentano nella media.

Nel primo caso si registrano sufficienti esportazioni nette con il mondo (5%) ed Investimenti diretti esteri soprattutto in entrata (3,76%, in uscita 2,19%) .

Nel caso della conoscenza ed innovazione, il gruppo presenta valori poco sotto la media per la spesa in R&S (quasi 1% del Pil, ma con una deviazione standard dello 0,46%), un numero di domande di brevetti abbastanza buono (75 per milione di abitanti, al terzo posto rispetto agli altri cluster) ed un basso livello di istruzione avanzata (4,13 laureati per migliaia di abitanti).

La variabile competitiva che spicca di più è l'incidenza della spesa pubblica sul totale del Pil, di ben 5,8%. A fronte della prima posizione in questo ambito, il gruppo presenta la più alta produzione interna che si può probabilmente spiegare in buona parte da un livello distinto di investimenti pubblici.

La massima produzione pro capite sembra basarsi sulla massima efficienza del lavoro ed incidenza degli investimenti pubblici sul Pil, questo è il profilo competitivo del cluster 2 che non presenta picchi nell'apertura verso l'estero o nella conoscenza e innovazione.

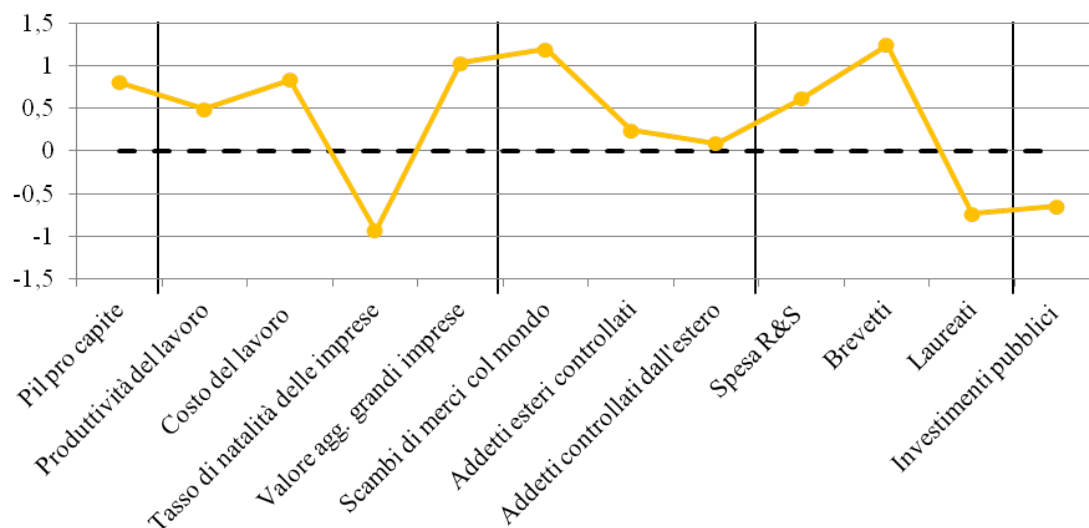
4.5.3 Gruppo 3: reddito alto, elevate esportazioni



Il cluster 3 comprende il Friuli Venezia Giulia, il Veneto e l'Emilia Romagna. La produzione media di questo gruppo (€ 30.324) si colloca nella fascia "Alta" insieme ai cluster 1 e 5, ovvero Lombardia, Piemonte e Lazio.

Per quanto il livello di Pil medio sia molto simile, questo gruppo presenta un suo particolare profilo competitivo che poco si rispecchia con quelli dei suoi pari.

Fig. 4.5.3 – Profilo competitivo di Friuli Venezia Giulia, Veneto ed Emilia Romagna



Uno degli aspetti di crescita reale che caratterizza queste tre regioni è un livello di efficienza del lavoro abbastanza vicino a quello dei cluster 1 e 5 ma che presenta la peggior combinazione dei due: a fronte di un costo del lavoro quasi pari a quello del Lazio (0,4869, il più alto in classifica) si registra una produttività apparente del lavoro molto più bassa degli altri due gruppi (€ 57.643). Poiché il livello di reddito medio appartiene alla stessa fascia, saranno dunque gli altri elementi competitivi a compensare questa mancanza.

Il contributo delle grandi imprese sul valore aggiunto (4,47%) risulta al secondo posto nella classifica con gli altri gruppi, è interessante collegare questo dato ad una bassa natalità delle imprese (5,68%, penultimo posto nella classifica) quasi a suggerire che la presenza di imprese di dimensioni maggiori rallenti l'entrata di nuove imprese.

Il grado di apertura internazionale evidenzia dati interessanti: per quanto i valori di Investimenti esteri diretti sia in entrata che in uscita risultino pressoché nella media (6,41% l'incidenza di addetti esteri controllati e 3,26% quella di addetti italiani a controllo estero sul totale degli addetti), questo gruppo presenta la più alta incidenza di esportazioni nette sul totale del Pil (12,26%), al primo posto nella classifica con un alto stacco da tutti gli altri cluster.

Il livello di conoscenza e la ricerca dell'innovazione si presentano molto simili a quelli del cluster 1. L'incidenza della spesa in Ricerca e Sviluppo è in terza posizione con l'1,3%, le

domande di brevetti europei sono pressoché pari a quelle di Lombardia e Piemonte (107,7 per milione di abitanti con una alta deviazione standard di 34,73) come anche il livello di istruzione avanzata (4,15 laureati per migliaia di abitanti).

L'incidenza della spesa pubblica sul Pil si pone intorno al 2%, un livello al di sotto della media (3%) che si ritrova anche nei cluster 4 e 5.

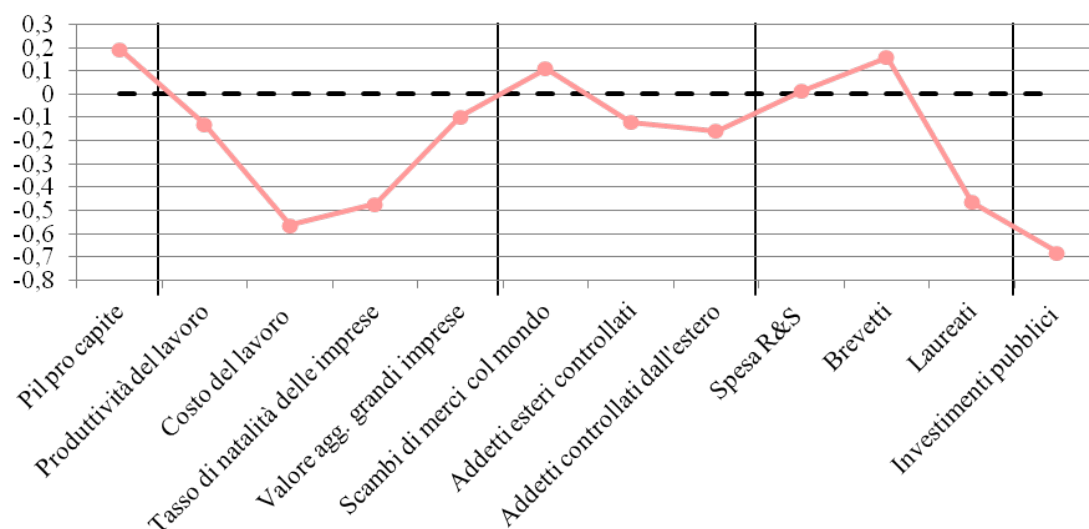
Il cluster 3 è il secondo nella fascia “Alta” in base al Pil, le sue qualità competitive si basano sul contributo delle grandi imprese, su alte esportazioni all'estero e su un buon livello di ricerca dell'innovazione.

4.5.4 Gruppo 4: reddito medio, profilo medio



Il cluster 4 comprende Liguria, Toscana, Umbria e Marche. Il livello di Pil di queste regioni (€ 26.501) è l'unico ad appartenere alla fascia “Medio”, infatti è molto vicino al valore medio registrato per l'insieme delle regioni italiane (€25.330). Ciò che caratterizza questo gruppo è la sua vicinanza ai valori medi in quasi tutte le variabili considerate, per questo possiamo definirlo un cluster “medio” a tutti gli effetti.

Fig. 4.5.4 – Profilo competitivo di Liguria, Toscana, Umbria e Marche



Come si può notare dal grafico, gli scostamenti dei valori standardizzati dalla media sono dell'ordine di grandezza di pochi decimali.

Il profilo della crescita reale è caratterizzato da una buona efficienza del lavoro. La produttività apparente del lavoro è poco sotto la media (€ 54.117 con una alta deviazione standard di €4.008) mentre il costo del lavoro è il più basso dopo il cluster 2 (0,4516). La percentuale di natalità delle imprese (6,11%) è anch'essa poco più bassa della media di tutte le regioni, mentre l'apporto delle grandi imprese al valore aggiunto complessivo quasi coincide con il valore medio (2,84%).

Il livello di internazionalizzazione presenta i seguenti valori: esportazioni nette del 3,5% sul totale del Pil (con una alta deviazione standard del 8,77%), incidenza di addetti esteri controllati del 4,16% e degli addetti controllati del 2,5% sul totale degli addetti. Tutti questi valori ricalcano il livello di apertura medio delle regioni italiane.

Riguardo gli indicatori della conoscenza, l'unico valore che si discosta molto dalla media è quello dell'istruzione avanzata. Secondo i dati del gruppo si registrano 4,4 laureati ogni mille abitanti. La spesa in R&S ammonta al 1,07% del Pil mentre le domande di brevetti sono 59,5 ogni milione di abitanti.

L'incidenza degli investimenti pubblici sul Pil risulta molto vicina a quella registrata per i cluster 3 e 5, ovvero il 2,07%, la fascia bassa ma non bassissima come quella del cluster 1. Questo è il cluster 4, a medi aspetti competitivi medie prestazioni. Nel suo basso profilo competitivo le mancanze che spiccano di più sono il tasso di laureati e l'incidenza della spesa pubblica sul Pil.

4.5.5 Gruppo 5: reddito alto, alta natalità delle imprese e spesa in R&S

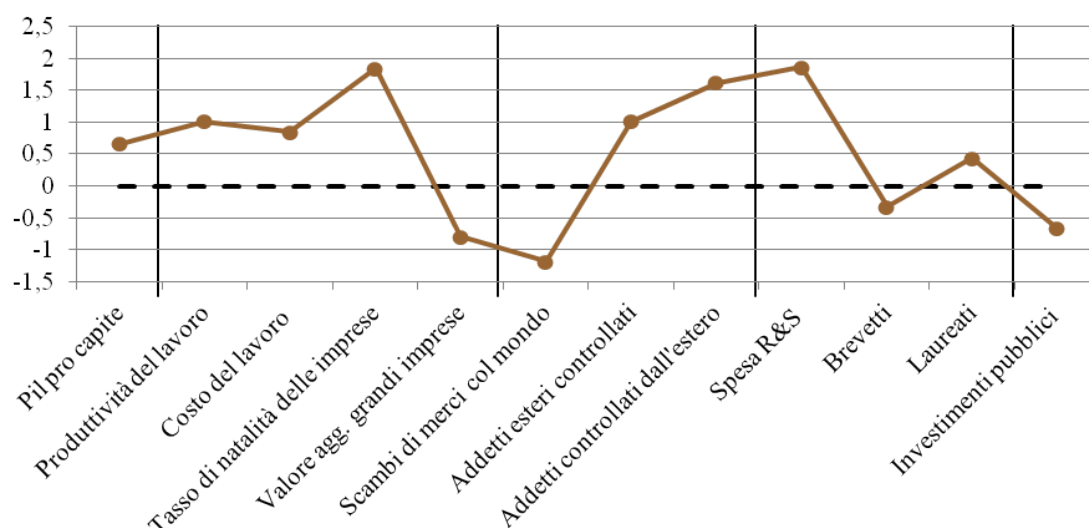


Il cluster 5 non si può propriamente chiamare così in quanto comprende una sola regione, il Lazio. I valori riportati in seguito non sono dunque valori medi, di cui si può riportare la deviazione standard per le variabili in cui risulta maggiore, ma semplicemente i valori registrati per questa regione. Il motivo per cui la *cluster analysis* elegge questa regione come gruppo a regione singola, il primo a crearsi (a partire dal raggruppamento a 6 cluster), è proprio

la particolare combinazione competitiva che caratterizza questa regione. Probabilmente possiamo far risalire questa particolare configurazione alla situazione politica e amministrativa specifica di questa regione che ospita non solo la capitale della nazione ma anche uno stato a sé stante.

Il Pil del Lazio rientra nella stessa fascia alta dei cluster 1 e 3, seppure si presenta come il più basso rispetto ai valori medi degli altri gruppi (ammonta a € 29.430).

Fig. 4.5.5 – Profilo competitivo del Lazio



La dinamica della crescita reale si presenta poco sostenuta: per quanto la produttività del lavoro è molto vicina a quella del cluster 1 (€ 60.568) e tra le più alte, il costo del lavoro è tuttavia il più alto (0,4872, poco più del valore registrato per il cluster 3). La natalità delle imprese, relazionata alla dimensione di queste, presenta una situazione opposta a quella del cluster 3: altissima natalità delle imprese (8,27%, primo posto in classifica) ma ridotta

dimensione media (il contributo al valore aggiunto da parte delle grandi imprese è solo del 1,82%, penultimo posto in classifica).

L'apertura verso l'estero risulta molto sviluppata: l'incidenza degli Investimenti diretti esteri, sia in entrata che in uscita, sul Pil è molto al di sopra della media (rispettivamente 7,57% e 11,17% in termini di addetti controllati e controllanti) nonché al secondo posto (in entrambi i casi) nella classifica dei cluster. Gli scambi di merci avvengono invece prevalentemente in entrata (le importazioni sono il -6,87% del Pil) ponendo il Lazio all'ultimo posto della classifica per esportazioni nette come percentuale del Pil.

Per quanto riguarda il livello di conoscenza, si registrano valori intorno alla media per domande di brevetti (38,09 per milione di abitanti) e numero di laureati (5,35 ogni mille abitanti), tuttavia il Lazio registra la maggior incidenza di spesa in Ricerca & Sviluppo sul totale del Pil (1,82%) ponendosi in cima alla classifica dei cluster per questo indicatore di innovazione.

L'ammontare di spesa pubblica come percentuale del Pil (2,09%) è molto al di sotto del valore medio, nella graduatoria dei cluster il Lazio è accompagnato dai gruppi 4 e 3 che registrano circa i medesimi valori.

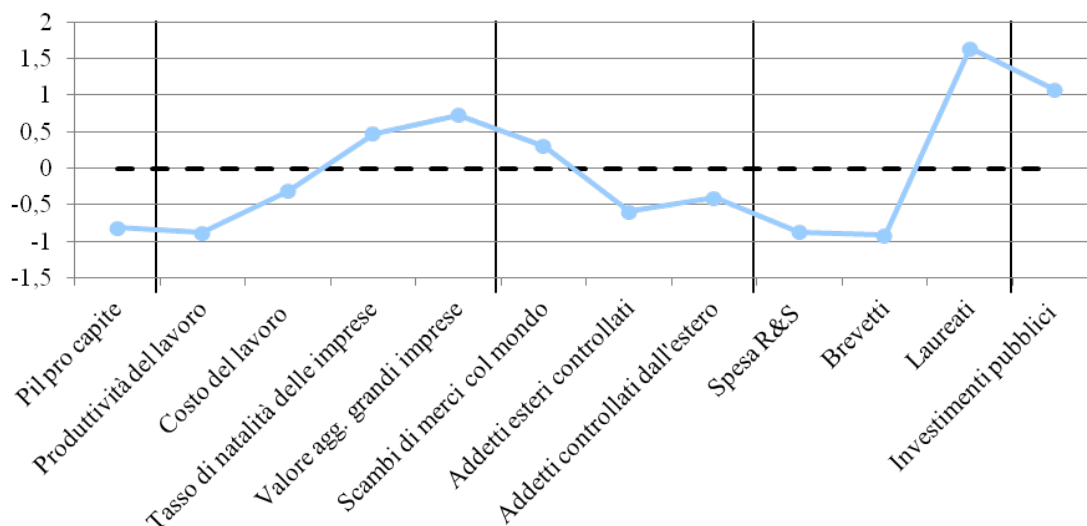
Nella sua particolarità il Lazio forma un cluster a sé, la sua produzione interna di fascia alta si può spiegare tramite un profilo competitivo caratterizzato da un'alta natalità delle imprese, un'ampia e stabile apertura verso l'estero e un'elevata spesa per l'innovazione.

4.5.6 Gruppo 6: reddito basso, numerosi laureati



Il cluster 6 comprende le regioni dell'Abruzzo, Molise e Basilicata. Il Pil pro capite medio di queste regioni ammonta a € 20.224 che, a confronto con i valori registrati dagli altri cluster, definiamo di classe "Basso". Questo gruppo e il seguente sono infatti quelli che dimostrano una performance al di sotto della media delle regioni italiane e di conseguenza sono caratterizzati da profili competitivi lacunosi.

Fig. 4.5.6 – Profilo competitivo di Abruzzo, Molise e Basilicata



Gli aspetti di crescita reale di questo gruppo indicano una efficienza del lavoro poco sostenuta. Infatti la produttività apparente del lavoro (€ 49.870) risulta molto al di sotto della media (al pari di quella del cluster 7) mentre il costo del lavoro (0,4579) non si discosta eccessivamente da questa. Vi sono poi delle potenzialità non sfruttate, quali un buon tasso di natalità delle imprese (6,99%) e un buon contributo delle grandi imprese sul valore aggiunto totale (4,03%), entrambi i valori sono al di sopra della media e al terzo posto nella classifica dei cluster. All'interno del cluster, le distribuzioni sono tuttavia poco omogenee in quanto si registrano per questi ultimi due valori delle deviazioni standard di, rispettivamente, 0,79% e 0,90%, molto alte per tali variabili.

Riguardo il grado di internazionalizzazione, sembrano prevalere i flussi di commercio anziché i flussi di investimenti a lungo termine. Mentre l'incidenza delle esportazioni nette sul Pil risulta sopra la media, seppur di poco, con il 5,1%, l'ammontare di Investimenti diretti esteri in entrata e uscita sul totale del Pil (1,86% e 1,23%) è molto più basso della media e, in entrambi i casi, in penultima posizione nelle classifiche dei cluster.

Lo sviluppo della conoscenza manifesta una particolare configurazione: massimo investimento nell'istruzione universitaria ma, allo stesso tempo, minima spesa per la ricerca nell'innovazione. Il tasso di laureati è di ben 6,6 per migliaia di abitanti (primo posto nella classifica dei gruppi) mentre la percentuale di R&S sul Pil e le domande di brevetti per milione di abitanti (rispettivamente di 0,71% e 11,94) toccano le ultime posizioni nelle classifiche dei cluster. Tuttavia anche nel caso dei laureati si registra un'alta deviazione standard (0,73) a dimostrare un'alta variabilità nel gruppo.

Infine, per quanto riguarda gli investimenti pubblici, il cluster 6 si pone al secondo posto in classifica con il 4,9% come incidenza della spesa pubblica sul Pil, altro valore marcato da una forte deviazione standard (1,5%), sempre rispetto alle varianze della stessa variabile.

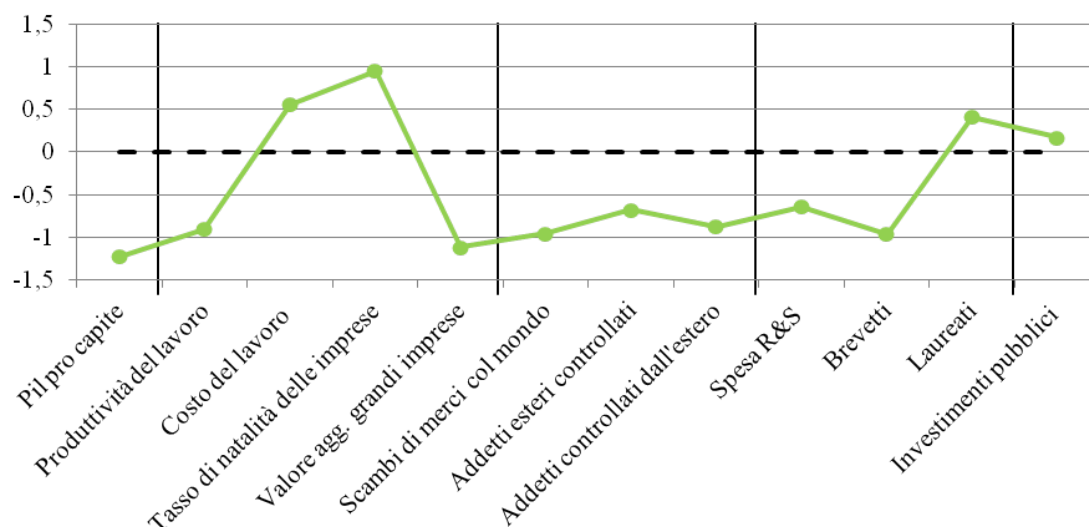
Ciò che definisce il profilo competitivo di questo cluster di reddito "Basso" sono una bassa efficienza del lavoro, ridotte relazioni con l'estero di lungo periodo e una bassa spesa nell'innovazione. I punti di forza che potrebbero essere invece sfruttati per rilanciare il sistema produttivo sono il tasso di imprese di grandi dimensioni, l'elevato numero di laureati e l'ammontare della spesa pubblica. Di questi aspetti positivi bisogna tuttavia tener conto della variabilità tra le regioni del cluster che risulta infatti la più alta nell'ambito di queste variabili.

4.5.7 Gruppo 7: reddito bassissimo, profonde lacune



Il cluster 7 comprende le regioni della Campania, Puglia, Calabria, Sicilia e Sardegna. La fascia di Pil pro capite di questo insieme di regioni è la più bassa in assoluto. Il valore medio di € 17.656 definisce la classe "Bassissimo". Come vedremo dal profilo competitivo di questo gruppo, la causa si può ritrovare nei valori al di sotto della media per quasi tutte le variabili considerate.

Fig. 4.5.7 – Profilo competitivo di Campania, Puglia, Calabria, Sicilia e Sardegna



La situazione della crescita reale si presenta come la peggiore: la produttività apparente del lavoro è la più bassa di tutte (€ 49.728) mentre il costo del lavoro è quasi il più alto di tutti (0,4799), una situazione opposta a quella più favorevole del cluster 2. Anche il contributo delle grandi imprese sul valore aggiunto si presenta come il più ridotto nella classifica (1,35%), accompagnato da un'alta natalità delle imprese (7,44%) quasi a confermare che queste non abbiano tempo per ingrandirsi ma siano subito sostituite da nuovi arrivati già agli inizi dei loro *businesses*.

L'apertura internazionale presenta il profilo meno esteso di tutti i gruppi: gli scambi con il mondo sono prevalentemente importazioni (-5,02%, esportazioni nette come percentuale del Pil), gli Investimenti diretti esteri sia in entrata che in uscita (rispettivamente 0,53% e 0,70% in termini di addetti controllati e controllanti) sono notevolmente al di sotto della media nonché all'ultimo posto delle classifiche dei cluster.

Anche il livello di conoscenza presenta profonde lacune. Sia l'incidenza di spesa in R&S sul Pil che il numero di domande di brevetti per milione di abitanti (di 0,80% e 9,56) si presentano molto al di sotto dei valori medi italiani e agli ultimi posti in classifica. Solo il numero medio di laureati per migliaia di abitanti sembra sollevarsi di poco dalla media con un valore di 5,33 seppur seguito da una deviazione standard di 0,7 molto alta per questa variabile, a dimostrazione di situazioni eterogenee tra regioni.

L'incidenza della spesa pubblica sul Pil risulta invece nella media con un valore del 3,45%, forse troppo basso per sostenere sufficientemente il sistema produttivo come avviene nel cluster 6.

Il gruppo 7 presenta sicuramente il profilo meno competitivo e più problematico, marcato da profondi *gap* in termini di efficienza del lavoro, dimensioni delle imprese, relazioni stabili con l'estero e ricerca nell'innovazione e nella conoscenza.

4.6 Conclusioni

Le possibilità di crescita dell'Italia, in particolar modo all'estero, evidenziano l'importanza di un'analisi competitiva del suo sistema produttivo nonché delle sue potenzialità e limiti all'espansione economica.

Conducendo questo tipo di ricerca sia a livello nazionale che a livello regionale, l'amministrazione del Paese può disporre di un insieme di indicatori sul quale basare una strategia di intervento mirata ed efficace.

Le lacune strutturali della nostra nazione sembrano interessare gli investimenti nelle tecnologie di innovazione e nel capitale umano, il sistema di protezione dei gruppi di lavoratori eccessivamente frenante per le imprese, la rete dei trasporti troppo legata alla modalità stradale ed eccessivamente frammentata, i tempi di risoluzione dei processi civili elevati e l'alta incidenza dell'economia sommersa sul totale del Pil.

Studiando un maggiore livello di dettaglio territoriale, le regioni italiane sono state valutate in termini di crescita reale, apertura internazionale, conoscenza e spesa pubblica. L'analisi dei gruppi si è basata su una combinazione di strumenti statistici che ha individuato 7 cluster nei quali sono state classificate le venti regioni. L'insieme di questi gruppi ha definito 5 fasce di reddito in base ai valori medi nei cluster della variabile Pil pro capite.

Nella fascia di reddito "Altissimo" rientrano le regioni del Trentino Alto Adige e Valle d'Aosta (cluster 2).

Queste sono le regioni che sembrano sfruttare al meglio il loro potenziale e la loro maturità: presentano infatti un'alta efficienza del fattore Lavoro, ovvero alta produttività e basso costo, e la quota più elevata di spesa pubblica. Dunque aprendo maggiormente questo cluster all'estero si può ipotizzare un'espansione delle possibilità di crescita giustificate da un'ottima efficienza del sistema produttivo.

Nella fascia "Alto", invece, rientrano tre gruppi di regioni che per quanto manifestano diverse combinazioni di elevata apertura all'estero, non riescono a sfruttare questo potenziale per espandere il proprio sistema produttivo e raggiungere la fascia di reddito più alta. Il cluster 1 (Piemonte e Lombardia) ad esempio, registra il più alto livello di Investimenti diretti esteri sia in entrata che in uscita di tutti i gruppi regionali, eppure non riesce a crescere verso l'estero utilizzando i suoi punti di forza quali la presenza di grandi imprese (250 addetti e oltre) e l'alta richiesta di brevetti europei. Il cluster 3 (Friuli Venezia Giulia, Veneto ed Emilia Romagna) presenta il più elevato saldo commerciale con il mondo ma, evidentemente, un profilo competitivo ancora da migliorare per sfruttare questo potenziale. Infine il cluster 5 (Lazio), che presenta un buon livello di Investimenti diretti esteri (il secondo dopo il cluster 1), potrebbe migliorare la sua performance sfruttando punti di forza quali gli alti investimenti in R&S e l'elevata natalità delle imprese.

Nella fascia "Medio" rientra solo il cluster 4, che comprende Liguria, Toscana, Umbria e Marche. Queste regioni sembrano essere piuttosto coerenti: registrano un reddito medio a fronte di un profilo competitivo nella media in quasi tutti gli aspetti economici.

Le regioni che più necessitano di un intervento sono quelle della fascia di reddito "Basso", ovvero Abruzzo, Molise e Basilicata (cluster 6), e soprattutto quelle di reddito "Bassissimo": Campania, Puglia, Calabria, Sicilia e Sardegna (cluster 7). Mentre le regioni del sesto gruppo presentano un punto di forza non sfruttato, la più alta frequenza di laureati, quelle dell'ultimo gruppo presentano quasi tutti i valori nettamente al di sotto della media.

Questa è la situazione produttiva italiana, caratterizzata da *gap* competitivi a livello nazionale ma anche e soprattutto da realtà estremamente variegata poichè nascondono problematiche nettamente diversificate.

Appendice

Istruzioni per l'utilizzo di R

Caricare i dati

Dopo aver costruito la matrice dei dati su un foglio elettronico come, ad esempio, *Excel*, si può salvare il lavoro in formato testo (.txt). Affinché R possa leggerlo correttamente occorre porre tra virgolette tutti i nomi delle unità (se lo possiedono) e tutti i nomi delle variabili, infine occorre sostituire tutte le virgole con i punti (secondo il sistema anglosassone dei numeri decimali). A questo punto, dopo aver impostato come *directory* la cartella in cui si trova il file di testo, si può digitare:

```
dati <- read.table("Matrice.dati.txt", header = TRUE)
```

In questo modo la matrice dei dati viene assegnata ad un oggetto chiamato “dati”, *header = TRUE* va specificato nel caso in cui la prima riga del file di testo sia composta dai nomi delle variabili.

Standardizzare la matrice

Con la seguente scrittura si assegna all’oggetto “standard” la standardizzazione della matrice “dati” precedentemente inserita.

```
standard <- scale(dati)
```

Cluster analysis gerarchica

Prima si calcola la matrice delle distanze con la seguente scrittura:

```
distanze <- dist(standard)
```

In questo modo si assegna ad un oggetto “distanze” la matrice delle distanze euclidee calcolata a partire dalla matrice degli scarti standardizzati precedentemente calcolata. Se si vuole utilizzare un altro tipo di distanza occorre scrivere *<- dist(standard, method = “x”)* dove *x* può essere, ad esempio, *manhattan*, *canberra*, *maximum*, *binary*.

L’analisi gerarchica si effettua scrivendo:

```
hc <- hclust(distanze, method = "average")
```

Si assegna così ad un oggetto “hc” una lista di elementi esplicativi della cluster analysis effettuata sulla matrice delle distanze “distanze” precedentemente calcolata. Altri metodi di raggruppamento sono *single*, *complete*, *ward*, *centroid*. Gli elementi esplicativi si possono richiamare con le seguenti scritture:

plot(hc) → espone il dendrogramma;

hc\$merge → espone il processo di agglomerazione, ovvero le unità che vengono aggregate per ogni fase (i numeri negativi sono le singole unità, i numeri positivi sono cluster che prendono il nome numerico dalla fase in cui sono stati creati aggregano unità e/o gruppi);

hc\$height → espone il livello di altezza al quale è avvenuta l'aggregazione per ogni fase del processo;

id <- cutree(hc, k = x, h = y) → assegna ad un oggetto "id" una vettore di allocazione delle unità, ovvero l'elenco delle unità su cui si è effettuata la cluster analysis con accanto ad ognuna di esse il numero del cluster alla quale appartiene. Il taglio del dendrogramma (cut-tree) si può stabilire come un valore della distanza (y), ovvero l'altezza alla quale il dendrogramma viene tagliato, oppure come un numero di classi (x) per la minor altezza possibile (scrivendo il valore di k si può omettere la h e viceversa).

Cluster analysis non gerarchica con il metodo delle k-medie

Il comando per effettuare la cluster analysis con il metodo delle k-medie è il seguente:

km <- kmeans(standard, centers)

Questo andrà ad assegnare ad un oggetto "km" una lista di elementi esplicativi del raggruppamento effettuato sulla matrice dei dati standardizzati "standard" precedentemente calcolata. L'input *centers* può avere due significati: il numero di gruppi da formare (compreso tra 1 e n, date n unità nella matrice dei dati) oppure i poli iniziali, che indicano non solo il numero di gruppi da formare ma anche i centri iniziali di aggregazione.

Nel primo caso, ovvero scrivendo al posto di *centers* un numero, i poli iniziali verranno scelti casualmente dal programma. È quindi da ricordare che ogni esecuzione dell'analisi con questo sistema non produrrà sempre lo stesso risultato. Per vedere la partizione ottenuta si possono richiamare gli elementi esplicativi con le seguenti scritture:

km\$cluster → espone il vettore di allocazione delle unità, ovvero l'elenco delle unità su cui si è effettuata la cluster analysis con accanto ad ognuna di esse il numero del cluster alla quale appartiene;

km\$center → espone la matrice dei centroidi, ovvero l'elenco dei *k* vettori (uno per ogni *cluster*) ognuno costituito dalla media di tutte le unità del suo *cluster* per ogni variabile;

km\$within → espone per ogni gruppo la devianza entro i gruppi.

Se invece si vogliono indicare i poli iniziali la procedura è più complessa. Nel nostro caso questi sono stati determinati a partire da un metodo gerarchico applicato alla medesima matrice dei dati: tagliando il dendrogramma prodotto dal metodo gerarchico alla minima altezza per formare *k* raggruppamenti, i poli iniziali possono essere i *k* valori medi dei *k* gruppi formati. Per tradurre su R questa procedura basta scrivere al posto di "centers" l'oggetto "initial" definito nel seguente modo:

initial <- tapply(standard, index, mean)

La funzione *tapply* applicherà ad ogni cella della matrice "standard", raggruppata secondo un indice "index", una funzione che in questo caso è "mean", ovvero la media aritmetica.

L'oggetto "index" si deve definire prima con il seguente comando:

index <- list(rep(cutree(hc,k),ncol(standard)),col(standard))

"Index" si costituisce di due oggetti (funzione *rep* e funzione *col*) resi un indice dalla funzione *list*.

La funzione *rep(a, b)* ripete gli elementi di un vettore *a* per *b* volte, in questo caso *a* è un vettore di allocazione dato dall'output di un metodo gerarchico (*hc*) tagliato con la

funzione *cutree* per avere k gruppi (alla minima altezza possibile). Tale oggetto viene replicato per un numero di volte pari al numero delle colonne della matrice “standard”. La funzione *col* darà come output una matrice, delle dimensioni della matrice “standard”, in cui in ogni colonna si ripete un unico valore che è il numero della colonna stessa. Indubbiamente l’oggetto “index” è molto complesso e poco intuitivo, ma “initial” è una matrice dei centroidi facilmente riconoscibile.

Calcolo della Devianza Totale

Partendo da una matrice di dati standardizzati, il calcolo della Devianza totale (T) risulta molto facile:

sum(standard^2)

Il risultato di questa scrittura è la somma di tutte le celle della matrice standardizzata “standard” ognuna di esse elevata al quadrato, è sufficiente questa operazione in quanto la media di tutte le variabili è zero e dunque lo scarto di ogni valore dalla media sarà sempre il valore stesso.

Calcolo della Devianza nei gruppi

Per calcolare la Devianza W, se abbiamo effettuato il metodo delle k-medie, sarà sufficiente scrivere:

sum(km\$within)

Il risultato è la somma di tutte le devianze entro i gruppi formati nel raggruppamento con il metodo delle k-medie chiamato “km”. Si ricorda tuttavia che richiamando l’oggetto “km” appariranno una serie di elementi esplicativi tra i quali anche il valore dell’indice R^2 .

Se invece abbiamo effettuato un metodo gerarchico, dobbiamo usare un sistema un po’ più complesso:

v <- aggregate(standard, list(id), var)[-1]

La funzione *aggregate* calcola per ogni gruppo, a partire da una matrice “standard” raggruppata secondo un vettore di allocazione delle unità “id” (indicizzato tramite la funzione *list*), la funzione scritta alla fine del comando, in questo caso *var* sta per varianza. L’output della scrittura è dunque una matrice (chiamata “v”) formata da tanti vettori quanti gruppi, ogni vettore si compone della somma (calcolata per ogni variabile, ovvero per ogni spazio del vettore) delle varianze tra ogni elemento nel gruppo e la media di tutti gli elementi della matrice “standard”. La scrittura *[-1]* serve per eliminare la prima colonna della matrice ottenuta con la funzione *aggregate*, in caso contrario si andrebbero a considerare i numeri indicativi dei gruppi come varianze di una immaginaria variabile iniziale.

A questo punto, per la definizione di varianza campionaria, si devono sommare tutti i numeri che compongono ogni vettore delle varianze e moltiplicare ogni somma per il numero di elementi che si trova in quel gruppo meno una unità. Il numero di elementi in ogni gruppo si può conoscere con il seguente comando:

table(id)

Essa espone la numerosità di ogni gruppo nel raggruppamento rappresentato dall'oggetto "id", ovvero un dendrogramma tagliato ad una data altezza.

La somma finale per ottenere W (ipotizzando k = 4 gruppi ognuno di numerosità 5) si può calcolare scrivendo:

$$\text{sum}(v[1,])*4 + \text{sum}(v[2,])*4 + \text{sum}(v[3,])*4 + \text{sum}(v[4,])*4$$

Dove $v[1,]$ è il vettore delle varianze del gruppo uno.

Bibliografia

Zani S., Cerioli A., (2000), *Analisi dei dati e data mining per le decisioni aziendali*, Giuffrè Editore

Istat, (2013), *Rapporto sulla competitività dei settori produttivi*

Istat, (2012), *Rapporto annuale 2012*

A.T. Kearney, Visa, Friedrich Schneider, (2013), *The Shadow Economy in Europe 2013*

Proietti T., *Analisi di mercato*

Ferrante M.R., *Unità 2 - Corso di Statistica aziendale*

Massari R., *Analisi dei dati con R*

Massari R., *Introduzione all'analisi statistica con R*

Siti Internet consultati

- <http://www.istat.it/it/>
<http://www.istat.it/it/archivio/75111>
<http://www.istat.it/it/archivio/16777>
- <http://www.unioncamere.gov.it/>
http://www.starnet.unioncamere.it/Appendice-statistica-al-Rapporto-2012_1A8055
- <http://www.ice.gov.it/>
http://actea.ice.it/short_stat.aspx
<http://actea.ice.it/ide.aspx>
- <http://www.r-project.org/>