

# Progetto sul dominio “AMBIENTE”

Birgauan Floarea

## INTRODUZIONE

Nel 2010 L' ISTAT, insieme ai rappresentanti delle parti sociali e della società civile, creò il progetto Bes (Benessere equo e sostenibile) allo scopo di misurare, non solo attraverso il Pil, il benessere e il progresso della società in diversi aspetti, come la qualità della vita e le condizioni materiali. Sono stati individuati 12 domini fondamentali e selezionati all'incirca 130 indicatori.

## Dominio d'Analisi: “Ambiente”

Comprendere i complessi processi che regolano il mondo e l'**ambiente** circostante ci aiuta a capire il ruolo e la responsabilità che ciascun individuo ha nei confronti del pianeta.

La qualità dell'ambiente influenza in maniera diretta e indiretta la qualità della vita degli individui e per definire in che misura, sono stati individuati degli indicatori, raggruppati in 6 dimensioni:

- ▶ Qualità dell'aria
- ▶ Qualità delle acque e delle risorse idriche
- ▶ Biodiversità
- ▶ Naturalità degli ecosistemi marini e terrestri
- ▶ Qualità del suolo
- ▶ Dimensione trasversale della valutazione soggettiva della situazione ambientale attraverso tre indicatori:
  1. Soddisfazione per la situazione ambientale
  2. Preoccupazione per la perdita di biodiversità
  3. Preoccupazione per i cambiamenti climatici

**Gli effetti dei cambiamenti climatici e dell'aumento dell'effetto serra rappresentano uno dei problemi climatici che preoccupano maggiormente le persone.** La sensibilizzazione su questo argomento è alta presso i cittadini di tutte le età, giovani compresi.

Secondo il Rapporto Bes, tale preoccupazione cresce in modo costante dal 58,7% del 2014 fino a oltre il 70% negli ultimi due anni. Nel 2020 nel nord si registra il 72,2%, nel Mezzogiorno il 67,5%, ma fa eccezione il Molise, che è la regione del Sud dove si rileva il maggiore numero di persone preoccupate per l'effetto serra e i cambiamenti climatici (77,4%).

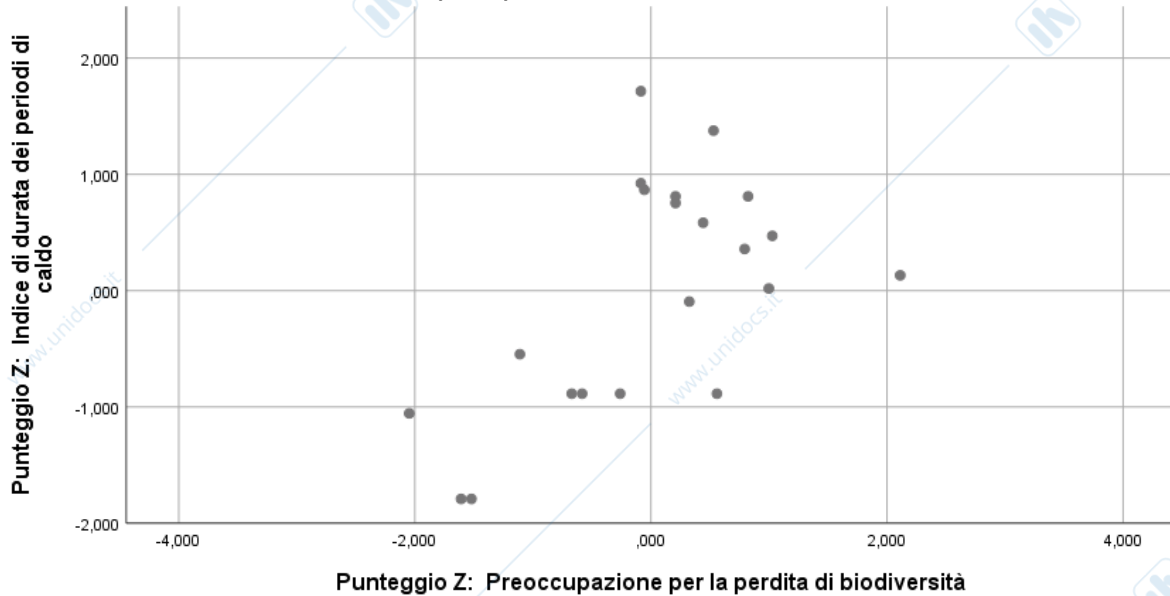
Uno degli effetti negativi dei cambiamenti climatici è l'aumento della durata dei periodi di caldo causato dall'effetto serra.

L'intensità dei giorni di caldo negli ultimi dieci anni risulta sempre maggiore rispetto alla mediana del periodo di riferimento 1981-2010, infatti nel 2020 si conferma il graduale aumento delle temperature minime e massime, più elevate nel Centro e nel Nord. Questo aumento provoca una forte preoccupazione per la popolazione non solo per i cambiamenti climatici, ma anche per l'effetto che questo aumento può provocare sull'ambiente, cioè la perdita della biodiversità. La preoccupazione per la scomparsa di specie animale e vegetale si concentra soprattutto nelle regioni del Nord e Centro.

Grazie al seguente grafico di dispersione è possibile visualizzare meglio questa relazione tra l'indicatore "Indice di durata dei periodi di caldo" e "la preoccupazione per la perdita della biodiversità".

*Coefficiente di Pearson: 0,598*

Dispersione semplice di Punteggio Z: Indice di durata dei periodi di caldo per Punteggio Z: Preoccupazione per la perdita di biodiversità



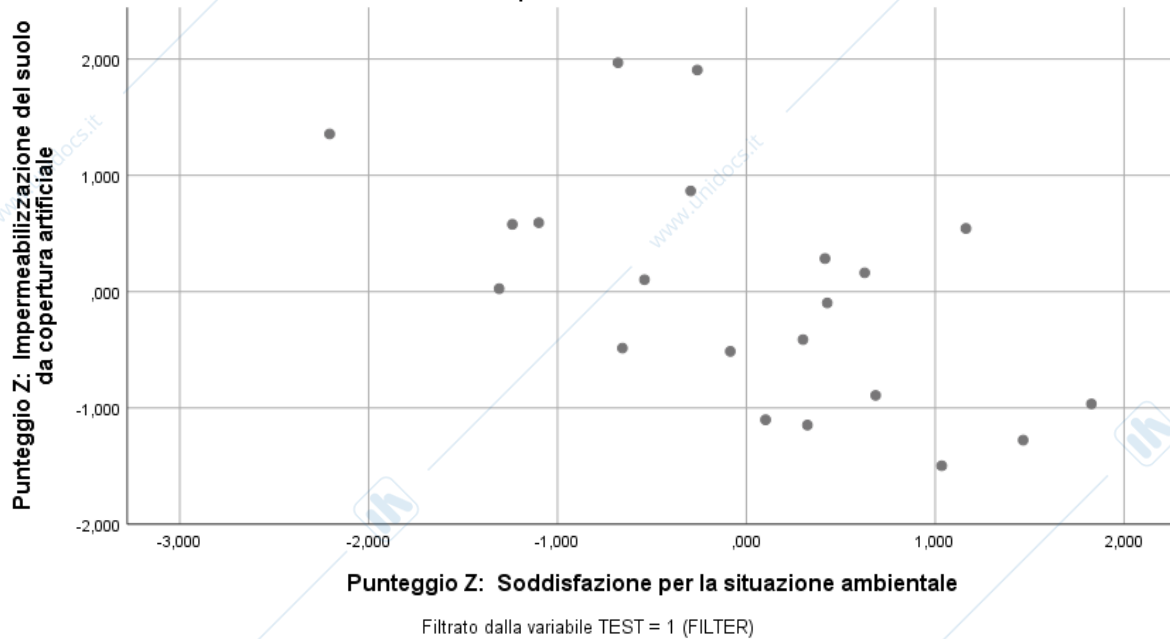
Filtrato dalla variabile TEST = 1 (FILTER)

Gli indicatori legati al governo del territorio sono in netto peggioramento: il consumo di suolo continua ad avanzare, soprattutto al Nord, infatti le regioni con maggior quota di suolo consumato sono Lombardia, Veneto e fa eccezione la Campania.

L'espansione di copertura artificiali sul suolo naturale riduce la permeabilità e lo sviluppo funzionale del terreno. Tale fenomeno si può considerare praticamente irreversibile, nel breve periodo, per questo motivo la copertura del suolo provoca delle conseguenze sulla qualità dell'ambiente.

*Coefficiente di Pearson: -0.588*

**Dispersione semplice di Punteggio Z: Impermeabilizzazione del suolo da copertura artificiale per Punteggio Z: Soddifazione per la situazione ambientale**

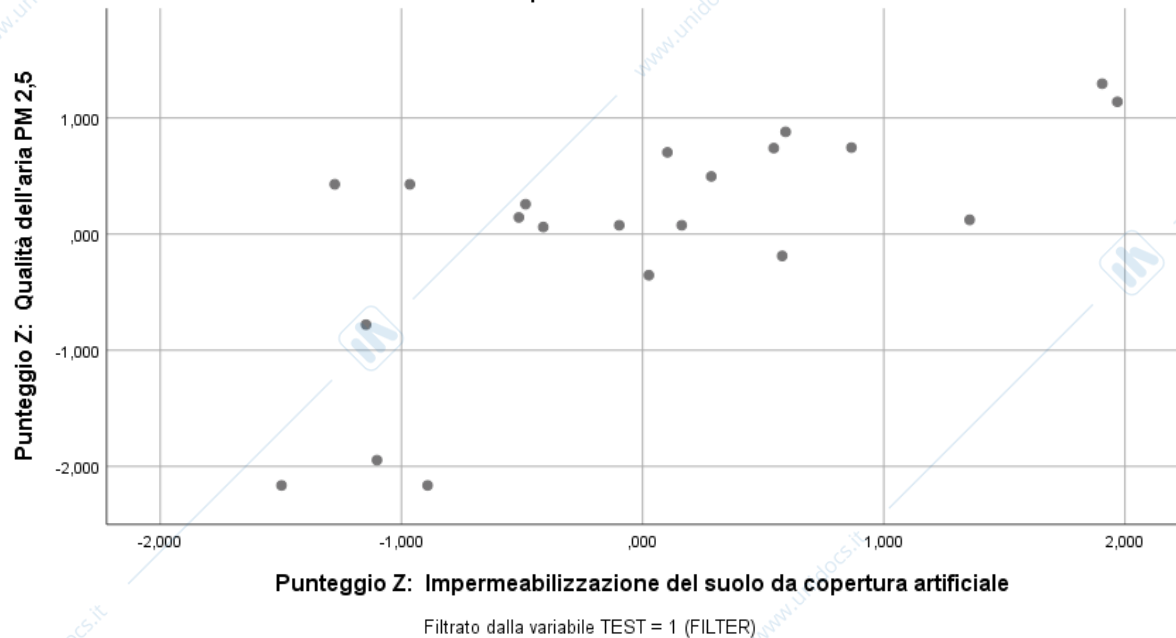


La forma più evidente e più diffusa di **copertura artificiale** è la **costruzione di edifici e strade con materiali artificiali** (quali asfalto o calcestruzzo). L'aumento di copertura implica un incremento dell'inquinamento dovuto ai processi di combustione negli impianti di riscaldamento, nelle attività industriali e il traffico veicolare. Questi processi producono un'emissione di particelle atmosferiche chiamate PM<sub>2,5</sub> che sono molto dannose per l'uomo perché l'esposizione prolungata nel tempo, anche a bassi livelli di queste particelle, è associata all'aumento di disturbi respiratori insieme ad effetti sul sistema cardio vascolare.

Oms, Organizzazione mondiale della sanità, ha imposto dei valori soglia per la salute ma in Italia, dal 2010, questa soglia è stata superata del 80%. L'indicatore raggiunge le percentuali più alte nelle regioni del Nord al contrario del Centro e Sud si osserva un miglioramento.

*Coefficiente di Pearson: 0.678*

Dispersione semplice di Punteggio Z: Qualità dell'aria PM 2,5 per Punteggio Z: Impermeabilizzazione del suolo da copertura artificiale



Gli indicatori sull'ambiente sono 21 però quelli trattati sono 20 dato che per l'indicatore "Emissioni di CO2 e altri gas clima alteranti" era a disposizione solo il dato "Italia".

1. **Consumo materiale interno:**  
Quantità di materiali trasformati in emissioni, rifiuti o nuovi stock del sistema antropico (in milioni di tonnellate).
2. **Dispersione da rete idrica comunale:**  
Perdite idriche totali nelle reti comunali di distribuzione dell'acqua potabile (percentuale del volume complessivo immesso in rete).
3. **Conferimento dei rifiuti urbani in discarica:**  
Percentuale di rifiuti urbani conferiti in discarica sul totale dei rifiuti urbani raccolti.
4. **Qualità dell'aria urbana - PM2,5:**  
Percentuale di misurazioni valide superiori al valore di riferimento per la salute, definito dall'Oms ( $10 \mu\text{g}/\text{m}^3$ ), sul totale delle misurazioni valide delle concentrazioni medie annuali di PM2,5 per tutte le tipologie di stazione (traffico urbano e suburbano, fondo urbano e suburbano, rurale).
5. **Coste marine balneabili:**  
Percentuale di coste balneabili autorizzate sul totale della linea litoranea ai sensi delle norme vigenti (l'indicatore tiene conto dei tratti di costa stabilmente interdetti alla balneazione a norma di legge e di quelli interdetti stagionalmente per livelli di contaminanti oltre le soglie di rischio per la salute).
6. **Disponibilità di verde urbano:**  
Metri quadrati di verde urbano per abitante.
7. **Soddisfazione per la situazione ambientale:**  
Persone molto o abbastanza soddisfatte della situazione ambientale (aria, acqua, rumore) della zona in cui vivono per 100 persone di 14 anni e più.
8. **Preoccupazione per la perdita di biodiversità:**  
Persone che ritengono l'estinzione di specie vegetali/ animali tra le 5 preoccupazioni ambientali prioritarie per 100 persone di 14 anni e più.
9. **Preoccupazione per i cambiamenti climatici:**  
Percentuale di persone di 14 anni e più che ritengono il cambiamento climatico o l'aumento dell'effetto serra e il buco dell'ozono tra le 5 preoccupazioni ambientali prioritarie per 100 persone di 14 anni e più.
10. **Siti contaminati:**  
Incidenza dei siti di interesse nazionale (Sin) e dei siti di competenza delle Regioni sulla superficie territoriale, valori per 1.000.
11. **Popolazione esposta al rischio di frane:**  
Percentuale della popolazione residente in aree con pericolosità da frane elevata e molto elevata sul totale della popolazione residente.
12. **Popolazione esposta al rischio di alluvioni:**  
Percentuale della popolazione residente in aree a pericolosità idraulica media (tempo di ritorno 100-200 anni ex D. Lgs. 49/2010).
13. **Trattamento delle acque reflue:**  
Quota percentuale dei carichi inquinanti confluiti in impianti secondari o avanzati, in abitanti equivalenti, rispetto ai carichi complessivi urbani (Aetu) generati.
14. **Aree protette:**  
Percentuale di superficie territoriale coperta da aree naturali protette terrestri incluse nell'elenco ufficiale delle aree protette (Euap) e nella Rete Natura 2000.

**15. Energia elettrica da fonti rinnovabili:**

Percentuale di consumi di energia elettrica coperti da fonti rinnovabili sul totale dei consumi interni lordi.

**16. Impermeabilizzazione del suolo da copertura artificiale:**

Percentuale di suolo impermeabilizzato sul totale della superficie territoriale.

**17. Raccolta dei rifiuti urbani:**

Rifiuti urbani raccolti per abitante.

**18. Indice di durata dei periodi di caldo:**

Numero di giorni nell'anno in cui la temperatura massima è superiore al 90° percentile della distribuzione nel periodo climatologico di riferimento (1981-2010), per almeno sei giorni consecutivi. A livello territoriale (regione, ripartizione geografica, Italia) tale indicatore corrisponde alla mediana dei valori di tutte le celle del gridded dataset che ricadono all'interno dell'area analizzata.

**19. Giorni con precipitazione estremamente intensa:**

Numero di giorni dell'anno in cui la cumulata giornaliera delle precipitazioni supera o è uguale al valore di 50 mm. A livello territoriale (regione, ripartizione geografica, Italia) tale indicatore corrisponde alla mediana dei valori di tutte le celle del gridded dataset che ricadono all'interno dell'area analizzata.

**20. Giorni consecutivi senza pioggia:**

Numero massimo di giorni consecutivi nell'anno con precipitazione giornaliera inferiore o uguale a 1 mm. A livello territoriale (regione, ripartizione geografica, Italia) tale indicatore corrisponde alla mediana dei valori di tutte le celle del gridded dataset che ricadono all'interno dell'area analizzata.

## Preparazione dei dati

Gli indicatori sono variabili con unità di misura o con unità di grandezza (variabilità) diverse e per la loro sintetizzazione si deve procedere con la trasformazione delle variabili, in modo da essere riportate a un'unità di misura comune per poter essere confrontabili, ovvero si procede con la **normalizzazione**.

In particolare, il processo di normalizzazione si basa sulla standardizzazione, cioè le variabili vengono trasformate in variabili normalizzate con media pari allo 0 e varianza pari a 1.

Un altro modo di trasformare le variabili quantitative è attraverso gli **Numeri Indice**, cioè si sceglie una unità come base (nel nostro caso abbiamo scelto il dato dell'Italia) e si ripropongono i valori delle variabili rispetto a questa base. Infatti, tutti i valori inferiori a 100 si trovano in una situazione peggiore rispetto alla base e i valori superiori in una situazione migliore.

Ma dato che la base può variare nel tempo, essendo una base mobile, e considerando che la variabilità è molto elevata, la scelta migliore è la standardizzazione, in quanto ha variabilità standard, cioè pari ad 1.

Lo stesso ragionamento vale anche per i numeri puri.

Statistiche descrittive

	N	Minimo	Massimo	Media	Deviazione std.
Coste marine balneabili	21	42,2	90,8	65,667	11,4642
Consumo materiale interno	21	0,5	240,2	45,233	67,4362
Dispersione da rete idrica comunale	21	22,1	55,6	41,667	9,3503
Conferimento dei rifiuti urbani in discarica	21	1,3	90,0	28,248	21,2117
Qualità dell'aria PM 2,5	21	33,3	100,0	75,029	19,2841
Disponibilità di verde urbano	21	7,2	555,5	74,176	138,9412
Soddisfazione per la situazione ambientale	21	56,4	91,0	75,329	8,5778
Preoccupazione per la perdita di biodiversità	21	17,5	31,7	24,490	3,4141
Preoccupazione per i cambiamenti climatici	21	62,2	77,4	69,971	3,5161
Siti contaminati	21	0,1	42,6	5,986	9,6646
Popolazione esposta al rischio di frane	21	0,1	12,1	3,252	2,8675
Popolazione esposta al rischio di alluvioni	21	0,1	63,7	8,905	13,9182
Trattamento delle acque reflue	21	43,9	99,7	61,486	12,0516
Aree protette	21	12,2	36,6	23,319	6,4592
Energia elettrica da fonti rinnovabili	21	8,5	297,2	63,110	68,2709
Impermeabilizzazione del suolo da copertura artificiale	21	2,2	12,1	6,427	2,8566
Rifiuti urbani raccolti	21	355,0	664,0	495,667	74,4871
Indice di durata dei periodi di caldo	21	0,0	31,0	15,833	8,8435
Giorni con precipitazione estremamente intensa	21	0	4	1,14	1,195
Giorni consecutivi senza pioggia	21	19,0	47,0	26,429	6,9125
Numero di casi validi (listwise)	21				

## La matrice di correlazione

Considerando la matrice di correlazione tra tutte le variabili standardizzate, si nota che i coefficienti di correlazione hanno valori molto bassi e questo significa che le variabili sono poco correlate fra loro. Questo comporta un'osservazione a priori, cioè che attraverso l'**analisi in componenti principali** sarà difficile trovare delle variabili che possano riassumere adeguatamente degli indici scarsamente correlati.

La maggior parte dei coefficienti di correlazione di Pearson ha valori inferiori al 0,4, addirittura osserviamo dei valori intorno allo zero e ciò implica che gli indicatori tendono ad essere incorrelati fra loro. Ad esempio, i casi più critici sono evidenziati in rosso, viceversa i valori più alti sono evidenziati in verde.

Matrice di correlazione\*

	Punteggio Z: Consumo energetico interno	Punteggio Z: Dispersione da rete idrica comunale	Punteggio Z: Conferimento dei rifiuti urbani in discarica	Punteggio Z: Coste marine balneabili	Punteggio Z: Qualità dell'aria PM2.5	Punteggio Z: Disponibilità di verde urbano	Punteggio Z: Soddisfazione per la situazione ambientale	Punteggio Z: Preoccupazione per la perdita di biodiversità	Punteggio Z: Preoccupazione per i cambiamenti climatici	Punteggio Z: Siti contaminati	Punteggio Z: Popolazione esposta al rischio di frana	Punteggio Z: Popolazione esposta al rischio di alluvioni	Punteggio Z: Trattamento delle acque reflue	Punteggio Z: Aree protette	Punteggio Z: Energia elettrica da fonti rinnovabili	Punteggio Z: Inquinamento del suolo da copertura artificiale	Punteggio Z: Rifiuti urbani raccolti	Punteggio Z: Indice di durata dei periodi di caldo	Punteggio Z: Giorni con precipitazione estremamente intensa	Punteggio Z: Giorni consecutivi senza pioggia
Correlazione	1,000	-0,441	-0,386	-0,330	0,285	0,263	0,409	0,376	0,070	-0,113	-0,285	-0,115	0,554	0,034	0,297	-0,175	0,149	0,197	0,445	-0,097
Punteggio Z: Consumo energetico interno		1,000	0,300	0,355	-0,105	-0,003	-0,398	-0,325	-0,052	-0,007	-0,198	-0,253	-0,337	0,184	-0,552	0,020	-0,480	-0,407	-0,224	0,409
Punteggio Z: Dispersione da rete idrica comunale	-0,441		1,000	0,282	-0,567	-0,084	0,099	-0,274	0,153	-0,319	0,384	-0,198	-0,379	0,117	0,094	-0,401	-0,348	-0,383	-0,300	
Punteggio Z: Conferimento dei rifiuti urbani in discarica	-0,386	0,300		1,000	0,282	-0,567	-0,084	0,099	-0,274	0,153	-0,319	0,384	-0,198	-0,379	0,117	0,094	-0,401	-0,348	-0,383	
Punteggio Z: Coste marine balneabili	-0,330	0,355	0,282		1,000	-0,182	0,318	-0,235	-0,552	-0,229	0,332	0,282	-0,109	-0,082	0,222	-0,051	-0,189	0,437	0,487	-0,416
Punteggio Z: Qualità dell'aria PM2.5	0,285	-0,105	-0,567	-0,182		1,000	-0,069	-0,201	0,094	-0,094	0,195	-0,615	0,183	0,042	-0,235	-0,476	0,678	0,247	0,062	0,154
Punteggio Z: Disponibilità di verde urbano	0,263	0,263	-0,084	0,318	-0,069		1,000	0,339	-0,398	0,102	-0,114	0,108	-0,176	0,097	0,040	0,158	-0,358	-0,301	-0,057	0,311
Punteggio Z: Soddisfazione per la situazione ambientale	0,409	-0,398	0,099	-0,235	-0,201	0,339		1,000	0,453	0,212	-0,155	0,194	0,003	0,235	-0,089	0,504	-0,588	0,144	0,447	0,518
Punteggio Z: Preoccupazione per la perdita di biodiversità	0,376	-0,325	-0,274	-0,552	0,094	-0,398	0,453		1,000	0,369	0,162	-0,203	0,351	0,468	-0,240	0,072	-0,002	0,455	0,598	0,201
Punteggio Z: Preoccupazione per i cambiamenti climatici	0,070	-0,052	0,153	-0,229	-0,094	0,102	0,212	0,369		1,000	0,011	-0,120	0,341	-0,074	-0,274	-0,268	0,121	0,123	0,134	0,061
Punteggio Z: Siti contaminati	-0,113	-0,007	-0,319	-0,332	0,195	-0,114	-0,155	0,162	0,011		1,000	-0,294	-0,083	0,044	-0,318	-0,233	0,156	-0,021	0,209	0,052
Punteggio Z: Popolazione esposta al rischio di frana	-0,285	-0,198	0,384	0,282	-0,615	0,108	0,194	-0,203	-0,120	-0,294		1,000	0,008	0,039	0,555	0,635	-0,504	-0,055	0,041	0,087
Punteggio Z: Popolazione esposta al rischio di alluvioni	-0,115	-0,253	-0,198	-0,109	-0,183	-0,176	-0,398	0,351	0,341	-0,083	0,008		1,000	0,003	-0,433	-0,175	0,227	0,673	0,244	-0,255
Punteggio Z: Trattamento delle acque reflue	0,554	-0,337	-0,379	-0,082	0,042	0,097	0,235	0,468	-0,074	0,044	0,039	0,003		1,000	0,108	0,442	-0,291	0,106	0,211	
Punteggio Z: Aree protette	0,034	0,184	0,117	0,222	-0,235	0,040	-0,089	-0,240	-0,274	-0,318	0,555	-0,433	0,108		1,000	0,321	-0,218	-0,395	-0,236	
Punteggio Z: Energia elettrica da fonti rinnovabili	0,297	-0,552	0,094	-0,051	-0,476	0,158	0,504	0,072	-0,268	-0,233	0,635	-0,175	0,442	0,321		1,000	-0,654	0,078	0,239	
Punteggio Z: Inquinamento del suolo da copertura artificiale	-0,175	0,020	-0,401	-0,118	0,678	-0,351	-0,588	-0,002	0,121	0,156	-0,504	0,227	-0,291	-0,218	-0,654		1,000	0,229	0,051	
Punteggio Z: Rifiuti urbani raccolti	0,149	-0,480	-0,348	-0,427	0,247	-0,301	0,144	0,455	0,123	-0,021	-0,065	0,673	0,106	-0,395	0,078	0,229		1,000	0,639	
Punteggio Z: Indice di durata dei periodi di caldo	0,197	-0,407	-0,383	-0,497	0,062	-0,057	0,447	0,598	0,134	0,209	0,041	0,244	0,211	-0,236	0,239	0,051	0,639		1,000	
Punteggio Z: Giorni con precipitazione estremamente intensa	0,445	-0,224	-0,302	-0,416	0,154	0,311	0,518	0,201	0,061	0,062	0,087	-0,255	0,111	0,390	0,333	-0,112	-0,049	0,322		
Punteggio Z: Giorni consecutivi senza pioggia	-0,097	0,409	-0,001	-0,084	-0,209	0,055	-0,347	-0,259	-0,070	0,149	-0,354	-0,065	-0,354	-0,255	-0,310	-0,024	-0,199	-0,219	-0,316	

\* Nella fase di analisi vengono utilizzati solo i casi per cui TEST1 = 1 (FILTER) = Selected

## Analisi fattoriale

L'analisi fattoriale consiste in un insieme di tecniche statistiche che hanno lo scopo di individuare le variabili di sintesi preservando la massima variabilità. Si propone quindi di determinare un certo numero di variabili "latenti" (fattori non direttamente misurabili nella realtà) più ristretto e riassuntivo rispetto al numero di variabili di partenza. Nel nostro caso il numero delle variabili è pari a 20 quindi è lecito utilizzare la tecnica di analisi fattoriale per ridurre le variabili. A tale scopo si presta l'**Analisi in componenti principali (ACP)**.

La tecnica utilizzata per l'analisi in componenti principali è la seguente:

Partendo da un numero p di variabili nella matrice dei dati, nel nostro caso 20, si cerca di sintetizzarle in un numero inferiore di variabili (variabili latenti) che siano date da una combinazione lineare della variabile di partenza, cioè si cerca di proiettare le variabili iniziali in un sottospazio vettoriale di dimensione inferiore, sotto la condizione di minima perdita delle informazioni rispetto alla struttura originale.

La tecnica utilizzata per estrarre i fattori si basa sull'impostare che gli autovalori siano maggiori di 1. Ha come vantaggio quello di vedere quante dimensioni emergono in fase esplorativa. Il limite di tale criterio è che tende a sovrastimare il numero di fattori se le variabili di partenza sono molte e, viceversa a sottostimarle se sono poche. Attraverso la tabella della varianza totale spiegata, considerando le nuove componenti, si osserva la percentuale di variabilità totale di partenza.

### VARIANZA TOTALE SPIEGATA

#### Varianza totale spiegata<sup>a</sup>

Componente	Totale	Autovalori iniziali		Caricamenti somme dei quadrati di estrazione		
		% di varianza	% cumulativa	Totale	% di varianza	% cumulativa
1	4,608	23,039	23,039	4,608	23,039	23,039
2	4,073	20,364	43,403	4,073	20,364	43,403
3	2,240	11,201	54,604	2,240	11,201	54,604
4	1,710	8,550	63,154	1,710	8,550	63,154
5	1,409	7,047	70,201	1,409	7,047	70,201
6	1,315	6,576	76,777	1,315	6,576	76,777
7	1,113	5,567	82,344	1,113	5,567	82,344
8	0,910	4,549	86,893			
9	0,755	3,775	90,668			
10	0,664	3,320	93,988			
11	0,411	2,055	96,043			
12	0,270	1,348	97,392			
13	0,193	0,965	98,357			
14	0,145	0,727	99,084			
15	0,101	0,503	99,587			
16	0,038	0,189	99,777			
17	0,028	0,141	99,918			
18	0,010	0,048	99,966			
19	0,004	0,022	99,988			
20	0,002	0,012	100,000			

Metodo di estrazione: Analisi dei componenti principali.

a. Nella fase di analisi vengono utilizzati solo i casi per cui TEST = 1 (FILTER) = Selected.

Nelle tre colonne sono riportati: gli autovalori corrispondenti a ogni fattore, la percentuale di varianza spiegata corrispondente e la relativa percentuale di varianza cumulata (Cumulative %). Invece le tre colonne riproposte accanto forniscono gli stessi dati, ma relativi solo ai fattori che si è deciso di estrarre per la soluzione fattoriale non ruotata.

Gli autovalori scelti sono pertanto 7 e, complessivamente, spiegano più del 82% della variabilità presente nello spazio originario a 20 dimensioni: c'è quindi un risparmio di dimensioni (da 20 a 7), a fronte di una perdita di informazione del 18%. In realtà si potevano scegliere anche 6 componenti dato che spiegano più del 76% della variabilità originaria.

Per vedere rispetto a quali indicatori questa perdita di informazione è maggiore, si può consultare la tabella della Comunalità.

**Comunalità<sup>a</sup>**

	Iniziale	Estrazione
Punteggio Z: Consumo materiale interno	1,000	0,816
Punteggio Z: Dispersione da rete idrica comunale	1,000	0,678
Punteggio Z: Conferimento dei rifiuti urbani in discarica	1,000	0,806
Punteggio Z: Coste marine balneabili	1,000	0,793
Punteggio Z: Qualità dell'aria PM2,5	1,000	0,880
Punteggio Z: Disponibilità di verde urbano	1,000	0,918
Punteggio Z: Soddisfazione per la situazione ambientale	1,000	0,838
Punteggio Z: Preoccupazione per la perdita di biodiversità	1,000	0,923
Punteggio Z: Preoccupazione per i cambiamenti climatici	1,000	0,782
Punteggio Z: Siti contaminati	1,000	0,579
Punteggio Z: Popolazione esposta al rischio di frane	1,000	0,902
Punteggio Z: Popolazione esposta al rischio di alluvioni	1,000	0,787
Punteggio Z: Trattamento delle acque reflue	1,000	0,852
Punteggio Z: Aree protette	1,000	0,836
Punteggio Z: Energia elettrica da fonti rinnovabili	1,000	0,916
Punteggio Z: Impermeabilizzazione del suolo da copertura artificiale	1,000	0,907
Punteggio Z: Rifiuti urbani raccolti	1,000	0,841
Punteggio Z: Indice di durata dei periodi di caldo	1,000	0,738
Punteggio Z: Giorni con precipitazione estremamente intensa	1,000	0,931
Punteggio Z: Giorni consecutivi senza pioggia	1,000	0,746

Metodo di estrazione: Analisi dei componenti principali.

a. Nella fase di analisi vengono utilizzati solo i casi per cui TEST = 1 (FILTER) = Selected.

**LA COMUNALITÀ**

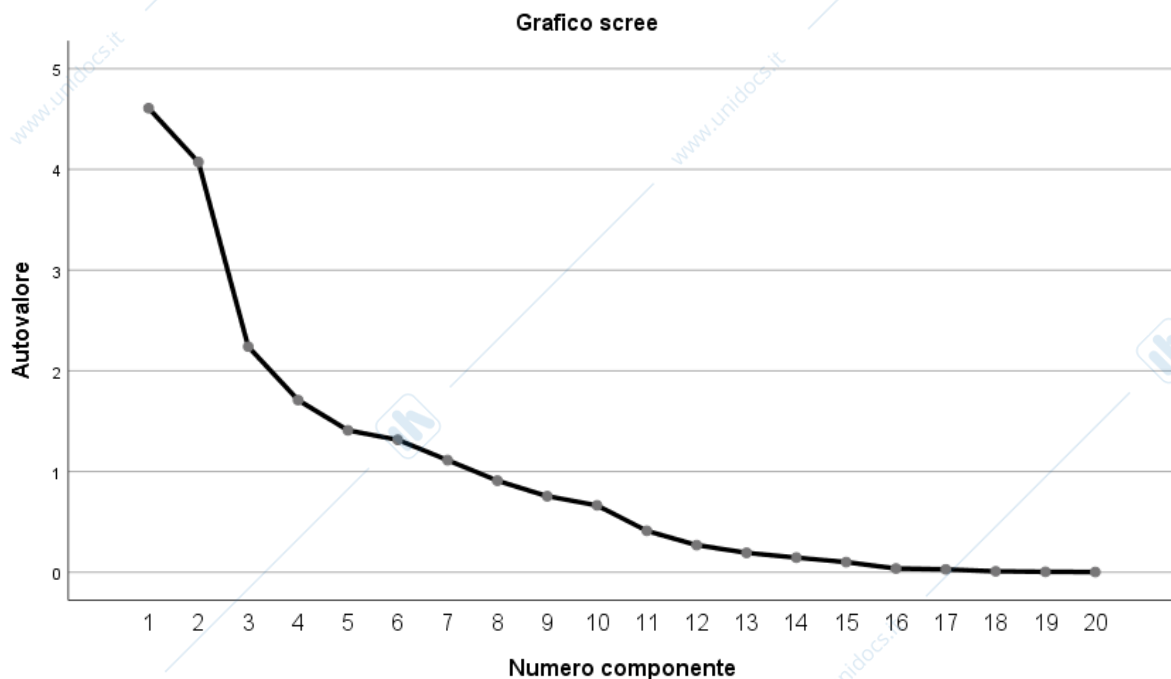
La comunalità è la percentuale di varianza della variabile spiegata dai fattori comuni, cioè considerando la seguente tabella della Comunalità, la prima colonna rappresenta la variabilità iniziale per ogni componente, invece la seconda colonna mostra quant'è la variabilità degli indicatori di partenza se stessi considerando solo 7 componenti (nel nostro caso).

Nel caso di studio in esame, il suo valore minimo è riportato dall' indice "Siti contaminati" con un valore pari al 58%, quindi un valore non critico.

## GRAFICO SCREE

Il Grafico Scree rappresenta graficamente la progressione decrescente delle grandezze degli autovalori associati ai fattori, riportando gli autovalori sull'asse delle ordinate e i fattori in ascissa. Dall'analisi dello Grafico Scree si può decidere di interrompere il processo di estrazione nel punto in cui la curva degli autovalori cambia pendenza (anche laddove tende a diventare piatta).

Possiamo osservare la profonda pendenza della curva tra la seconda e la terza componente, infatti il valore dell'autovalore è diminuito da 4 fino a 2. Dalla terza componente in poi continua a diminuire fino alla settima componente invece oltre, la curva tende a diventare piatta.



## LA MATRICE DELLE COMPONENTI

La matrice delle componenti è una matrice che mette in correlazione gli indicatori originari con le nuove variabili. Nella seguente tabella ho eliminato tutti i coefficienti di correlazione che hanno valore inferiore allo 0.3 in valore assoluto.

Osserviamo che la prima componenti è correlata fortemente con l'indicatore "Preoccupazione per la perdita di biodiversità". I coefficienti perdono di peso all'aumentare delle componenti, questo è dovuto al fatto che la prima e la seconda componente hanno un'alta variabilità, invece dalla terza componente inizia a diminuire.

Matrice dei componenti<sup>a,b</sup>

	Componente						
	1	2	3	4	5	6	7
Punteggio Z: Preoccupazione per la perdita di biodiversità	0,782						-0,457
Punteggio Z: Indice di durata dei periodi di caldo	0,753						
Punteggio Z: Rifiuti urbani raccolti	0,690		-0,441				
Punteggio Z: Dispersione da rete idrica comunale	-0,683						
Punteggio Z: Coste marine balneabili	-0,665				0,485		
Punteggio Z: Consumo materiale interno	0,584		0,501				
Punteggio Z: Conferimento dei rifiuti urbani in discarica	-0,566	0,313	-0,490				
Punteggio Z: Trattamento delle acque reflue	0,481	0,354				-0,428	-0,461
Punteggio Z: Energia elettrica da fonti rinnovabili		0,856					
Punteggio Z: Impermeabilizzazione del suolo da copertura artificiale		-0,810		-0,348		0,318	
Punteggio Z: Popolazione esposta al rischio di frane		0,730	-0,395	-0,328			
Punteggio Z: Soddisfazione per la situazione ambientale	0,508	0,605		0,404			
Punteggio Z: Qualità dell'aria PM 2,5	0,350	-0,592	0,477				
Punteggio Z: Aree protette	-0,312	0,541	0,304	-0,404		0,334	
Punteggio Z: Popolazione esposta al rischio di alluvioni	0,386	-0,334	-0,593		0,380		
Punteggio Z: Giorni consecutivi senza pioggia	-0,358	-0,354		0,506		-0,419	
Punteggio Z: Preoccupazione per i cambiamenti climatici			-0,334	0,505		0,464	
Punteggio Z: Disponibilità di verde urbano		0,382	0,382	0,360	0,567		0,407
Punteggio Z: Siti contaminati		-0,368			-0,492		
Punteggio Z: Giorni con precipitazione estremamente intensa	0,426	0,398	0,500			0,508	

Metodo di estrazione: Analisi dei componenti principali.

a. 7 componenti estratti.

b. Nella fase di analisi vengono utilizzati solo i casi per cui TEST = 1 (FILTER) = Selected.

## CLUSTER ANALYSIS

La Cluster Analysis consiste in un insieme di tecniche statistiche che hanno lo scopo di ridurre la complessità della matrice dei dati. A differenza dell'analisi fattoriale, la cluster analysis opera sulle unità statistiche (le righe) raggruppando le unità che hanno tipologie di comportamento simile, quindi nello stesso gruppo le unità devono essere omogenee fra loro, ma i gruppi fra loro devono essere eterogenei.

La relazione tra le unità si può determinare in due modi: la similarità oppure le distanze. Nel nostro caso i caratteri sono quantitativi e si preferisce usare le misure di distanza.

### 1. TECNICA GERARCHICA

**Il processo utilizzato:** si inizia con  $n$  gruppi di una sola unità. Per ogni passo si raggruppano le unità più simili fino ad arrivare ad un unico gruppo di " $n$ " unità (aggregazione gerarchica **ascendente**: algoritmi **aggregativi**).

Un altro modo è partire da un solo gruppo con " $n$ " unità e si arriva agli " $n$ " gruppi con una sola unità (aggregazione gerarchica **discendente**: algoritmi **scissori**).

Nel nostro caso dato che le unità sono 21 la tecnica gerarchica è la più adatta. La metrica usata è quella euclidea (le distanze al quadrato), cioè si sta considerando il **metodo Ward**.

Attraverso la **matrice di prossimità** possiamo osservare le varie distanze euclidee al quadrato tra le diverse regioni.

**Matrice di prossimità**  
Distanza euclidea al quadrato

Caso	1:Piemonte	2:Valle d'Aosta/Valle d'Aoste	3:Liguria	4:Lombardia	6:Bozano/Bozen	7:Trento	8:Veneto	9:Friuli-Venezia Giulia	10:Emilia-Romagna	11:Toscana	12:Umbria	13:Marche	14:Lazio	15:Abruzzo	16: Molise	17:Campania	18:Puglia	19:Basilicata	20:Calabria	21:Siria	22:Sardegna
1:Piemonte	0,000	74,913	24,574	18,662	53,003	54,297	27,530	21,437	45,916	29,440	30,838	33,428	25,916	44,313	62,254	38,688	25,859	61,816	53,961	48,149	34,496
2:Valle d'Aosta/Valle d'Aoste	74,913	0,000	46,206	74,358	57,383	57,778	74,837	71,440	85,611	51,725	56,431	56,827	65,688	54,157	56,463	69,414	70,491	74,564	68,557	92,323	71,192
3:Liguria	24,574	46,206	0,000	18,158	44,485	37,656	13,160	20,188	31,308	12,654	14,846	14,185	11,815	14,107	27,835	23,357	18,709	47,463	32,680	35,244	28,686
4:Lombardia	18,662	74,358	18,158	0,000	46,447	42,349	6,711	21,275	27,223	20,092	25,399	20,250	17,012	40,551	62,687	28,247	21,222	63,123	48,676	40,030	40,318
6:Bozano/Bozen	53,003	57,383	44,485	46,447	0,000	31,348	56,685	53,702	68,355	58,752	49,915	57,401	50,897	55,889	75,084	70,821	53,099	80,309	78,153	92,874	65,752
7:Trento	54,297	57,778	37,656	42,349	31,348	0,000	38,740	34,294	69,928	46,902	45,295	46,243	44,301	42,502	60,692	61,201	54,771	48,372	62,472	68,564	57,268
8:Veneto	27,530	74,837	13,160	6,711	56,685	38,740	0,000	15,804	33,610	17,994	21,506	17,326	11,987	26,187	50,676	22,378	19,562	54,949	36,541	34,084	36,557
9:Friuli-Venezia Giulia	21,437	71,440	20,188	21,275	53,702	34,294	15,804	0,000	50,610	29,694	31,225	33,847	27,507	38,207	61,389	43,737	37,859	68,257	54,381	47,702	37,823
10:Emilia-Romagna	45,916	85,611	31,308	27,223	68,355	69,928	33,610	50,610	0,000	19,725	42,444	37,764	38,755	60,773	77,904	60,993	49,304	86,258	76,713	65,125	55,722
11:Toscana	29,440	51,725	12,654	20,092	58,752	46,902	17,994	29,694	19,725	0,000	11,829	13,537	18,595	31,515	41,823	39,073	25,865	53,414	40,112	38,547	25,959
12:Umbria	30,838	56,431	14,846	25,399	49,915	45,295	21,506	31,225	42,444	11,829	0,000	13,777	11,862	20,896	37,094	28,890	11,787	35,082	24,865	30,210	20,602
13:Marche	33,428	56,827	14,185	20,250	57,401	46,243	17,326	33,847	37,764	13,537	13,777	0,000	17,038	21,323	29,172	28,404	13,481	36,433	19,344	25,866	32,964
14:Lazio	25,916	65,688	11,815	17,012	50,897	44,301	11,987	27,507	38,755	18,595	11,862	17,038	0,000	11,485	35,263	12,207	9,981	41,186	27,122	25,498	22,202
15:Abruzzo	44,313	54,157	14,107	40,551	55,889	42,502	26,187	38,207	60,773	31,515	20,896	21,323	11,485	0,000	23,849	17,502	18,402	32,153	20,155	37,967	38,899
16:Molise	62,254	56,463	27,835	62,687	75,084	60,692	50,676	61,389	77,904	41,823	37,094	29,172	35,263	23,849	0,000	55,572	44,798	42,048	42,870	42,706	41,170
17:Campania	38,688	69,414	17,709	28,247	70,821	61,201	22,378	43,737	60,993	39,073	28,890	28,404	12,207	17,502	55,572	0,000	11,973	43,411	18,781	32,360	39,938
18:Puglia	25,859	70,491	18,709	21,222	53,099	54,771	19,562	37,859	49,304	25,865	11,787	18,402	9,981	18,402	44,798	11,973	0,000	37,257	12,891	23,517	30,765
19:Basilicata	61,816	74,564	47,463	63,123	80,309	48,372	54,949	68,257	86,258	53,414	35,082	36,433	41,186	42,048	48,411	37,257	0,000	26,946	50,082	45,820	
20:Calabria	53,961	68,557	32,680	48,676	78,153	62,472	36,541	54,381	76,713	40,112	24,865	19,344	27,122	20,155	42,870	18,781	12,891	26,946	0,000	25,515	38,382
21:Siria	48,149	92,323	35,244	40,030	92,874	68,564	34,084	47,702	65,125	38,547	30,210	25,866	25,498	37,967	42,706	32,360	23,517	50,082	25,515	0,000	23,324
22:Sardegna	34,496	71,192	28,686	40,318	65,752	57,268	36,557	37,823	55,722	25,959	20,602	32,964	22,202	38,899	41,170	39,938	30,765	45,820	38,382	23,324	0,000

Questa è una matrice di dissimilarità

Attraverso la seguente tabella “**Pianificazione di agglomerazione**” possiamo osservare passo dopo passo ogni raggruppamento delle unità.

**Stadio 1:** Lombardia – Veneto (con coefficiente di distanza di Ward pari a 3,355) → Nord

**Stadio 2:** Lazio-Puglia (con coefficiente di distanza di Ward pari a 8,346) → Centro e Sud

**Stadio 3:** Toscana-Umbria (con coefficiente di distanza di Ward pari a 14,260) → Centro

Già dai primi passi ci rendiamo conto che si andrà a creare dei gruppi: Nord e Centro insieme al Sud

Inoltre, i coefficienti di distanza aumentano sempre di più, questo è dovuto al fatto che si stanno raggruppando unità fortemente diversificate.

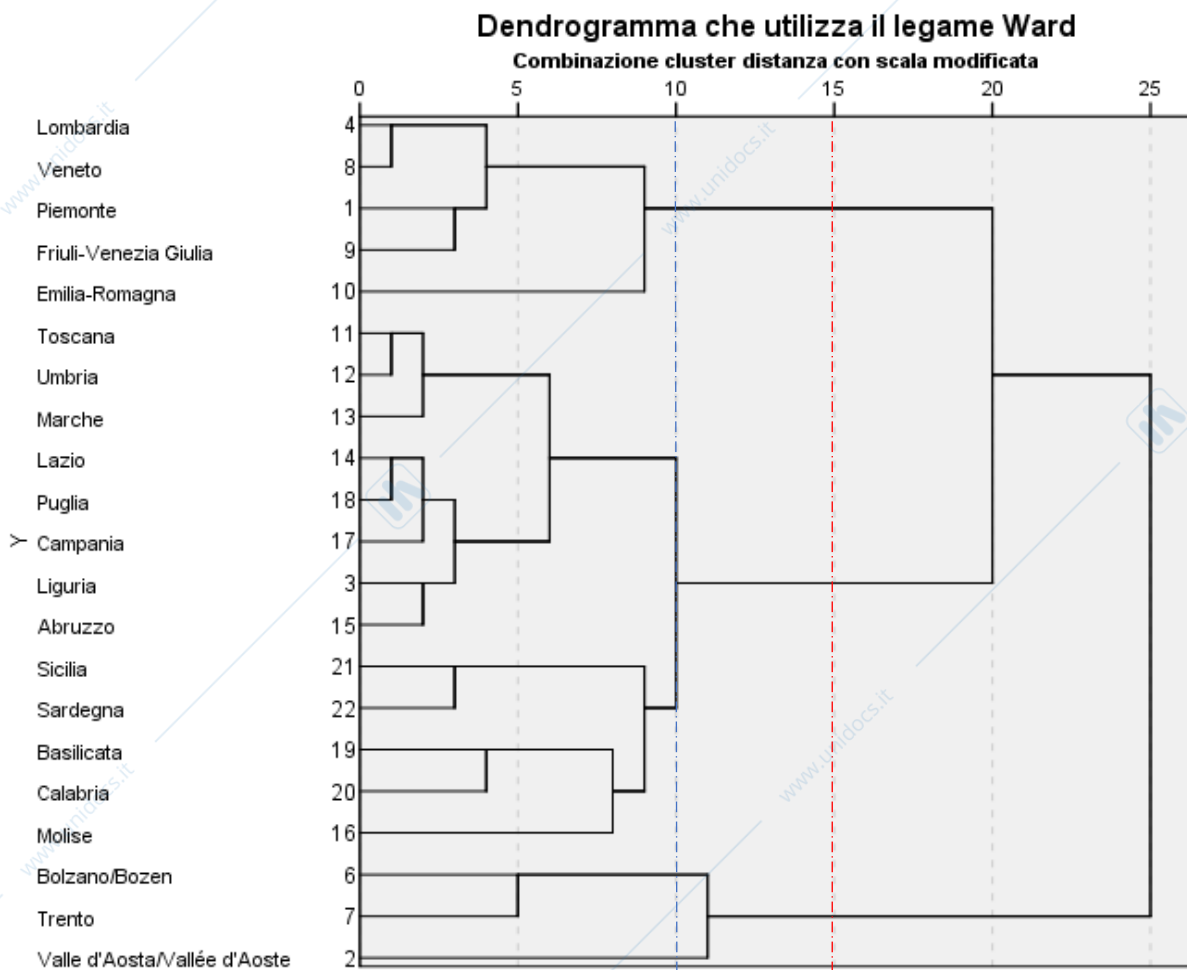
### Pianificazione di agglomerazione

Stadio	Combinato in cluster		Coefficienti	Stadio prima apparizione cluster		Stadio successivo
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	4	8	3,355	0	0	11
2	14	18	8,346	0	0	4
3	11	12	14,260	0	0	6
4	14	17	20,657	2	0	8
5	3	15	27,710	0	0	8
6	11	13	34,844	3	0	13
7	1	9	45,562	0	0	11
8	3	14	57,029	5	4	13
9	21	22	68,691	0	0	16
10	19	20	82,164	0	0	14
11	1	4	95,945	7	1	15
12	6	7	111,619	0	0	18
13	3	11	131,050	8	6	17
14	16	19	154,865	0	10	16
15	1	10	180,766	11	0	19
16	16	21	207,589	14	9	17
17	3	16	237,501	13	16	19
18	2	6	270,664	0	12	20
19	1	3	327,215	15	17	20
20	1	2	400,000	19	18	0

## Dendrogramma

Il Dendrogramma viene utilizzato per fornire una rappresentazione grafica del processo di raggruppamento delle unità attraverso un albero di classificazione. Tagliando l'albero ad un certo livello possiamo individuare i gruppi.

- Tagliando al livello 15 andremmo a crea 3 gruppi: una parte del Nord, il Centro unito al sud e infine Bolzano, Trento e Valle d'Aosta.



Si può osservare che il numero di cluster aumenta se tagliamo al livello 10 ma attraverso questa tabella " **Appartenenza Cluster**" possiamo osservare che non è conveniente.

### Appartenenza cluster

Caso	6 cluster	5 cluster	4 cluster	3 cluster
1:Piemonte	1	1	1	1
2:Valle d'Aosta/Vallée d'Aoste	2	2	2	2
3:Liguria	3	3	3	3
4:Lombardia	1	1	1	1
6: Bolzano/Bozen	4	4	4	2
7:Trento	4	4	4	2
8:Veneto	1	1	1	1
9:Friuli-Venezia Giulia	1	1	1	1
10:Emilia-Romagna	1	1	1	1
11:Toscana	3	3	3	3
12:Umbria	3	3	3	3
13:Marche	3	3	3	3
14:Lazio	3	3	3	3
15:Abruzzo	3	3	3	3
16:Molise	5	5	3	3
17:Campania	3	3	3	3
18:Puglia	3	3	3	3
19:Basilicata	5	5	3	3
20:Calabria	5	5	3	3
21:Sicilia	6	5	3	3
22:Sardegna	6	5	3	3

## 2. TECNICA NON GERARCHICA

Il collettivo iniziale formato da "n" unità saranno collocate in "g" gruppi (g strettamente minore di n), tale che i gruppi formino una partizione dell'insieme delle unità. Successivamente determineremmo i centri dei gruppi o in modo casuale (poco consigliata nel caso in cui il collettivo è numeroso) o attraverso il *baricentro*. In seguito, avremmo lo spostamento delle unità fino a quando non si avrà la minima distanza tra le unità all'interno dei gruppi (varianza within) e massima distanza tra i gruppi (varianza between).

### Aggregazione in 5 clusters

#### Appartenenza cluster

Numero di caso	Regioni ripartizioni geografiche	Cluster	Distanza
1	Piemonte	1	4,355
2	Valle d'Aosta/Vallée d'Aoste	2	0,000
3	Liguria	1	2,672
4	Lombardia	1	3,658
6	Bolzano/Bozen	5	2,799
7	Trento	5	2,799
8	Veneto	1	3,028
9	Friuli-Venezia Giulia	1	4,426
10	Emilia-Romagna	4	0,000
11	Toscana	1	3,471
12	Umbria	1	2,716
13	Marche	1	2,802
14	Lazio	1	2,177
15	Abruzzo	1	3,518
16	Molise	3	0,000
17	Campania	1	3,652
18	Puglia	1	2,544
19	Basilicata	1	5,487
20	Calabria	1	4,068
21	Sicilia	1	4,348
22	Sardegna	1	4,193

#### Numero di casi in ciascun cluster

Cluster	1	16,000
	2	1,000
	3	1,000
	4	1,000
	5	2,000
Valido		21,000
Mancante		0,000

### Aggregazione in 3 clusters

#### Appartenenza cluster

Numero di caso	Regioni ripartizioni geografiche	Cluster	Distanza
1	Piemonte	1	4,347
2	Valle d'Aosta/Vallée d'Aoste	3	4,702
3	Liguria	1	2,374
4	Lombardia	1	3,611
6	Bolzano/Bozen	3	3,647
7	Trento	3	3,665
8	Veneto	1	3,015
9	Friuli-Venezia Giulia	1	4,442
10	Emilia-Romagna	1	5,621
11	Toscana	1	3,194
12	Umbria	1	2,712
13	Marche	1	2,704
14	Lazio	1	2,196
15	Abruzzo	1	3,450
16	Molise	1	5,305
17	Campania	1	3,875
18	Puglia	1	2,784
19	Basilicata	1	5,518
20	Calabria	1	4,315
21	Sicilia	2	0,000
22	Sardegna	1	4,287

#### Numero di casi in ciascun cluster

Cluster	1	17,000
	2	1,000
	3	3,000
Valido		21,000
Mancante		0,000

Mettendo a confronto “Aggregazione in 5 cluster” e “Aggregazione in 3 cluster” notiamo che:

- le provincie autonome: Bolzano e Trento non cambiano, cioè fanno parte sempre dello stesso gruppo.

Considerando 5 cluster, Molise e Emilia Romagna erano due gruppi distinti, formati da un sola unità, ma passando a 3 cluster esse si sono raggruppate al grande gruppo “1”. La Sicilia, che prima faceva parte nel “1” gruppo, passando all’aggregazione in 3 cluster ha formato un gruppo a sé stante, cioè con una sola unità.

Inoltre, osserviamo che sia con 5 cluster sia con i 3 cluster si è creato un gruppo di grande dimensioni formato da 16 unità nel primo caso e 17 nel secondo.

**Confrontando la tecnica gerarchica e non gerarchico osserviamo che:**

- 1) Non c’è più la distinzione tra Nord e Centro insieme al Sud, infatti con il metodo non gerarchico questi due gruppi vengo raggruppati in un solo gruppo.
- 2) In entrambi i casi le provincie autonome hanno mantenuto lo stesso gruppo.

Osservazione:

Per quanto riguarda la cronologia delle iterazioni, sia per raggruppamento in 5 cluster e sia per 3 cluster, è evidente come già dalla quinta iterazione gli spostamenti delle unità tendono a zero: questo vuol dire che non ci sono stati cambiamenti nei baricentri dei gruppi; ciò implica una distinzione abbastanza netta tra di essi.

### Aggregazione in 5 cluster

#### Cronologia delle iterazioni<sup>a</sup>

Iterazione	Modifica nei centri del cluster				
	1	2	3	4	5
1	3,437	0,000	0,000	0,000	1,866
2	0,202	0,000	0,000	0,000	0,622
3	0,012	0,000	0,000	0,000	0,207
4	0,001	0,000	0,000	0,000	0,069
5	4,115E-05	0,000	0,000	0,000	0,023
6	2,421E-06	0,000	0,000	0,000	0,008
7	1,424E-07	0,000	0,000	0,000	0,003
8	8,376E-09	0,000	0,000	0,000	0,001
9	4,927E-10	0,000	0,000	0,000	0,000
10	2,898E-11	0,000	0,000	0,000	9,482E-05

a. Le iterazioni sono state arrestate perché è stato eseguito il numero massimo di iterazioni. Convergenza delle iterazioni non riuscita. La modifica di coordinata assoluta massima per un centro è 5,073E-5. L’iterazione corrente è 10. La distanza minima tra i centri iniziali è 7,455.

### Aggregazione in 3 cluster

#### Cronologia delle iterazioni<sup>a</sup>

Iterazione	Modifica nei centri del cluster		
	1	2	3
1	5,309	0,000	2,749
2	0,295	0,000	0,687
3	0,016	0,000	0,172
4	0,001	0,000	0,043
5	5,057E-05	0,000	0,011
6	2,809E-06	0,000	0,003
7	1,561E-07	0,000	0,001
8	8,671E-09	0,000	0,000
9	4,817E-10	0,000	4,194E-05
10	2,676E-11	0,000	1,048E-05

a. Le iterazioni sono state arrestate perché è stato eseguito il numero massimo di iterazioni. Convergenza delle iterazioni non riuscita. La modifica di coordinata assoluta massima per un centro è 5,299E-6. L’iterazione corrente è 10. La distanza minima tra i centri iniziali è 8,070.