

## **LO STOCK IMMOBILIARE NELLE REGIONI ITALIANE: ANALISI DELLA DISTRIBUZIONE E DEI CLUSTER**

Maurizio FESTA<sup>1</sup>, Alessandra STORNILO<sup>2</sup>, Gregorio VENDITTI<sup>3</sup>

### **SOMMARIO**

Questo studio propone un'analisi delle diversità e delle similitudini tra le regioni italiane utilizzando le informazioni relative allo stock immobiliare edificato nei diversi ambiti territoriali. Le informazioni sulla numerosità e sulla classificazione degli immobili residenti negli archivi del catasto italiano, gestito dall'Agenzia del Territorio, consentono di caratterizzare ogni regione attraverso la destinazione economica dei suoi immobili. L'opportuno esame dello stock immobiliare censito nelle regioni italiane ha consentito, infatti, di individuare quelle variabili capaci di mostrare, dopo un adeguato processo di elaborazione, similitudini e diversità di tali ambiti territoriali. Si è scelto a tal fine di procedere con un'elaborazione utilizzando la tecnica della cluster analysis. È stata svolta un'elaborazione dei dati regionali relativi alle numerosità dello stock immobiliare delle diverse tipologie immobiliari, rapportata alla popolazione residente, con l'obiettivo di mostrare le relazioni tra le diverse regioni che i dati censuari, demografici ed economici del territorio evidenziano. I risultati dell'analisi così condotta consentono di raggiungere un obiettivo conoscitivo nuovo, fotografando il territorio nazionale dal punto di vista della "densità" immobiliare delle sue regioni. Questa fotografia dell'Italia potrà supportare politiche di sviluppo e di governo del territorio, apportando il suo contributo al miglioramento della conoscenza delle diversità regionali.

---

<sup>1</sup> Agenzia del Territorio, Largo Leopardi, 5, 00185, Roma e-mail: maurizio.festa@agenziaterritorio.it

<sup>2</sup> Agenzia del Territorio, Largo Leopardi, 5, 00185, Roma e-mail: alessandra.stornio@agenziaterritorio.it

<sup>3</sup> Agenzia del Territorio, Largo Leopardi, 5, 00185, Roma, e-mail: gregorio.venditti@agenziaterritorio.it

## Introduzione

L'Italia possiede un vasto patrimonio immobiliare che riflette il suo passato di civiltà urbana e contemporaneamente rispecchia le diverse tradizioni rurali delle regioni. Tale patrimonio fisico costituisce, a sua volta, il patrimonio di dati che la banca dati dell'Agenzia del Territorio possiede in termini di stock ovvero di popolazione di immobili e terreni del loro uso, scambio e valore su tutto il territorio.

Recentemente alcuni studi hanno messo in risalto il primato delle famiglie italiane in termini di ricchezza patrimoniale di cui parte fondamentale è il patrimonio immobiliare. Il nostro paese si trova ai primi posti nel mondo per percentuale di famiglie proprietarie dell'alloggio di residenza.

Ciò riflette anche il grande ventaglio di tipologie edilizie presenti sul territorio che offrono un'offerta molto articolata sia per valore e pregio, che per adattamento agli stili di vita ed alle attività economiche dei proprietari. L'analisi delle tipologie abitative presenti al catasto ed un confronto effettuato con quelle commerciali e produttive mette in luce anche lo stretto rapporto tra residenza ed occupazione.

Obiettivo di questo lavoro è l'esame della distribuzione dello stock nelle regioni italiane al fine di mettere in risalto le disparità regionali. Le fonti utilizzate sono gli archivi della banca dati catastale dell'Agenzia del Territorio. E' elaborato lo *stock* dei fabbricati, così come censito nella banca dati del catasto urbano aggiornato al 31 dicembre 2009. Si tratta di informazioni che riguardano un totale di oltre 65 milioni di unità immobiliari aggregate in diverse tipologie edilizie secondo la loro destinazione d'uso.

Il contributo proposto affronta il tema dell'analisi a livello regionale delle tipologie immobiliari presenti, residenziali e non, per lo studio e l'interpretazione dei fenomeni territoriali delle diverse realtà regionali, considerando anche i livelli di valore delle abitazioni. In particolare, sono utilizzati metodi di rappresentazione cartografica e strumenti di statistica descrittiva. I dati analizzati evidenziano le disparità regionali in termini di tipologie edilizie residenziali, di tipologie terziarie, commerciali e produttive. Si tratta, in questo caso, di un'elaborazione dei dati regionali relativi alle numerosità dello stock immobiliare delle diverse tipologie immobiliari, rapportata alla popolazione residente, con l'obiettivo di mostrare le relazioni tra le diverse regioni che i dati censuari, demografici ed economici del territorio evidenziano.

L'analisi così condotta permette di raggiungere un obiettivo conoscitivo nuovo, fotografando il territorio nazionale dal punto di vista della "densità" immobiliare delle sue regioni. Questa fotografia dell'Italia potrà supportare politiche di sviluppo e di governo del territorio, apportando il suo contributo al miglioramento della conoscenza delle diversità regionali.

## **1 Presentazione della base dati**

La base dati è stata costituita attraverso l'elaborazione di due diverse fonti informative:

- la banca dati del Catasto Edilizio Urbano;
- la banca dati delle quotazioni dell'Osservatorio del Mercato Immobiliare;

La banca dati del Catasto Edilizio Urbano è gestita dall'Agenzia del Territorio. Essa contiene l'inventario dei beni immobili edificati (unità immobiliari urbane) dell'intero territorio nazionale, ad esclusione delle province di Trento e Bolzano, le quali gestiscono autonomamente il Catasto dei fabbricati secondo modalità, regole e metodologie del tutto analoghe a quelle del catasto nazionale. Per questa occasione l'inventario dei beni immobili edificati, rilevati al 31/12/2009, è stato comunque integrato anche con le informazioni relative alle proprietà immobiliari ubicate nei territori delle province di Trento e Bolzano, grazie alla disponibilità delle province autonome. Si tratta di informazioni che riguardano un totale di oltre 65 milioni di beni fra unità immobiliari urbane ed altre tipologie immobiliari che non producono reddito e delle quali si forniscono: la numerosità dello stock, la sua consistenza fisica («vani», superfici o volumi a secondo delle categorie tipologiche) e la correlata base imponibile fiscale determinata dal Catasto (la «rendita catastale»). Lo stock elaborato in questo lavoro comprende le unità censite nelle categorie catastali dei gruppi A, B, C, D che sono state ritenute più rappresentative per le finalità di questo studio.

Le unità nelle categorie del gruppo E ed F non sono state prese in considerazione in quanto trattasi, in larga parte, di unità non corredate dall'informazione legata alla propria destinazione d'uso.

I miglioramenti acquisiti nelle banche dati del catasto e quelli ancora in corso come interventi di bonifica, di fotoidentificazione e di revisione dei classamenti tendono a rendere sempre più corrispondente la situazione rappresentata negli archivi a quella realmente presente sul territorio nazionale.

Il processo di valorizzazione statistica delle informazioni contenute negli archivi dell'Agenzia fornisce sia un utile quadro informativo ai fini della trasparenza del mercato immobiliare sia un supporto allo studio ed alla analisi economico-sociale del territorio.

L'Osservatorio del Mercato Immobiliare è gestito anch'esso dall'Agenzia del Territorio e rileva sull'intero territorio nazionale, con periodicità semestrale, i valori medi di compravendita e locazione di immobili ordinari. I segmenti di mercato rilevati sono rappresentati tramite la zonizzazione del territorio, in quanto la localizzazione è la caratteristica che incide maggiormente sul valore di mercato degli immobili. Pertanto, ciascun comune dell'intero territorio nazionale è suddiviso in zone omogenee (zone OMI) nelle quali si registra uniformità di apprezzamento per condizioni economiche e socio-ambientali. In ciascuna zona sono rilevati, su base campionaria, i valori medi di mercato dei fabbricati distinti per destinazione d'uso (residenziale, commerciale, terziaria, produttiva) e per tipologia di immobile (abitazioni economiche, ville e villini, box, uffici, negozi, ecc.).

In questo studio per le unità immobiliari a destinazione residenziale (abitazioni) sono stati elaborati i valori medi regionali sulla base delle quotazioni rilevate nel 2° semestre 2009, mediando i valori medi comunali e utilizzando una ponderazione con lo stock abitativo.

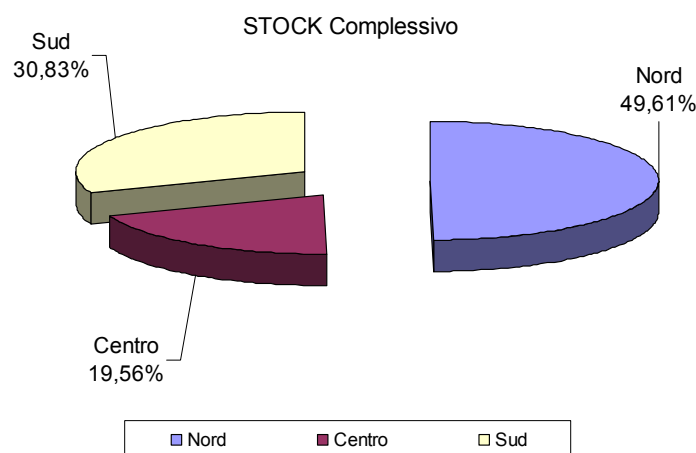
La numerosità dello stock immobiliare censito negli archivi catastali al 31.12 2009 è riportato nella tabella che segue.

**Figura 1**

<b>STOCK COMPLESSIVO 2009</b>									
<b>N° Unità Immobiliari per gruppi di categorie catastali</b>									
	<b>A</b>	<b>A10</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>	<b>Altro</b>	<b>Totale</b>
<b>Nord</b>	14.820.121	338.192	81.791	12.846.511	636.488	37.621	951.817	2.939.163	<b>32.651.704</b>
<b>Centro</b>	6.252.134	132.539	41.148	4.643.172	200.414	16.003	470.355	1.121.557	<b>12.877.322</b>
<b>Sud</b>	11.411.816	135.586	45.246	6.010.828	244.418	18.929	1.128.715	1.297.101	<b>20.292.639</b>
<b>Italia</b>	<b>32.484.071</b>	<b>606.317</b>	<b>168.185</b>	<b>23.500.511</b>	<b>1.081.320</b>	<b>72.553</b>	<b>2.550.887</b>	<b>5.357.821</b>	<b>65.821.665</b>

La tabella restituisce il dettaglio nazionale e nelle macro aree geografiche del paese (Nord, Centro, Sud) del numero totale di unità immobiliari distinto per gruppi di categorie catastali. Si evidenzia un totale nazionale di quasi 66 milioni di unità di cui composto da 32,6 milioni unità (quasi il 50%) al Nord, 12,9 milioni (quasi il 20%) al Centro e 20,3 milioni (circa il 30%) al Sud.

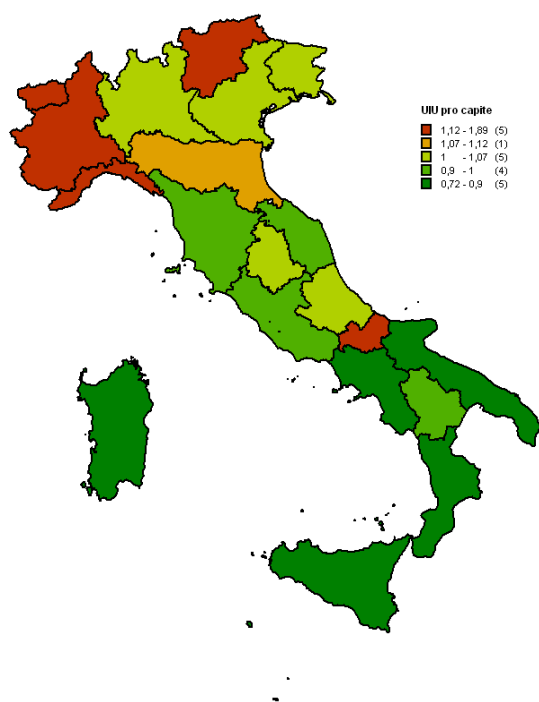
**Figura 2**



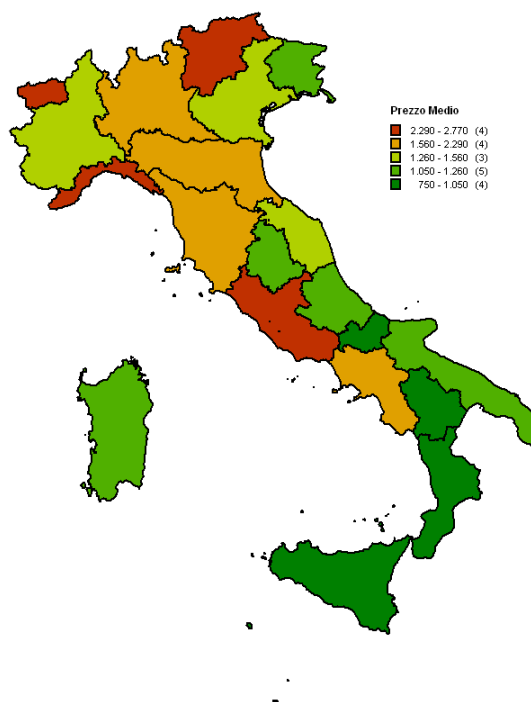
La mappa in Figura 3 riporta la distribuzione regionale delle unità immobiliari pro capite, evidenziando un valore superiore all'unità per le regioni del Nord, mentre al Centro, al Sud e nelle Isole i valori sono tutti inferiori, con le uniche eccezioni che riguardano le regioni Abruzzo, Umbria e Molise.

In Figura 4 è riportata la mappa dei valori medi di mercato, per unità di superficie, regionali delle abitazioni, costruite sulla base delle quotazioni immobiliari dell'Osservatorio del Mercato Immobiliare e pesate, nell'ambito regionale, con il relativo stock delle unità residenziali. Sono il Lazio, la Liguria, il Trentino Alto Adige e la Valle d'Aosta le regioni con i prezzi medi delle case più elevati. Di contro Molise, Basilicata, Sicilia e Calabria presentano quotazioni medie regionali inferiori a 1.050 €/mq.

**Figura 3**



**Figura 4**



## 2 Procedura di cluster

Un esame preliminare dello stock immobiliare censito nelle regioni italiane ha consentito di individuare quelle variabili in grado di mostrare, effettuato un adeguato processo di elaborazione, similitudini e diversità degli ambiti territoriali considerati. La tecnica di elaborazione prescelta è quella della cluster analysis.

La cluster analysis (analisi delle classi) consiste in una tecnica di analisi multivariata atta ad individuare gruppi di unità tra loro simili rispetto ad un insieme di caratteri presi in considerazione e secondo uno specifico criterio. Con la cluster analysis è possibile quindi raggruppare le unità statistiche, in modo da minimizzare la "*lontananza logica*" interna a ciascun gruppo e di massimizzare quella tra i gruppi. La "*lontananza logica*" viene quantificata per mezzo di misure di similarità/dissimilarità definite tra le unità statistiche.

L'obiettivo che ci si pone è sostanzialmente quello di riunire diverse unità statistiche in più sottoinsiemi tendenzialmente omogenei e mutuamente esaustivi. Le unità statistiche vengono, in altri termini, suddivise in un certo numero di gruppi a seconda del loro livello di "somiglianza" valutata a partire dai valori che una serie di variabili prescelte assume in ciascuna unità. L'applicazione della cluster analysis nel presente studio si è articolata nelle 4 fasi che saranno presentate nei paragrafi successivi e che sono sintetizzate di seguito:

- **scelta delle variabili di classificazione:** sono state individuate variabili legate allo stock immobiliare secondo macroaggregazioni che riflettono la densità regionale dello spazio costruito differenziato secondo le diverse destinazioni economiche dei beni immobiliari;
- **scelta della misura della distanza** tra le unità statistiche: è stato utilizzato il quadrato della distanza euclidea (corrispondente al concetto geometrico di distanza nello spazio multidimensionale), misura che assegna un peso progressivamente maggiore agli oggetti che stanno oltre una certa distanza;
- **scelta dell'algoritmo di raggruppamento** delle unità statistiche: è stata scelta una tecnica gerarchica che conduce, cioè, ad un insieme di gruppi ordinabili secondo livelli crescenti, con un numero di gruppi da  $n$  ad 1 (la distinzione che normalmente viene proposta è fra metodi gerarchici e non gerarchici);
- **scelta del numero ottimale di gruppi** attraverso la valutazione della partizione ottenuta;
- **presentazione e interpretazione dei risultati** ottenuti (connotazione dei gruppi).

## 2.1 Scelta delle variabili di classificazione

La prima fase è stata condotta analizzando lo stock immobiliare delle singole regioni italiane preliminarmente aggregato sotto il profilo della destinazione economica del bene.

Sono state innanzitutto determinate 8 macro aggregazioni (7 + 1 residuo) dello stock immobiliare censito nelle diverse tipologie, individuate dalle categorie catastali, che riguardano le possibili destinazioni d'uso degli immobili.

**Figura 5**

<i>COD</i>	<i>Aggregazione</i>	<i>Categorie catastali</i>	<i>Numero unità</i>
ABI	abitazioni	da A1 a A11 (esclusa A10)	33.095.409
PRT	unità pertinenziali	C2, C6 e C7	21.451.490
COM	unità commerciali	C1 e D5	2.112.783
PRO	unità produttive	C3, D1, D7 e D10	1.430.592
TER	unità terziarie	A10, B4 e D5	678.435
RCU	unità ricreative e culturali	B6, C4, C5, D2, D3 e D6	120.862
SIS	unità per la sanità e istruzione	B2, B5 e D4	64.548
AL	altre destinazioni d'uso	B1, B3, B7, B8, D9	76.726

Escludendo il residuo di unità contenuto in altro (AL), che riguarda immobili peculiari (prigioni, caserme, cappelle, edifici galleggianti), si sono ritenute le altre 7 aggregazioni significative per identificare le abitazioni sul territorio, le dotazioni infrastrutturali (scuole ed ospedali, unità ricreative, turistiche e culturali) e gli spazi produttivi, terziari e commerciali.

È necessario precisare che, generalmente, le variabili da utilizzare in un'elaborazione di *clustering* debbono essere espresse nella stessa unità di misura. Se le variabili quantitative da utilizzare nella cluster sono espresse in unità di misura diverse o hanno ordini di grandezza diversi è opportuno standardizzare le variabili.

A tal fine, per omogeneizzare immobili che, nell'ambito della stessa macroaggregazione, presentano caratteristiche dimensionali molto diverse, seppur con stessa destinazione d'uso economica del bene (ad esempio A10 e D5), sono stati utilizzati coefficienti di ragguaglio calcolati come rapporto tra i corrispondenti valori nazionali delle rendite delle diverse categorie (ad esempio il valore delle rendite delle unità in D5 risulta 8,4 volte il valore complessivo delle rendite delle unità A10). In Figura 6 è riportata la tabella con i valori calcolati dei coefficienti utilizzati.

**Figura 6**

		Categorie catastali	Rendita Catastale Media x unità (€)	Coefficiente di ragguaglio
PRO	Unità produttive	C3	514	1,0
		D1	6.083	11,8
		D7	10.262	20,0
TER	Unità terziarie	D5	19.403	8,4
		B4	6.275	2,7
		A10	2.310	1,0
COM	Unità commerciali	D8	12.598	7,5
		C1	1.686	1,0
RCU	Unità per attività ricreative, turistiche e culturali	B6	10.505	8,9
		C4	1.175	1,0
		C5	1.135	1,0
		D6	8.069	6,9
		D2	19.260	16,4
		D3	11.347	9,7

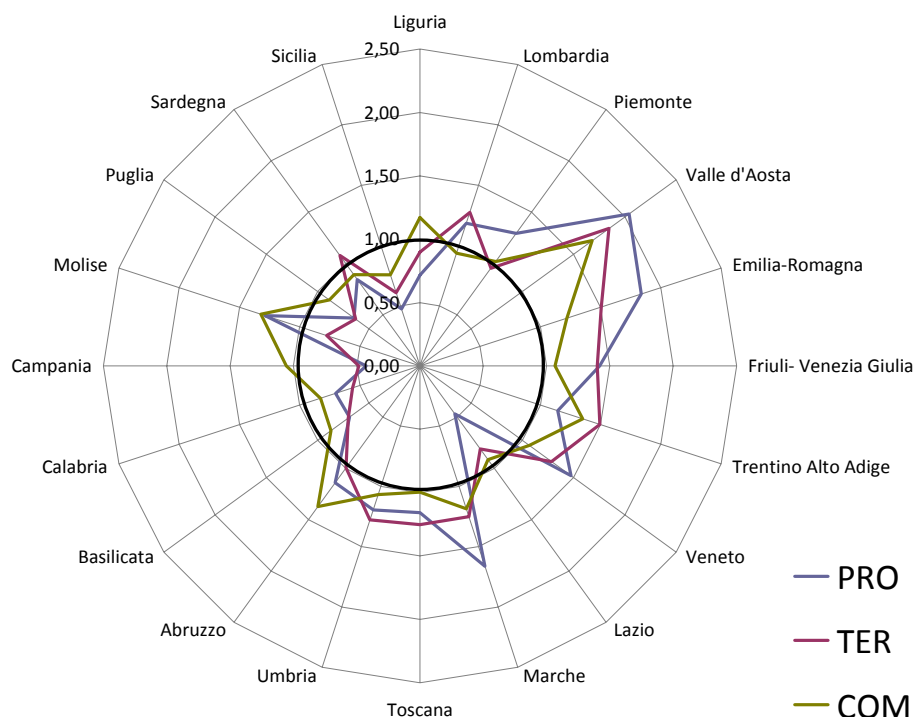
L'analisi delle distribuzioni regionali delle macro aggregazioni ha portato poi ad escludere dalla classificazione in cluster quelle relative alle unità pertinenti (PRT) in quanto, non essendo possibile scorporare (se non con ulteriori elaborazioni con ipotesi basate sulle dimensioni delle unità) la quota di tali unità relativa agli immobili residenziali (box, posti auto, cantine o soffitte) da quella legata all'uso non residenziale (autorimesse, magazzini), si ritiene alto il rischio di una distorsione dei risultati delle elaborazioni. Le variabili sono state quindi standardizzate costruendo, a partire dallo stock immobiliare per abitante in ogni regione, il rapporto del valore regionale rispetto a quello nazionale. Tale rapporto, in altri termini, restituisce per ogni regione il differenziale rispetto al dato procapite nazionale. Le variabili regionali utilizzate nella classificazione in cluster sono pertanto le seguenti:

1. PRO – stock produttivo
2. TER – stock terziario
3. COM – stock commerciale
4. SIS – stock per sanità ed istruzione
5. RCU – stock per attività turistiche, ricreative e culturali
6. ABI – stock abitazioni

Nelle figure che seguono sono riportati, in appositi grafici, i valori delle variabili individuate per le 20 regioni italiane. La Figura 7 mette in risalto i valori delle variabili relative allo stock destinato alle attività terziarie, commerciali e produttive e in particolare evidenzia che le regioni del Nord Est e la Valle d'Aosta sono quelle con un valore medio dello stock pro capite più alto rispetto alla media nazionale. Risaltano inoltre il valore dello stock produttivo procapite della regione Marche e, di contro, i valori, inferiori alla media nazionale, relativi allo stock non residenziale delle regioni del Sud e delle Isole.

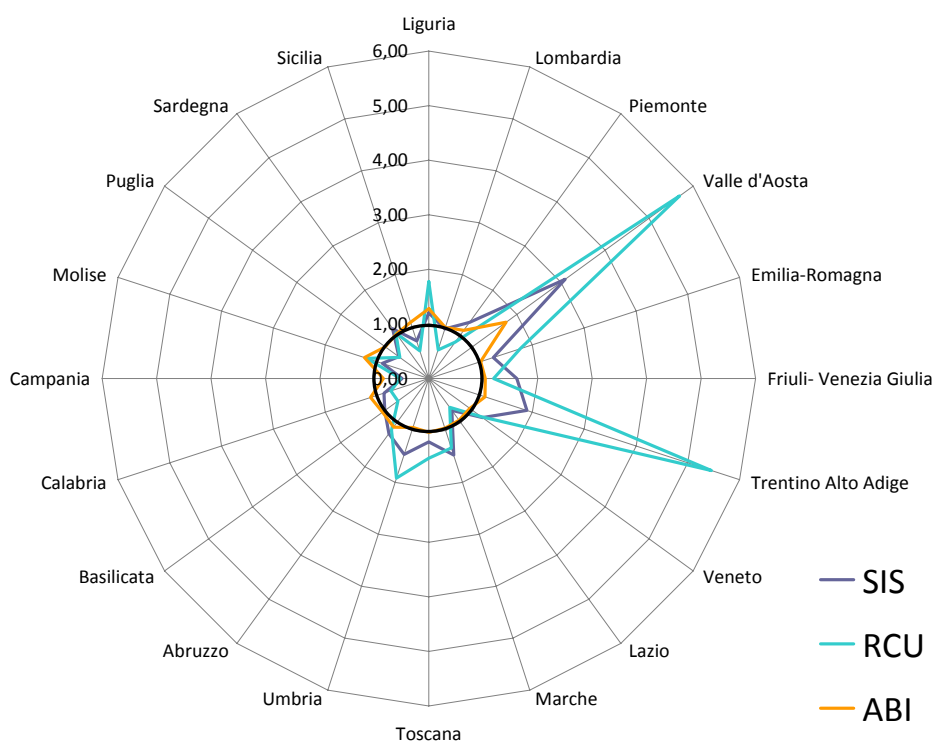


**Figura 7**



In Figura 8, la distribuzione dei valori delle variabili relative allo stock delle abitazioni (ABI), delle infrastrutture sanitarie e scolastiche (SIS) e degli immobili per le attività ricreative, turistiche e culturali (RCU), mette in evidenza i valori decisamente oltre il valore nazionale delle regioni autonome del Trentino alto Adige e della Valle d'Aosta, soprattutto relativamente a RCU (quasi 6 volte il dato nazionale).

**Figura 8**



## 2.2 *Classificazione del territorio con la cluster analysis*

Esistono molti metodi di classificazione che utilizzano diverse misure di distanza o dissimilarità. Si utilizzerà, nelle elaborazioni del presente lavoro, come si è già detto, la distanza quadratica euclidea per misurare le distanze e la tecnica di clustering con il legame di Ward per la classificazione dei gruppi.

La tecnica di clustering di Ward (1963) si distingue da tutti gli altri metodi in quanto utilizza un approccio di analisi della varianza per valutare le distanze tra i cluster. In breve, questo metodo tenta di minimizzare la somma dei quadrati (SS) di due (ipotetici) cluster che si possono formare ad ogni passo.

Il metodo di Ward, quindi, opera considerando, ad ogni passo dell'analisi, l'unione di tutte le possibili coppie di cluster e i due cluster, il cui raggruppamento fornisce il minimo aumento di errore nella somma dei quadrati, vengono combinati. Ad ogni stadio dell'analisi la perdita di informazione (information loss) che deriva dal raggruppamento di elementi può essere misurata dalla somma delle deviazioni quadratiche di tutti i punti.

Una partizione si considera tanto migliore quanto più le classi risultano omogenee al loro interno e differenti l'una dall'altra. In altri termini, quanto più è elevata la varianza tra le classi, e bassa la varianza interna (alle classi). È noto che la varianza totale di un insieme di unità, si può scomporre nella somma di due quantità: varianza interna (ai cluster) e varianza esterna (cioè tra i cluster). In maniera analoga si scompone la matrice di varianze e covarianze  $S$ . In simboli:

$$S = S_W + S_B$$

Dove  $S$  è la matrice di varianze e covarianze totali;

$S_W$  è la matrice delle varianze e covarianze "interne";

$S_B$  è la matrice delle varianze e covarianze "esterne".

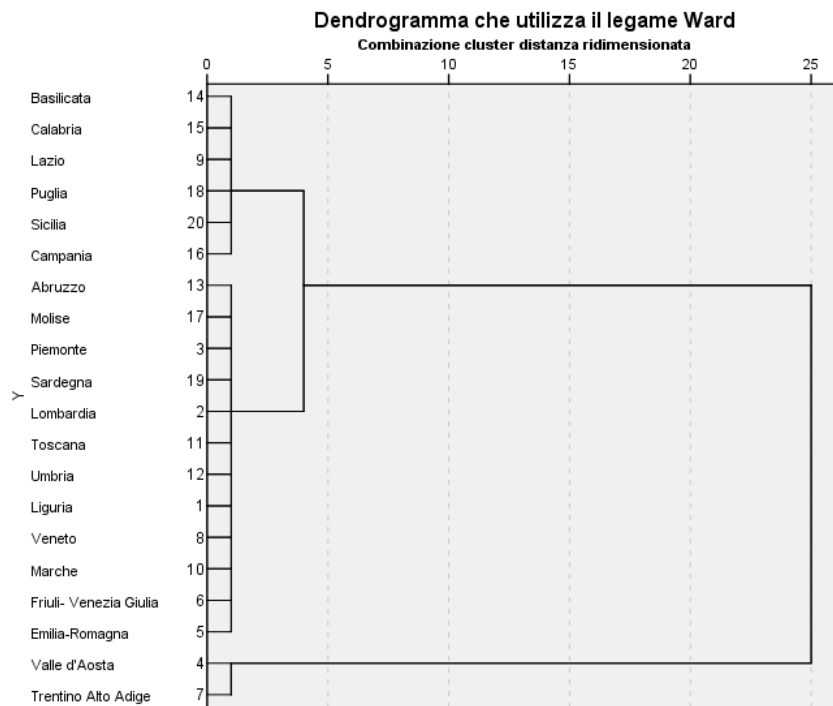
In generale, questo metodo è considerato molto efficiente, tuttavia, tende a creare cluster di piccole dimensioni.

Le elaborazioni sono state effettuate avvalendosi del software SPSS.

### ***Elaborazione 1***

L'elaborazione 1 ha riguardato le 6 variabili presentate nel paragrafo 2.1 e ha prodotto l'aggregazione gerarchica indicata nel dendrogramma riportato in Figura 9.

**Figura 9**



### ***Elaborazione 2***

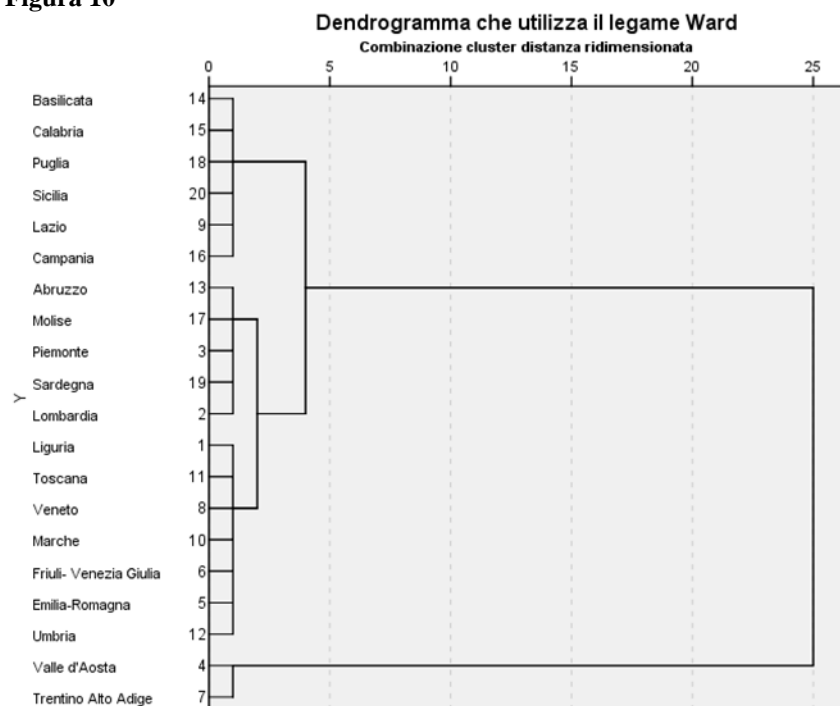
Una seconda elaborazione è stata effettuata inserendo una nuova variabile con il fine di identificare gli ambiti territoriali regionali con un valore monetario legato agli immobili.

La nuova variabile è costruita con il valore medio (per unità di superficie) regionale delle abitazioni elaborato sulla base delle quotazioni immobiliari desunte dalla banca dati dell'Osservatorio del Mercato Immobiliare (OMI). La variabile, denominata VAL, è rappresentata dal differenziale rispetto al valore medio nazionale delle abitazioni.

VAL – differenziale livello prezzi abitazioni

L'elaborazione 2 ha quindi riguardato 7 variabili e ha prodotto l'aggregazione gerarchica indicata nel dendrogramma riportato in Figura 10.

**Figura 10**

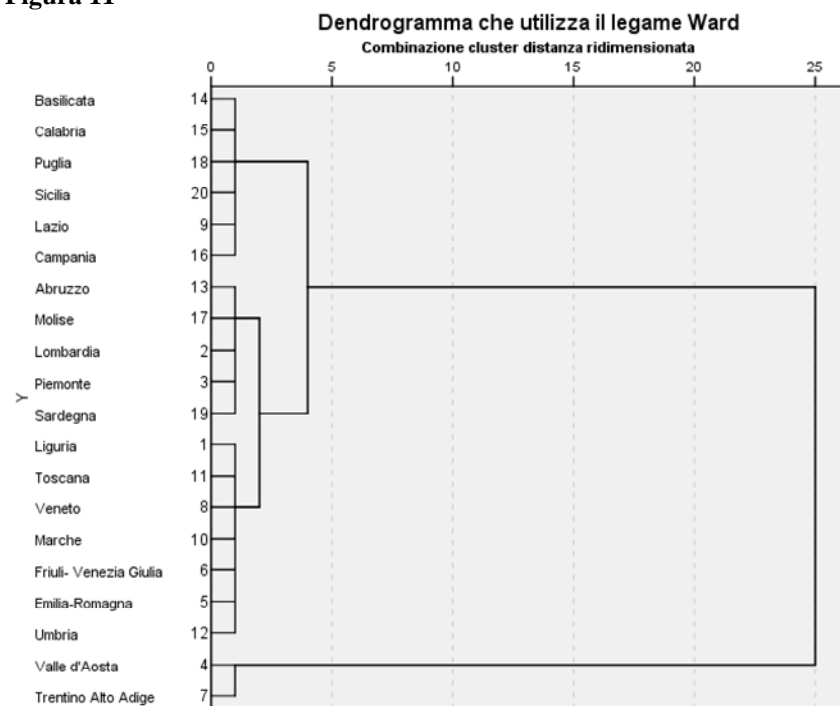


### **Elaborazione 3**

È stata effettuata un'ulteriore elaborazione, inserendo ancora una variabile “economica” legata al valore del prodotto interno lordo regionale nel 2009 (fonte ISTAT).

La variabile standardizzata, denominata PIL, è rappresentata dal differenziale regionale del valore procapite al dato procapite nazionale. L'elaborazione 3 ha quindi riguardato 8 variabili e ha prodotto l'aggregazione gerarchica indicata nel dendrogramma riportato in Figura 11.

**Figura 11**

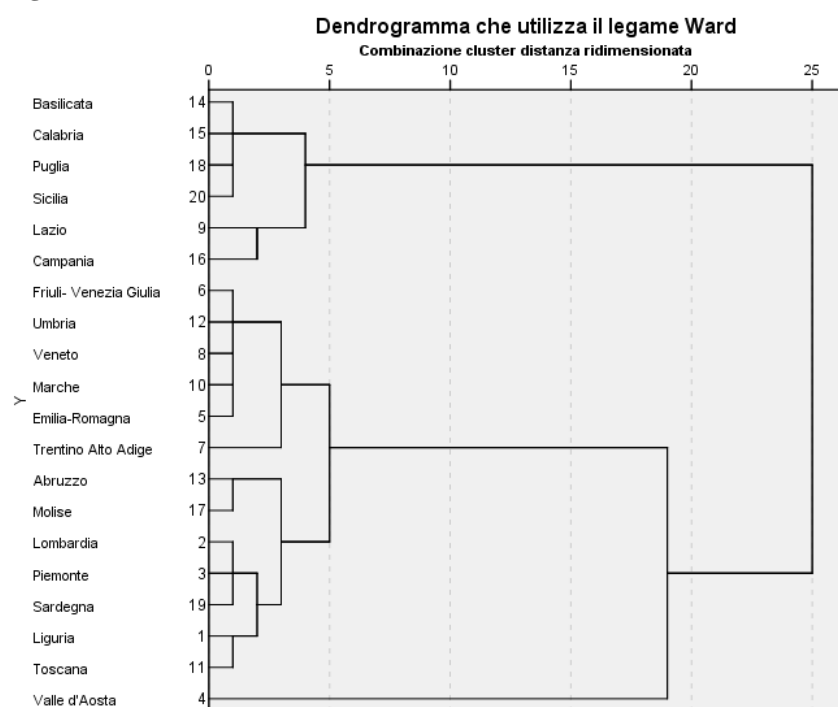


#### Elaborazione 4

L'ultima elaborazione è stata effettuata, eliminando dalle variabili fin qui elaborate, quella che presenta la maggiore dispersione nei valori regionali. La variabile eliminata dalla cluster analysis è la RCU che presenta una varianza 5,4 volte la varianza complessiva delle variabili esaminate.

L'elaborazione 4 ha quindi riguardato 7 variabili e ha prodotto l'aggregazione gerarchica indicata nel dendrogramma riportato in Figura 12.

**Figura 12**



In appendice sono riportate le tabelle contenenti i dati risultanti dalle elaborazioni di clustering.

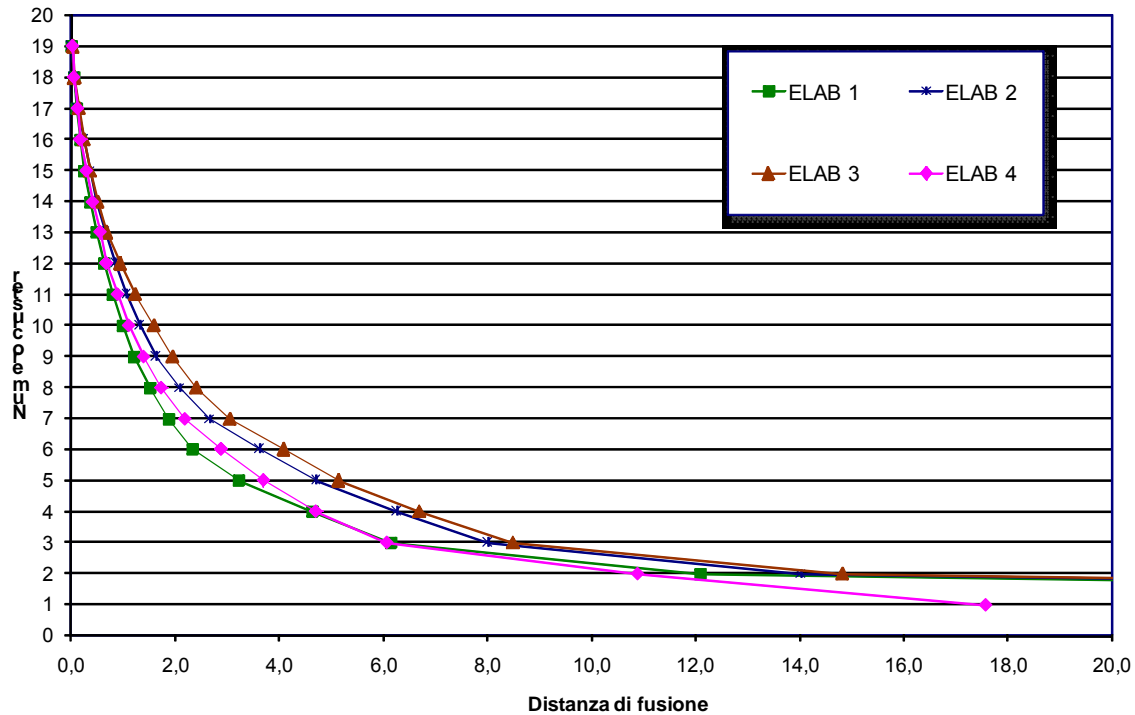
#### 2.3 Scelta del numero di gruppi

Svolte le elaborazioni di *clustering* con il legame di Ward, sono stati esaminati i risultati con il fine di pervenire ad una partizione tale da ottenere classi sufficientemente omogenee al loro interno e differenti l'una dall'altra, minimizzando la perdita di informazioni (*"loss information"*).

Nel caso di una cluster gerarchica la scelta del numero di cluster può essere effettuata utilizzando in primo luogo la distanza di fusione, che può essere letta nel dendrogramma: se nel passaggio da K gruppi a K+1 si registra un forte incremento della distanza di fusione si deve "tagliare l'albero" a K gruppi.

Con la distanza di fusione si possono costruire gli scree plot, grafici in cui viene posto in ordinata il numero di gruppi ed in ascissa la distanza di fusione. Il grafico in Figura 13 riporta il grafico scree plot con le risultanze delle quattro elaborazioni e “suggerisce” di tagliare a 3 gruppi. Infatti, nel passaggio a 2 gruppi si registra un consistente incremento della distanza di fusione.

**Figura 13**



In generale quest’analisi per la scelta del numero di gruppi si basa comunque su una di discontinuità dei dati osservata e questo può risultare una procedura eccessivamente soggettiva.

È possibile valutare l’entità dell’incremento della distanza di fusione calcolando l’incremento relativo della distanza di fusione:

$$\delta_k = (d_k - d_{k+1}) / d_{k+1}$$

verrà scelto il K in corrispondenza del  $\delta_k$  massimo.

Adottando questo tipo di metodo saranno scelti 10 gruppi per le prime 3 elaborazioni e 7 gruppi per l’elaborazione 4.

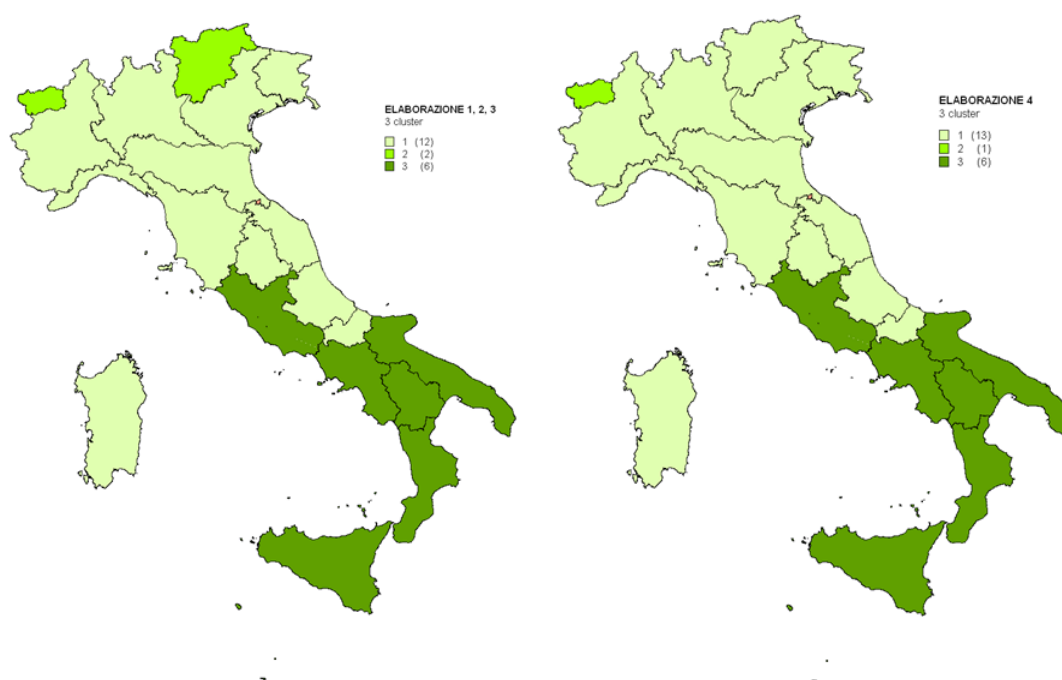
## 2.4 Presentazione dei risultati

In questo paragrafo, individuate le variabili, concluse le elaborazioni di clustering ed effettuata la scelta di classi sufficientemente omogenee al loro interno e differenti l’una dall’altra, si presentano i risultati.

Le mappe in Figura 14 mostrano i cluster risultanti dalle 4 elaborazioni nel caso di taglio a 3 gruppi in tutti i casi. Le mappe evidenziano suddivisioni in gruppi identiche nelle prime 3 elaborazioni, caratterizzate da un primo cluster contenente tutte le regioni del Nord, quelle del Centro, ad eccezione del Lazio, e l'Abruzzo, il Molise e la Sardegna. Il secondo cluster isola le peculiarità del Trentino Alto Adige e della Valle d'Aosta. Infine nella terza classe si trovano oltre il Lazio le rimanenti regioni del Sud.

I cluster prodotti dall'elaborazione 4 si differenziano solo per l'isolamento della Valle d'Aosta nella classe 2, con il Trentino Alto Adige che entra nella classe del Nord.

**Figura 14**



La Figura 15 mostra i cluster risultanti dalle 4 elaborazioni nel caso di taglio al  $\delta_k$  massimo. Risalta immediatamente che le elaborazioni 2 e 3 forniscono, una volta effettuata la scelta del numero di gruppi, un'identica classificazione delle regioni italiane.

Inoltre le mappe evidenziano in tutte le classificazioni la peculiarità delle regioni autonome del Trentino Alto Adige e Valle d'Aosta i nessun caso "simili" ad altre regioni italiane.

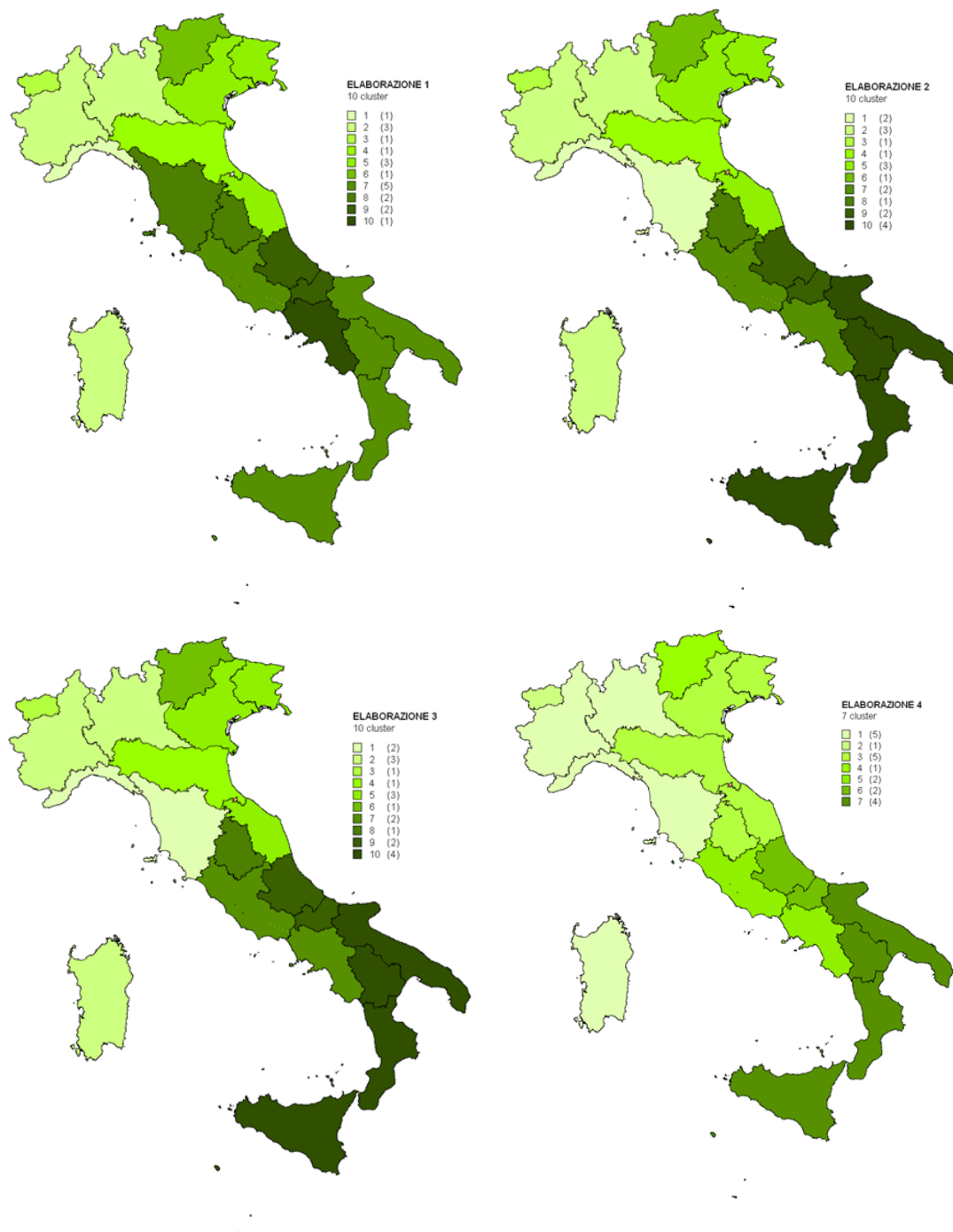
L'elaborazione 1, basata su variabili costruite tutte sulla distribuzione dello stock immobiliare, evidenzia altre regioni singolari: Liguria, Emilia Romagna e Campania. Risultano nello stesso cluster la Lombardia, il Piemonte e la Sardegna; il Friuli Venezia Giulia, il Veneto e le Marche. Toscana e Umbria al Centro e Abruzzo e Molise al Sud sono i cluster con solo due regioni. Nel cluster più numeroso rientrano il Lazio, la Basilicata, la Calabria, la Puglia e la Sicilia.

L'aver introdotto le variabili VAL e PIL (prezzo abitazioni e prodotto interno lordo) evidenzia una diversa classificazione delle regioni, fermo restando la singolarità di cui si è già detto del Trentino e della Valle d'Aosta.

Sono in tal caso Umbria ed Emilia Romagna le regioni “singolari”; la Campania trova una “vicinanza” con il Lazio e la Liguria con la Toscana. Risultano invariati i cluster con Lombardia, Piemonte e Sardegna; Friuli Venezia Giulia, Veneto e Marche; Abruzzo e Molise. Le regioni Basilicata, Calabria, Puglia e Sicilia restano nello stesso gruppo che però perde il Lazio.

L’elaborazione 4 fornisce, come si è detto, un raggruppamento in 7 cluster che, come mostra la mappa in Figura 15 individuano aree territoriali generalmente contigue del nostro paese.

**Figura 15**





### 3 Conclusioni

Le analisi proposte in questo studio hanno condotto ad evidenziare alcune diversità e similitudini tra le regioni italiane, utilizzando le informazioni relative alla densità degli immobili articolati secondo le principali destinazioni d'uso.

Il patrimonio informativo dell'Agenzia del Territorio, che gestisce, tra l'altro, i dati censuari di oltre 65 milioni di unità immobiliari urbane e l'archivio delle quotazioni immobiliari dell'Osservatorio del Mercato Immobiliare, ha consentito di caratterizzare le regioni italiane con variabili standardizzate costruite a partire dallo stock immobiliare per abitante in ogni regione e sul livello dei prezzi medi delle abitazioni.

Il procedimento di clustering utilizzato ha proposto una serie di risultati non immediatamente riconoscibili. Infatti un aspetto critico di un'elaborazione di *clustering* è sicuramente quello della scelta del "taglio" del dendrogramma risultante.

Sono stati proposti due metodi di scelta: il primo sulla base dello scree plot delle distanze di fusione dei gruppi il secondo sulla base dell'incremento relativo della distanza di fusione in due stadi di elaborazione successivi.

Il primo metodo ha restituito, nelle 4 elaborazioni proposte, 3 cluster che individuano un'Italia già ampiamente nota nelle sue diversità tra Nord e Sud, ma con un'interessante vicinanza di talune regioni del Centro e del Sud a quelle del Nord, che necessita sicuramente approfondimenti in studi successivi.

Il secondo metodo, meno soggettivo ma forse troppo cautelativo, evidenzia invece cluster meno ampie, dove vi si ritrovano generalmente regioni contigue.

In ogni caso, un'evidenza risalta su tutto: le regioni autonome del Trentino Alto Adige e della Valle d'Aosta hanno caratteristiche peculiari e per questo molto distanti dalle altre.

Si ritiene di aver fornito con questo studio un utile contributo alla comprensione delle dinamiche territoriali, affidando a successivi lavori gli sviluppi necessari per approfondire le ragioni di talune diversità e similarità proposte dai risultati raggiunti.

## Appendice

### Elaborazione 1

CLUSTER PRO TER COM SIS RCU ABI

METHOD WARD

MEASURE=SEUCLID

Stadio	Cluster accorpati		Distanza di fusione	$\bar{O}_k$	Stadio di formazione del cluster		Stadio successivo
	Cluster 1	Cluster 2			Cluster 1	Cluster 2	
1	14	15	,023	,000	0	0	7
2	9	18	,057	,000	0	0	3
3	9	20	,103	,000	2	0	7
4	8	10	,172	,000	0	0	5
5	6	8	,259	,000	0	4	13
6	13	17	,366	,000	0	0	14
7	9	14	,484	,000	3	1	11
8	3	19	,637	,000	0	0	10
9	11	12	,794	,000	0	0	12
<b>10</b>	<b>2</b>	<b>3</b>	<b>,986</b>	<b>,000</b>	<b>0</b>	<b>8</b>	<b>14</b>
11	9	16	1,199	,000	7	0	18
12	1	11	1,522	,000	0	9	15
13	5	6	1,872	,000	0	5	15
14	2	13	2,332	,000	10	6	16
15	1	5	3,224	,000	12	13	16
16	1	2	4,644	,000	15	14	18
17	4	7	6,138	,000	0	0	19
18	1	9	12,077	,000	16	11	19
19	1	4	54,365	,000	18	17	0

### Elaborazione 2

CLUSTER PRO TER COM SIS RCU ABI VAL

METHOD WARD

MEASURE=SEUCLID

Stadio	Cluster accorpati		Distanza di fusione	$\bar{O}_k$	Stadio di formazione del cluster		Stadio successivo
	Cluster 1	Cluster 2			Cluster 1	Cluster 2	
1	14	15	,024	-,626	0	0	5
2	18	20	,064	-,523	0	0	5
3	8	10	,134	-,444	0	0	4
4	6	8	,241	-,343	0	3	12
5	14	18	,366	-,262	1	2	14
6	13	17	,496	-,236	0	0	13
7	3	19	,649	-,231	0	0	8
8	2	3	,844	-,202	0	7	13
9	9	16	1,057	-,202	0	0	14
<b>10</b>	<b>1</b>	<b>11</b>	<b>1,325</b>	<b>-,193</b>	<b>0</b>	<b>0</b>	<b>15</b>
11	5	12	1,643	-,217	0	0	12
12	5	6	2,097	-,216	11	4	15
13	2	13	2,676	-,265	8	6	17
14	9	14	3,639	-,225	9	5	18
15	1	5	4,698	-,250	10	12	17
16	4	7	6,262	-,218	0	0	19
17	1	2	8,011	-,429	15	13	18
18	1	9	14,031	-,752	17	14	19
19	1	4	56,663	,000	18	16	0

### Elaborazione 3

CLUSTER PRO TER COM SIS RCU ABI VAL PIL

METHOD WARD

MEASURE=SEUCLID

Stadio	Cluster accorpati		Distanza di fusione	$\bar{D}_k$	Stadio di formazione del cluster		Stadio successivo
	Cluster 1	Cluster 2			Cluster 1	Cluster 2	
1	14	15	,026	-,605	0	0	5
2	18	20	,066	-,541	0	0	5
3	8	10	,144	-,428	0	0	4
4	6	8	,252	-,334	0	3	12
5	14	18	,379	-,255	1	2	14
6	13	17	,509	-,252	0	0	13
7	2	3	,680	-,283	0	0	9
8	1	11	,949	-,232	0	0	15
9	2	19	1,235	-,223	7	0	13
<b>10</b>	<b>9</b>	<b>16</b>	<b>1,590</b>	<b>-,183</b>	<b>0</b>	<b>0</b>	<b>14</b>
11	5	12	1,945	-,190	0	0	12
12	5	6	2,401	-,212	11	4	15
13	2	13	3,046	-,253	9	6	17
14	9	14	4,077	-,206	10	5	18
15	1	5	5,136	-,233	8	12	17
16	4	7	6,700	-,212	0	0	19
17	1	2	8,501	-,426	15	13	18
18	1	9	14,803	-,743	17	14	19
19	1	4	57,662	,000	18	16	0

### Elaborazione 4

CLUSTER PRO TER COM SIS ABI VAL PIL

METHOD WARD

MEASURE=SEUCLID

Stadio	Cluster accorpati		Distanza di fusione	$\bar{D}_k$	Stadio di formazione del cluster		Stadio successivo
	Cluster 1	Cluster 2			Cluster 1	Cluster 2	
1	14	15	,026	-,552	0	0	5
2	18	20	,057	-,534	0	0	5
3	6	12	,123	-,353	0	0	11
4	8	10	,191	-,369	0	0	6
5	14	18	,303	-,284	1	2	16
6	5	8	,423	-,235	0	4	11
<b>7</b>	<b>13</b>	<b>17</b>	<b>,553</b>	<b>-,196</b>	<b>0</b>	<b>0</b>	<b>15</b>
8	2	3	,687	-,231	0	0	9
9	2	19	,894	-,198	8	0	13
10	1	11	1,114	-,201	0	0	13
11	5	6	1,395	-,199	6	3	14
12	9	16	1,742	-,206	0	0	16
13	1	2	2,193	-,242	10	9	15
14	5	7	2,893	-,216	11	0	17
15	1	13	3,691	-,217	13	7	17
16	9	14	4,716	-,223	12	5	19
17	1	5	6,071	-,442	15	14	18
18	1	4	10,889	-,380	17	0	19
19	1	9	17,567	,000	18	16	0

Cluster risultanti dalle elaborazioni

ELAB 1	10 cluster	ELAB 2	10 cluster
Liguria	1	Liguria	1
Lombardia	2	Toscana	1
Piemonte	2	Lombardia	2
Sardegna	2	Piemonte	2
Valle d'Aosta	3	Sardegna	2
Emilia-Romagna	4	Valle d'Aosta	3
Friuli- Venezia Giulia	5	Emilia-Romagna	4
Veneto	5	Friuli- Venezia Giulia	5
Marche	5	Veneto	5
Trentino Alto Adige	6	Marche	5
Lazio	7	Trentino Alto Adige	6
Basilicata	7	Lazio	7
Calabria	7	Campania	7
Puglia	7	Umbria	8
Sicilia	7	Abruzzo	9
Toscana	8	Molise	9
Umbria	8	Basilicata	10
Abruzzo	9	Calabria	10
Molise	9	Puglia	10
Campania	10	Sicilia	10

ELAB 3	10 cluster	ELAB 4	7 cluster
Liguria	1	Liguria	1
Toscana	1	Lombardia	1
Lombardia	2	Piemonte	1
Piemonte	2	Toscana	1
Sardegna	2	Sardegna	1
Valle d'Aosta	3	Valle d'Aosta	2
Emilia-Romagna	4	Emilia-Romagna	3
Friuli- Venezia Giulia	5	Friuli- Venezia Giulia	3
Veneto	5	Veneto	3
Marche	5	Marche	3
Trentino Alto Adige	6	Umbria	3
Lazio	7	Trentino Alto Adige	4
Campania	7	Lazio	5
Umbria	8	Campania	5
Abruzzo	9	Abruzzo	6
Molise	9	Molise	6
Basilicata	10	Basilicata	7
Calabria	10	Calabria	7
Puglia	10	Puglia	7
Sicilia	10	Sicilia	7

Tabella delle macro aggregazioni delle categorie catastali

<b>RESIDENZIALE</b>	
A1	abitazioni di tipo signorile
A2	abitazioni di tipo civile
A3	abitazioni di tipo economico
A4	abitazioni di tipo popolare
A5	abitazioni di tipo ultrapopolare
A6	abitazioni di tipo rurale
A7	abitazioni in villini
A8	abitazioni in ville
A9	castelli, palazzi di eminenti pregi artistici o storici
A11	abitazioni ed alloggi tipici dei luoghi

<b>TERZIARIO</b>	
A10	uffici e studi privati
B4	uffici pubblici
D5	istituti di credito, cambio e assicurazione

<b>COMMERCIALE</b>	
C1	negozi e botteghe
D8	fabbricati costruiti per un'attività commerciale e non modificabili

<b>PRODUTTIVO</b>	
C3	laboratori per arti e mestieri
D1	opifici
D7	fabbricati costruiti per un'attività industriale e non modificabili
D10	fabbricati per funzioni produttive connesse alle attività agricole

<b>SANITA' E ISTRUZIONE</b>	
B2	case di cura e ospedali pubblici
B5	scuole e laboratori scientifici
D4	case di cura e ospedali privati

<b>ATTIVITA' RICREATIVE, TURISTICHE E CULTURALI</b>	
B6	biblioteche, pinacoteche, musei, gallerie, accademie
C4	fabbricati e locali per esercizi sportivi
C5	stabilimenti balneari e di acque curative
D2	alberghi e pensioni
D3	teatri, cinematografi, sale per concerti e spettacoli e simili
D6	fabbricati e locali per esercizi sportivi

## 4 Bibliografia

Agenzia del Territorio (2009), *Manuale della Banca dati dell'Osservatorio del Mercato immobiliare*

Agenzia del Territorio (2010), *Statistiche catastali 2009*

Dipartimento delle Finanze, Agenzia del Territorio (2010), *Gli immobili in Italia 2010*

Andenberg M. (1973), *Cluster analysis for applications*, New York Academic Press.

Fabbris L. (1983), *Analisi esplorativa di dati multidimensionali*, Cleup editore.

Hartigan J.A. (1975), *Clustering Algorithms*, Wiley.

Johanson S.C.(1967), *Hierarchical clustering schemes*, Psychometrika.

Kendall M. (1975), *Multivariate analysis*, Charles Griffin & Company, London.

Morrison D.F. (1976), *Multivariate statistical methods*, Mc Graw Hill.

## **ABSTRACT**

This study explores the differences and the similarities between the Italian regions using information about the buildings in different geographical areas. The information on the number and classification of real estate is in the archives of Italian cadastre, managed by the Agenzia del Territorio, allow to characterize each region through the economic use of his property. The examination of the real estate in the Italian regions has allowed, in fact, to identify those variables able to show, after an appropriate processing, similarities and differences. It was chosen for this purpose to proceed using the processing technique of cluster analysis. Data processing was performed on the regional numerousness of different types of real estate, compared to the resident population, with the aim to show the relationships between the different regions that demographic and economic data were able to show. The results of the analysis then allows to gain new knowledge about the Italian regions, in terms of "density" of its real estate composition. This picture of Italy will support policy development and local governance, bringing its contribution to improving the understanding of regional multiplicity.