



UNIVERSITÀ
DEGLI STUDI DI BARI
ALDO MORO

DIPARTIMENTO DI ECONOMIA E FINANZA
CORSO DI LAUREA TRIENNALE IN
SCIENZE STATISTICHE
L-41 D.M. 270/04

TESI DI LAUREA
IN
STATISTICA MULTIVARIATA

ANALISI STATISTICA DELLA RACCOLTA DIFFERENZIATA IN ITALIA

Relatore:
Prof. Alessio Pollice

Laureando:
Davide Cacciatore

ANNO ACCADEMICO 2020/2021

INDICE

1	Introduzione.....	3
2	Raccolta differenziata pro-capite per le principali frazioni merceologiche.....	4
3	Regioni omogenee per raccolta pro-capite	
	3.1 Analisi Cluster.....	6
	3.2 Analisi delle componenti principali.....	7
4	Costi di gestione dei rifiuti urbani	
	4.1 Analisi fattoriale.....	9
5	Regioni per raccolta pro-capite e costi di gestione	
	5.1 Analisi della correlazione canonica.....	11
6	Analisi della dipendenza della percentuale di raccolta differenziata	
	6.1 Correlazione spaziale.....	13
	6.2 Modello di regressione spaziale.....	14
7	Conclusioni.....	18
8	Riferimenti bibliografici.....	19

1. Introduzione

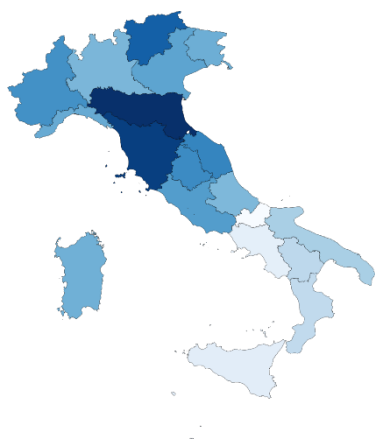
L'analisi svolta in questa tesi studia i dati della raccolta differenziata in Italia, forniti dal Catasto dei rifiuti, organizzato per via informatica dall'Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA).

Il Catasto ha lo scopo di fornire un quadro conoscitivo completo e facilmente accessibile in materia di rifiuti.

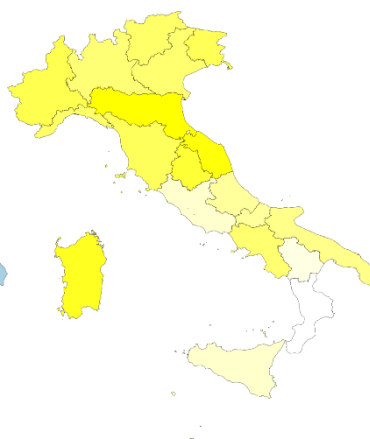
I dati su cui è stata svolta l'analisi si riferiscono al 2019 e riguardano la produzione e la raccolta differenziata dei rifiuti urbani e i loro costi di gestione.

2. Raccolta differenziata pro-capite per le principali frazioni merceologiche

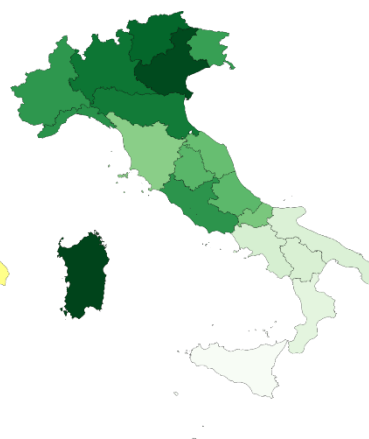
CARTA E CARTONE



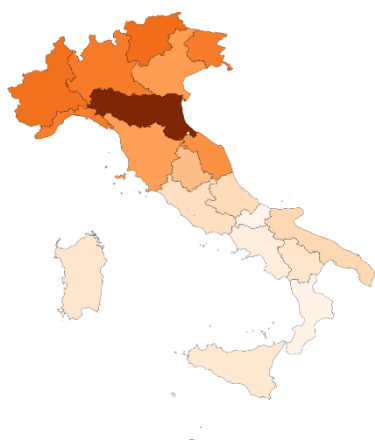
PLASTICA



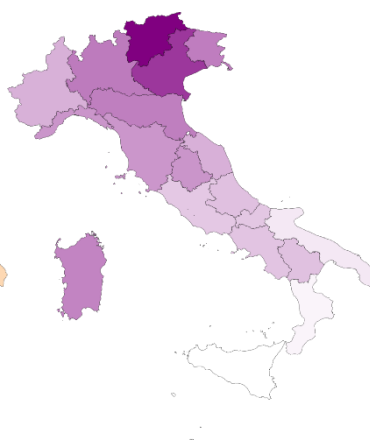
VETRO



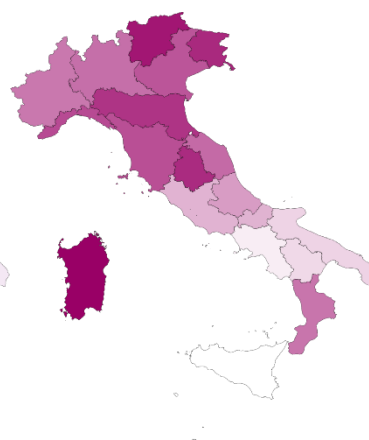
LEGNO



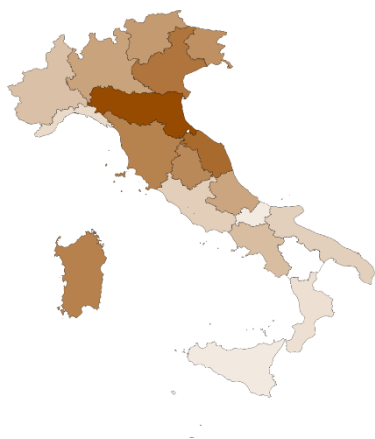
METALLO



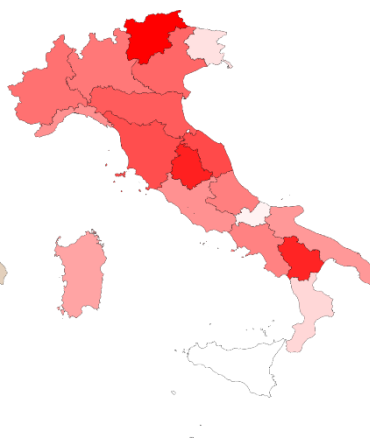
RAEE



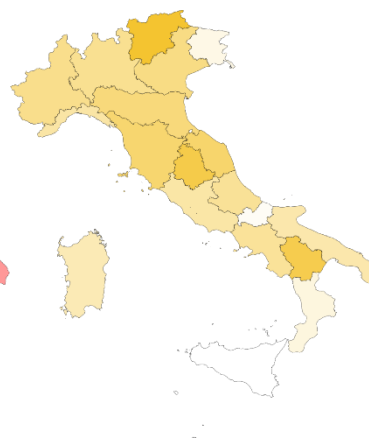
FRAZ ORGANICA



TESSILI



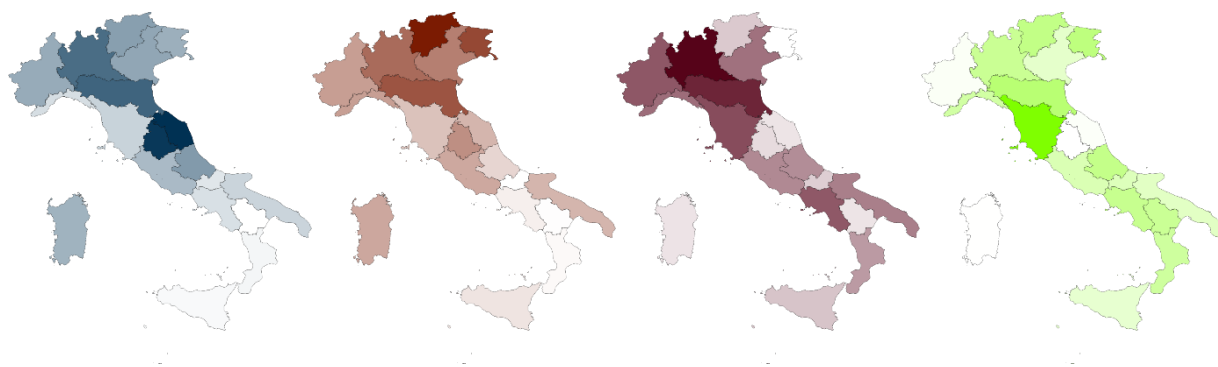
SELETTIVA



Sono state considerate le quantità pro-capite delle principali frazioni merceologiche di rifiuti. Nelle mappe un colore più intenso rappresenta una quantità maggiore di rifiuti di quella precisa categoria nella regione interessata.

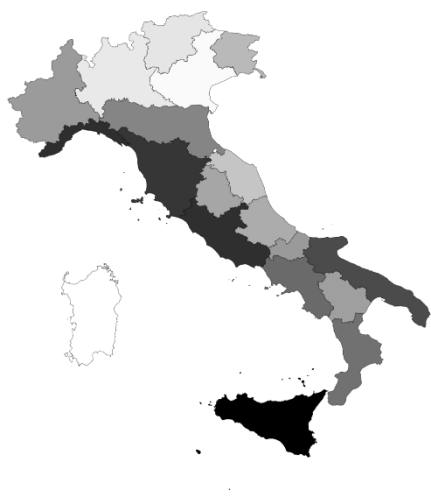
In particolare, osservando le principali frazioni merceologiche di raccolta differenziata (Carta e cartone, Plastica, Vetro, Legno, Metallo, Raee, Frazione organica, Tessili e Selettiva), si può già notare una sostanziale differenza presente tra le regioni del Nord e quelle del Sud Italia. Queste ultime, infatti, hanno quasi sempre colori poco intensi, ad indicare bassi valori di rifiuti raccolti per quella frazione.

PULIZIA STRADALE COST. E DEMOLIZ. INGOMBRANTI ALTRO



Inoltre, sono state considerati anche altre 4 variabili, che fanno riferimento alla quantità di rifiuti provenienti dalla pulizia stradale e dalla costruzione e demolizione, ai rifiuti ingombranti e ad altre categorie di rifiuti, sempre in valori pro-capite.

INDIFFERENZIATO



Infine, si è osservata anche la variabile relativa alla quantità di indifferenziato pro-capite raccolto nelle varie regioni. Si può notare come la mappa presenti una colorazione inversa rispetto alle precedenti, in quanto le regioni ad avere colori più intensi sono quelle centro-meridionali, che evidentemente mantengono un basso livello di differenziazione dei rifiuti.

3. Regioni omogenee per raccolta pro-capite

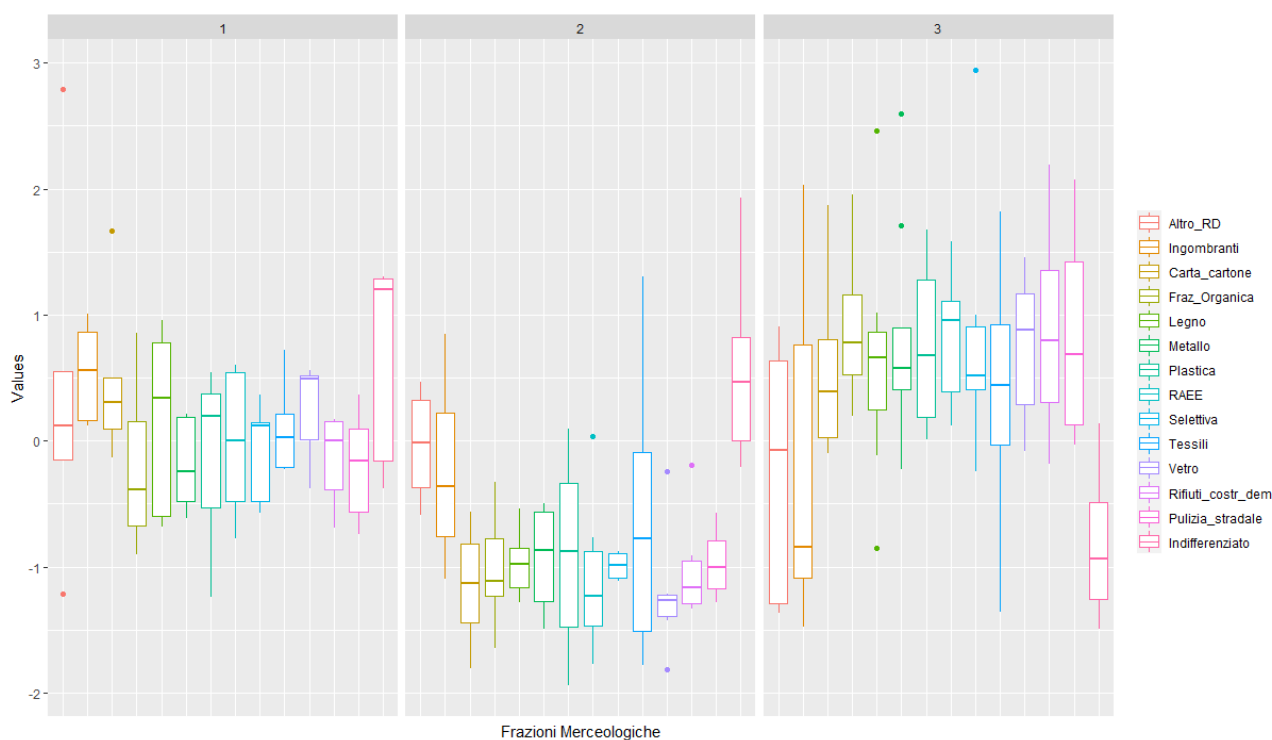
3.1 Analisi cluster

Sulla base delle 14 variabili precedentemente descritte è stata effettuata una analisi cluster, in modo da classificare le regioni in gruppi omogenei.

Tramite il metodo del legame completo sono stati costituiti 3 gruppi:

- Gruppo 1: Abruzzo, Lazio, Liguria, Toscana e Piemonte.
- Gruppo 2: Basilicata, Calabria, Campania, Molise, Puglia e Sicilia.
- Gruppo 3: Emilia-Romagna, Friuli-Venezia Giulia, Lombardia, Marche, Sardegna, Trentino-Alto Adige, Umbria e Veneto.

Per mancanza di dati, non è stato possibile considerare la regione Valle d'Aosta.



Dai boxplot relativi alle variabili osservate all'interno dei gruppi si possono trarre informazioni interessanti riguardo la classificazione delle regioni.

Il Gruppo 3 è formato da regioni che presentano valori più alti in relazione alle variabili che si riferiscono a categorie di rifiuti differenziati, mentre valori più bassi per quanto riguarda l'indifferenziato.

Il Gruppo 2, al contrario, è formato da regioni con valori più bassi per le variabili della raccolta differenziata e valori più alti per l'indifferenziato.

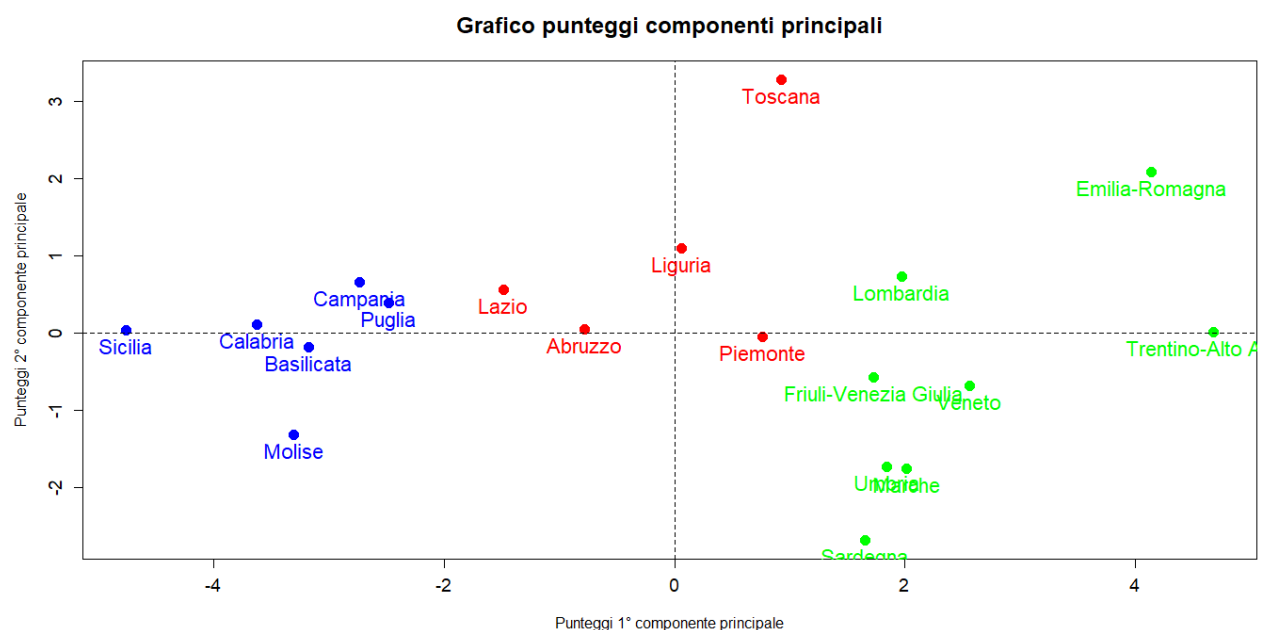
Infine, il Gruppo 1 è formato da regioni con performance intermedie.

Si nota come, nel Gruppo 3 siano state inserite le regioni italiane del Nord-Est (più la Sardegna) mentre nel Gruppo 2 ci siano tutte le regioni meridionali. Viene, quindi, confermata l'analisi esplorativa effettuata in precedenza osservando le mappe.

3.2 Analisi delle componenti principali

Successivamente, è stata effettuata una analisi delle componenti principali per provare a sintetizzare l'insieme di variabili considerato.

Le prime due componenti principali arrivano a spiegare complessivamente il 68% della varianza totale.



Le regioni sono state rappresentate nel grafico sulla base dei punteggi ottenuti nelle prime due componenti.

In particolare, la prima componente principale, che spiega il 55% della variabilità complessiva, contrappone le variabili relative alle quantità pro-capite di raccolta differenziata con la quantità pro-capite di indifferenziato.

Infatti, sull'asse orizzontale si distinguono con chiarezza i tre gruppi precedentemente ottenuti con l'analisi cluster. Partendo da sinistra, individuiamo il Gruppo 2 in blu, formato dalle regioni che fanno meno raccolta differenziata; il Gruppo 1 in rosso in una posizione intermedia e il Gruppo 3 in verde formato dalle regioni che riescono a differenziare una maggiore quantità di rifiuti.

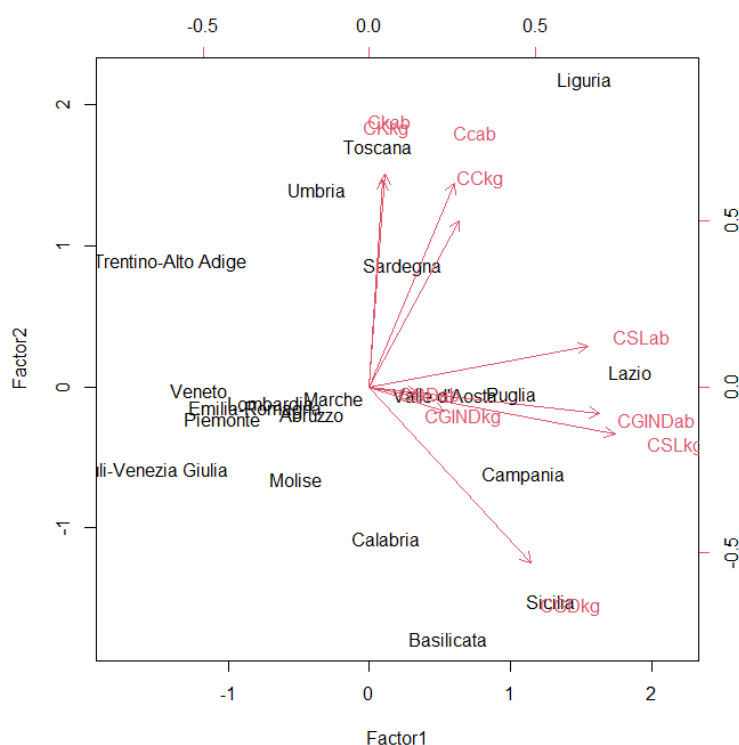
Infine, la seconda componente principale, che spiega appena il 13% della variabilità complessiva, è influenzata positivamente dalla variabile relativa ai rifiuti ingombranti e quella relativa alle altre categorie di rifiuti.

4. Costi di gestione dei rifiuti urbani

4.1 Analisi fattoriale

Una ulteriore analisi è stata effettuata riguardo i diversi costi di gestione dei rifiuti urbani. In particolare, le variabili considerate, sia pro-capite, sia per kg, sono: costi di gestione dell'indifferenziato, costi di gestione della raccolta differenziata, costi di spazzamento e lavaggio urbano, costi comuni e costi del capitale.

Tramite un'analisi fattoriale con rotazione ortogonale sono stati ottenuti 6 fattori comuni e sono stati stimati i punteggi ottenuti dalle regioni attraverso il metodo di Bartlett.



Nel biplot qui riportato vengono considerati solamente i primi due fattori comuni che arrivano a spiegare il 57% della varianza totale.

Per quanto riguarda il primo fattore, le variabili che maggiormente lo influenzano positivamente, e che si mostrano anche abbastanza correlate tra loro, sono: i costi di

spazzamento e lavaggio urbano pro-capite e per kg e i costi di gestione dell'indifferenziato pro-capite.

Nella parte sinistra del grafico, quindi, troviamo quelle regioni che avranno bassi costi di gestione in riferimento a quest'ultime variabili.

Per il secondo fattore, invece, vengono contrapposti i costi comuni e i costi del capitale al costo di gestione della raccolta differenziata per kg.

Nella parte inferiore del grafico troviamo regioni come Sicilia, Basilicata, Calabria e Campania che sono quindi caratterizzate da bassi costi del capitale (quindi bassi costi di gestione di impianti, macchinari e stabilimenti dediti alla gestione dei rifiuti) e da un elevato costo di gestione della raccolta differenziata.

Una interpretazione possibile di questo risultato è che quest'ultime regioni effettuino scarsi investimenti in questo settore e che debbano poi affrontare costi eccessivi per la gestione dei singoli rifiuti differenziati. Questa può essere una parziale motivazione alle basse percentuali di raccolta differenziata che vengono registrate proprio in queste regioni.

l'asse troviamo anche i costi di gestione dell'indifferenziato per abitante, i costi di spazzamento e lavaggio e i costi di gestione della raccolta differenziata per kg.

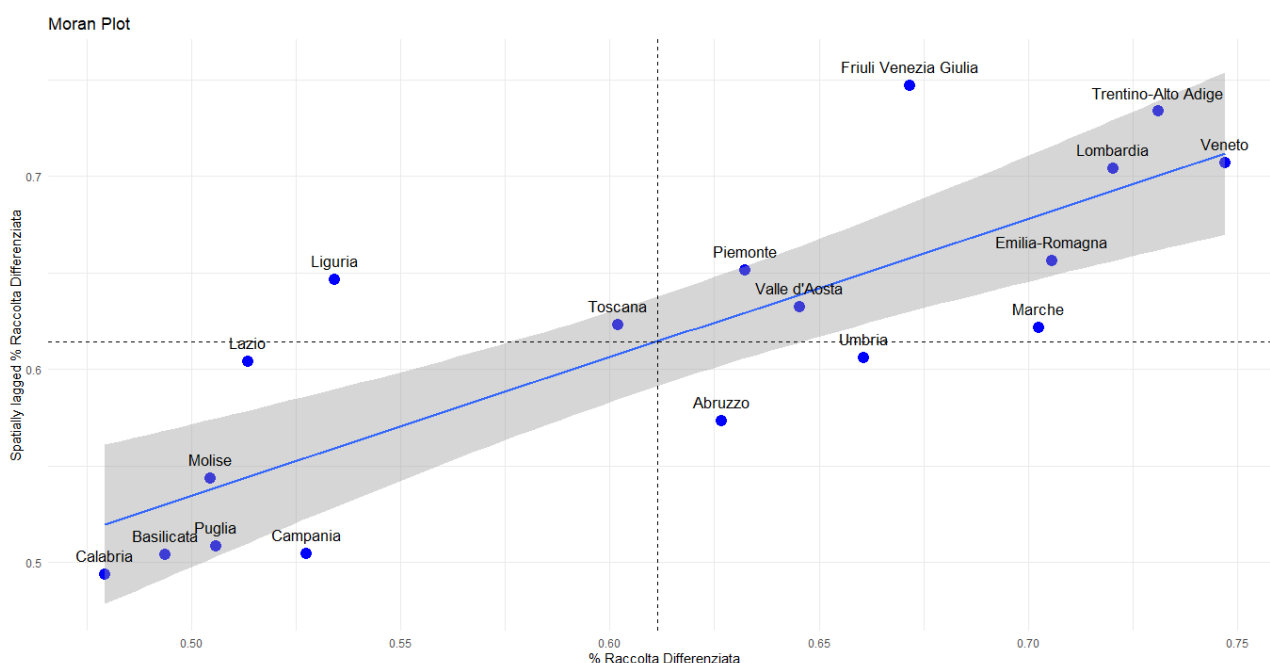
Si può interpretare questo risultato mettendo in evidenza come, le regioni che riescono ad avere buone percentuali di raccolta differenziata si trovino a dover sostenere costi inferiori sia per i rifiuti urbani indifferenziati per abitante sia di raccolta differenziata per kg.

In pratica, per regioni come, ad esempio, Sicilia, Calabria, Campania e Puglia, un kg di raccolta differenziata costa molto più che a regioni come Emilia-Romagna, Trentino-Alto Adige, Lombardia e Marche.

6. Analisi della dipendenza della percentuale di raccolta differenziata

6.1 Correlazione spaziale

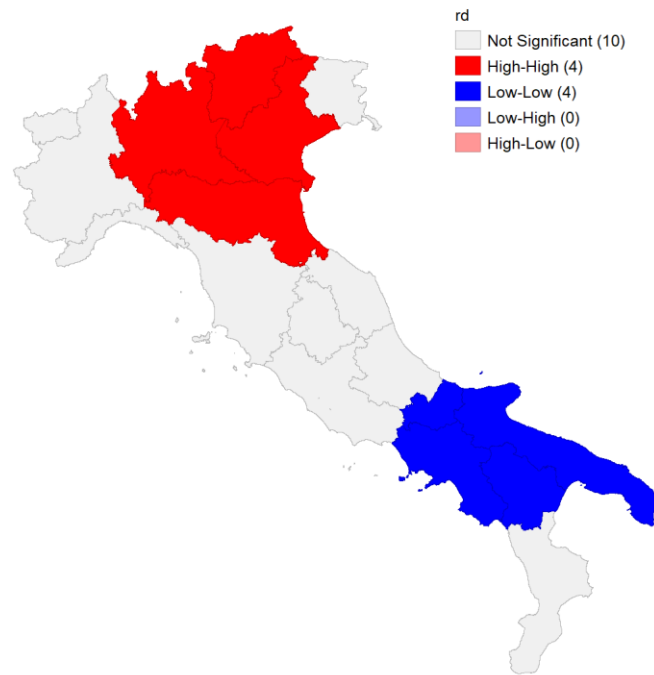
Trattandosi di dati territoriali, si è analizzato se la percentuale di raccolta differenziata di ogni regione fosse correlata con la media delle percentuali delle regioni confinanti.



Tramite l'Indice I di Moran, pari a 0.718, è stata individuata una correlazione positiva. Perciò, regioni con elevata percentuale di raccolta differenziata saranno confinanti con altre regioni dalla percentuale elevata e viceversa.

Inoltre, tramite il Local Moran's Test si sono evidenziate delle particolari zone di influenza dove questa correlazione positiva risulta più marcata.

Nelle zone in rosso troviamo regioni con percentuali elevate all'interno ed elevate nelle confinanti, mentre la zona in blu evidenzia regioni con basse percentuali interne e basse anche tra le regioni confinanti.



Poiché per lo studio della dipendenza spaziale è necessario che i territori siano confinanti, per queste analisi non sono state considerate le isole Sicilia e Sardegna.

La presenza di una effettiva correlazione spaziale ci permette di proseguire lo studio con un'analisi della regressione spaziale.

6.2 Regressione spaziale

Il modello di regressione spaziale utilizzato è lo Spatial Error Model (SEM).

$$y = X\beta + u, \quad u = \lambda W + \varepsilon$$

In questo modello si presuppone che i residui possano essere espressi tramite una combinazione lineare di un termine spaziale λ con pesi relativi alle regioni confinanti contenuti nella matrice W più una componente stocastica ε .

Nel modello utilizzato, la variabile dipendente è la percentuale di raccolta differenziata regionale, mentre come variabili indipendenti sono stati considerati i valori dell'età media regionale e del tasso di occupazione regionale, entrambi forniti dall'Istat.

Di seguito, i risultati principali ottenuti con il modello:

- I residui non sono molto elevati e mostrano un andamento abbastanza simmetrico.

Min	1° Q	Mediana	3°Q	Max
-0,1171	-0,0227	0,0037	0,0283	0,057

- I coefficienti e l'intercetta risultano tutti significativi, dati i bassi valori ottenuti dai p-value. In particolare, l'età media sembra influenzare negativamente la percentuale di raccolta differenziata, mentre il tasso di occupazione la influenza positivamente.

	Stima	Std. Error	Z Value	Pr(> z)
<i>Intercetta</i>	1,2357	0,3116	3,966	0,00007
<i>Età</i>	-0,027	0,0073	-3,718	0,0002
<i>Occupazione</i>	0,0098	0,0009	11,095	< 2,2e-16

- Il valore dell'effetto spaziale stimato $\hat{\lambda}$ è pari a -0,384 ma osservando i risultati dei test effettuati (Likelihood Ratio Test, Test Z e Test di Wald) non risulta essere significativo.

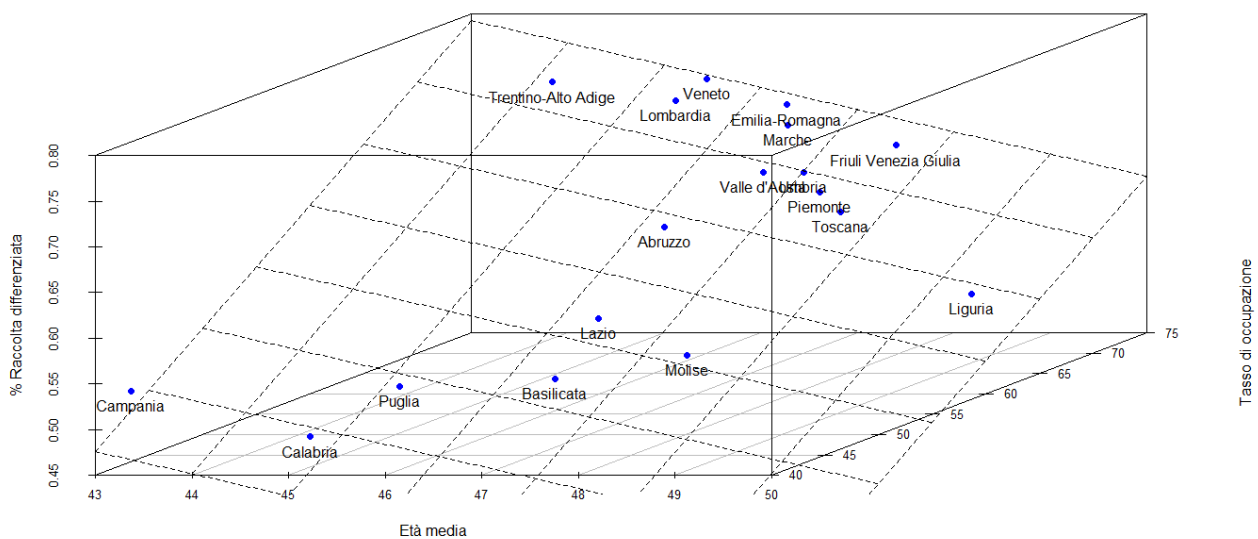
	$\hat{\lambda}$
<i>Lambda</i>	-0,384

	Test value	P-value
<i>LR Test</i>	<i>1,112</i>	<i>0,292</i>
<i>Test Z</i>	<i>-1,364</i>	<i>0,173</i>
<i>Wald</i>	<i>1,86</i>	<i>0,173</i>

La mappa dei valori dei residui, infatti, ha un aspetto del tutto casuale e non sembrano in nessun modo essere influenzati dalla posizione delle regioni.



Dunque, la variabilità spaziale della percentuale di raccolta differenziata risulta essere ben rappresentata dalle variabili antecedenti scelte: età media e tasso di occupazione.



Osservando il grafico tridimensionale risulta chiaro come la percentuale di raccolta differenziata aumenti al diminuire dell'età media e all'aumentare del tasso di occupazione.

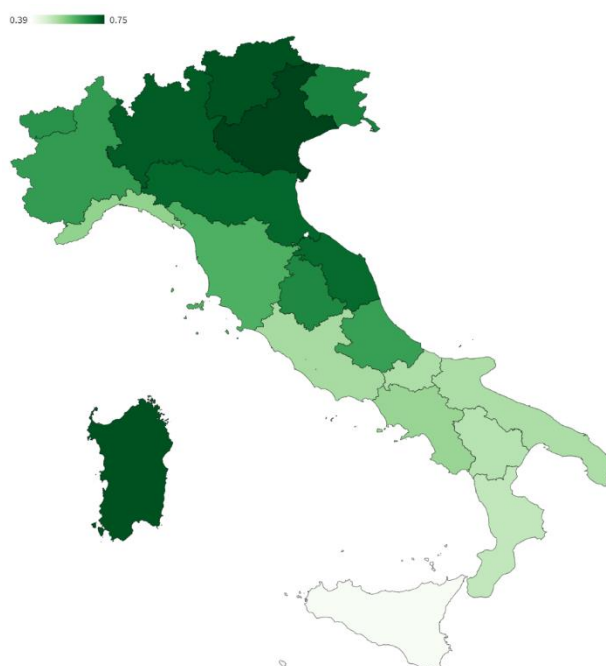
La relazione negativa tra età media regionale e raccolta differenziata può essere dovuta al fatto che soltanto le generazioni più giovani hanno avuto una maggiore educazione ambientale e sul riciclo dei rifiuti, mentre c'è il bisogno di far comprendere al meglio l'importanza di queste tematiche anche alle classi di età più elevate.

La variabile del tasso di occupazione racchiude in sé l'informazione di tanti indicatori di benessere sociale ed economico di un dato territorio, poiché è correlata, ad esempio, con il reddito medio regionale e con il livello di istruzione. Si può, quindi, affermare che la percentuale di raccolta differenziata sia una variabile legata strettamente alla situazione socio-economica di una regione.

7. Conclusioni

Nonostante l'ottava posizione raggiunta nella classifica stilata dall'Eurostat delle nazioni europee che riciclano più rifiuti, il nostro paese ha ancora della strada da fare davanti a sé ed esistono margini di miglioramento importanti.

L'obiettivo nazionale era raggiungere il 65% totale di rifiuti urbani differenziati entro il 2012, ma nel 2019 il livello raggiunto era del 61,28%.



Osservando la mappa della percentuale regionale di raccolta differenziata, si nota subito un paese sostanzialmente diviso in due. Il primo passo, quindi, sarà provare a eliminare le croniche differenze presenti tra Nord e Sud Italia che colpiscono anche questo fenomeno.

In un mondo che punta ad essere sempre più “green” ed eco-sostenibile fare la raccolta differenziata è un'azione semplice, alla portata di tutti e che ci può aiutare moltissimo nel ridurre il nostro impatto ambientale.

8. Riferimenti bibliografici

1. González, Ignacio, Déjean, Sébastien, Martin, Pascal, AND Baccini, Alain. "CCA: An R Package to Extend Canonical Correlation Analysis" *Journal of Statistical Software [Online]*, Volume 23, Issue 12 (17 January 2008).
2. Bivand, Roger, AND Piras, Gianfranco. "Comparing Implementations of Estimation Methods for Spatial Econometrics" *Journal of Statistical Software [Online]*, Volume 63, Issue 18 (16 February 2015).
3. Anselin, Luc. 1996. "The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association." In *Spatial Analytical Perspectives on Gis in Environmental and Socio-Economic Sciences*, edited by Manfred Fischer, Henk Scholten, and David Unwin, 111–25. London: Taylor; Francis.
4. Anselin, Luc. 2019. "A Local Indicator of Multivariate Spatial Association, Extending Geary's c ." *Geographical Analysis* 51 (2): 133–50.

