

### 3. Studio di un caso: i paesi membri dell'OECD – anno 2001

#### 3.1 Fonte dei dati – anno 2001

I dati analizzati sono reperibili presso la sezione Statistics Portal del sito dell'OECD, Organization for Economic Cooperation and Development, ([www.oecd.org](http://www.oecd.org)). Essi fanno riferimento all'anno 2001.

**Tabella 1:** *dataset analizzato – anno 2001*

PAESE	AREAGEO	VARPIL	DISOC	D.L.D.	VARDIS	SUSSIDI	VARDOM
Australia	no	2,69	6,84	21,98	6,15	bassi	1,44
Austria	si	0,89	3,60	23,33	4,12	bassi	0,58
Belgio	si	0,71	6,21	51,65	8,03	alti	-0,44
Canada	no	1,78	7,28	9,49	7,73	bassi	1,28
Repubblica Ceca	si	3,10	8,18	52,74	7,51	m. bassi	2,49
Danimarca	si	0,70	4,21	22,22	5,21	alti	3,82
Finlandia	si	2,28	9,12	26,18	9,76	alti	0,01
Francia	si	1,79	8,81	37,59	8,66	medi	6,87
Germania	si	1,64	7,91	50,42	7,78	medi	2,03
Grecia	si	4,20	10,41	52,80	10,39	bassi	1,68
Ungheria	si	3,71	5,73	46,56	6,37	m. bassi	-0,11
Islanda	si	3,92	2,32	12,51	2,99	bassi	4,12
Irlanda	si	4,83	3,74	33,13	8,06	m. bassi	1,97
Italia	si	1,86	9,59	63,36	8,85	m. bassi	-2,11
Giappone	no	0,18	5,24	26,58	3,93	bassi	3,79
Korea	no	3,97	4,18	2,33	3,99	m. bassi	2,04
Lussemburgo	si	2,52	1,81	28,45	3,13	medi	1,43
Messico	no	-0,92	2,63	0,97	3,15	m. bassi	1,02
Olanda	si	1,99	2,51	3,45	3,78	medi	3,68
Nuova Zelanda	no	2,48	5,36	16,74	5,50	bassi	4,54
Norvegia	si	1,99	3,49	5,46	3,77	bassi	-0,43
Polonia	si	1,28	18,59	43,06	16,37	bassi	2,28
Portogallo	si	1,97	4,30	38,13	5,95	bassi	1,66
Slovacchia	si	3,48	19,33	53,72	16,52	bassi	0,58
Spagna	si	3,67	10,53	44,04	12,45	alti	-1,02
Svezia	si	1,41	5,08	22,34	7,43	medi	1,69
Svizzera	si	1,15	2,52	29,91	3,23	m. bassi	8,21
Regno Unito	si	3,15	4,78	27,75	5,85	m. bassi	3,79
Stati Uniti	no	1,08	4,79	6,10	5,44	bassi	0,37

*Legenda:*

**PAESE:** nome degli stati appartenenti all'Oecd

**AREAGEO:** è uno stato europeo?

**VARPIL:** tasso di crescita del pil nel 2001

**DISOC:** tasso di disoccupazione nel 2001

**D.L.D.:** incidenza della disoccupazione di lunga durata (oltre 12 mesi) sul totale dei disoccupati (%)

**VARDIS:** tasso di crescita della disoccupazione nel 2001

**SUSSIDI:** spesa pubblica in sussidi per la disoccupazione in rapporto al pil (%)

(*molto bassi:* sussidi/pil <0,5 ; *bassi:* 0,5<= sussidi/pil <1; *medi:* 1<=sussidi/pil <1,5; *alti:* sussidi/pil =>1,5)

**VARDOM:** tasso di crescita della domanda interna di beni e servizi

Ai fini dell'analisi, la variabile **AREAGEO**, di tipo dicotomico, viene riclassificata ponendo la modalità NO pari a 0 e la modalità SI pari a 1.

Seguendo lo stesso principio, la modalità **SUSSIDI**, rilevata su scala ordinale, viene riclassificata sostituendo le modalità molto bassi, bassi, medi, alti con i valori 1, 2, 3, 4.

### 3.2 Analisi preliminare

Prima di procedere con la cluster analysis, si effettuano alcune analisi preliminari sulle variabili che verranno utilizzate: nello specifico, si produrrà la tabella di correlazione di Pearson al fine di individuare la presenza di eventuali correlazioni elevate tra le variabili.

Nel caso di elevata correlazione, le variabili in questione verranno escluse dall'analisi, con lo scopo di evitare che l'elevata interdipendenza distorca eccessivamente i risultati della stessa.

**Tabella 2:** correlazioni di Pearson – anno 2001

		Correlazioni						
		Areageo	Varpil	Disoc	dld	Vardis	Sussidi	Vardom
Areageo	Correlazione di Pearson	1	,250	,180	,556**	,296	,262	-,024
	Sig. (2-code)		,191	,351	,002	,118	,169	,902
	N	29	29	29	29	29	29	29
Varpil	Correlazione di Pearson	,250	1	,164	,236	,251	-,206	-,081
	Sig. (2-code)	,191		,395	,217	,189	,283	,677
	N	29	29	29	29	29	29	29
Disoc	Correlazione di Pearson	,180	,164	1	,594**	,954**	,103	-,236
	Sig. (2-code)	,351	,395		,001	,000	,595	,218
	N	29	29	29	29	29	29	29
dld	Correlazione di Pearson	,556**	,236	,594**	1	,629**	,053	-,207
	Sig. (2-code)	,002	,217	,001		,000	,785	,281
	N	29	29	29	29	29	29	29
Vardis	Correlazione di Pearson	,296	,251	,954**	,629**	1	,198	-,293
	Sig. (2-code)	,118	,189	,000	,000		,303	,123
	N	29	29	29	29	29	29	29
Sussidi	Correlazione di Pearson	,262	-,206	,103	,053	,198	1	-,107
	Sig. (2-code)	,169	,283	,595	,785	,303		,582
	N	29	29	29	29	29	29	29
Vardom	Correlazione di Pearson	-,024	-,081	-,236	-,207	-,293	-,107	1
	Sig. (2-code)	,902	,677	,218	,281	,123	,582	
	N	29	29	29	29	29	29	29

\*\* . La correlazione è significativa al livello 0,01 (2-code).

Arbitrariamente, si decide di stabilire il valore massimo di Pearson accettato affinché le variabili vengano utilizzate nell'analisi pari a  $|0,900|$ , ciò implicherebbe un'alevata correlazione tra le variabili.

Poiché la variabile **VARDIS** è positivamente correlata con la variabile **DISOC** con un valore, significativo, pari a 0,954, la stessa verrà esclusa dall'analisi. Le restanti variabili non presentano valori eccessivamente elevati di correlazione, pertanto verranno tutte impiegate nel procedimento.

Dopo aver stabilito quali variabili utilizzare, mediante le statistiche descrittive delle stesse si possono fare alcune considerazioni preliminari sul modo in cui queste influenzeranno l'analisi e la creazione dei clusters.

**Tabella 3:** statistiche descrittive – anno 2001

**Statistiche descrittive**

	N	Intervallo	Minimo	Massimo	Media	Deviazione std.	Varianza
Areageo	29	1	0	1	,76	,435	,190
Varpil	29	5,75	-,92	4,83	2,1897	1,33587	1,785
Disoc	29	17,52	1,81	19,33	6,5203	4,25861	18,136
dld	29	62,39	,97	63,36	29,4134	17,94284	321,946
Sussidi	29	3	1	4	2,17	1,002	1,005
Vardom	29	10,32	-2,11	8,21	1,9745	2,24783	5,053
Validi (listwise)	29						

La tabella mostra che la variabile **DLD** è quella che presenta una maggiore deviazione standard con un valore pari a 17,94284. Questo indica che è in questa variabile che si presenta una maggiore dispersione dei paesi esaminati e quindi è ipotizzabile che la stessa contribuisca in maniera rilevante alla formazione dei gruppi.

A seguire, si possono identificare la variabile **DISOC**, con un valore di standard deviation pari a 4,26861, seguita da **VARDOM** con livelli inferiori di deviazione standard rispetto a **DLD**, ma comunque significativi.

### 3.3 Un apporccio gerarchico – il metodo di Ward

Dopo aver verificato che, a causa dell'elevata correlazione tra **DISOC** e **VARDIS**, la seconda verrà esclusa dal procedimento, si procede ora con un'analisi basata sul metodo di Ward. La distanza utilizzata sarà la distanza euclidea quadratica e i valori iniziali, in quanto osservati su scale differenti, verranno standardizzati nei valori Z (sottraendo agli stessi la loro media e dividendo per lo scarto quadratico), in modo da eliminare eventuali distorsioni derivanti dalle diverse tipologie di misurazione.

Nello specifico, verranno analizzati il programma di agglomerazione, l'icicle diagram, il dendrogramma e, in ultimo, le caratteristiche peculiari dei clusters formati.

#### 3.3.1 Il programma di agglomerazione

Il programma di agglomerazione fornisce informazioni circa gli elementi e i gruppi combinati ad ogni stadio e la distanza a cui si verifica l'unione. Esso rappresenta una sintesi numerica del processo di agglomerazione.

Il cluster risultante dall'unione di due elementi prenderà il nome dell'elemento indicato nella seconda colonna nominata "cluster 1".

L'ultima colonna sulla destra, denominata "Stadio successivo", indica in quale stadio il cluster formato verrà nuovamente accorpato ad altri elementi.

Le due colonne riguardanti lo stadio di formazione del cluster indicano, rispettivamente, a quale stadio si è formato il primo cluster e a quale stadio si è formato il secondo cluster che vengono uniti allo stadio indicato nella prima colonna. Tali colonne riporteranno un valore pari a zero fino a quando non verranno accorpati due gruppi di cui almeno uno costituito da più di un elemento.

**Tabella 4:** *programma di agglomerazione – anno 2001*

**Programma di agglomerazione**

Stadio	Cluster accorpati		Coefficienti	Stadio di formazione del cluster		Stadio successivo
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	4	29	,408	0	0	15
2	17	26	1,113	0	0	8
3	2	23	1,909	0	0	8
4	5	11	2,907	0	0	9
5	1	20	3,973	0	0	15
6	13	28	5,166	0	0	12
7	7	25	6,363	0	0	18
8	2	17	7,659	3	2	14
9	5	10	9,050	4	0	19
10	3	9	10,475	0	0	18
11	6	19	12,068	0	0	21
12	12	13	13,740	0	6	23
13	22	24	15,574	0	0	24
14	2	21	17,578	8	0	21
15	1	4	19,669	5	1	16
16	1	15	22,178	15	0	20
17	8	27	25,643	0	0	23
18	3	7	29,251	10	7	25
19	5	14	32,956	9	0	24
20	1	16	36,958	16	0	22
21	2	6	41,489	14	11	25
22	1	18	47,379	20	0	28
23	8	12	56,984	17	12	26
24	5	22	66,692	19	13	27
25	2	3	78,385	21	18	26
26	2	8	98,660	25	23	27
27	2	5	126,745	26	24	28
28	1	2	168,000	22	27	0

Osservando i dati della tabella si possono evincere alcune considerazioni:

- a) l'ordine di aggregazione dipende dalla distanza tra i casi, quindi si aggregeranno per primi i casi più simili tra loro. Nel caso esaminato, i casi più simili tra loro sono il 4 e 29, seguiti dagli elementi 17 e 26 e, successivamente, dal 2 e 23.
- b) i primi due elementi che si fondono in un cluster sono il 4 e il 29 (rispettivamente Canada e Stati Uniti) che si raggruppano ad un livello di distanza pari a 0,408; se la distanza di aggregazione fosse stata pari a 0,000, questo avrebbe significato che i due elementi presentavano lo stesso valore per tutte le variabili analizzate. Il gruppo, così formato tra gli elementi 4 e 29, d'ora in avanti verrà identificato con il numero del primo elemento cioè 4.
- c) Il gruppo formato da questi due elementi verrà raggruppato nuovamente allo stadio 15 (colonna "stadio successivo") con il gruppo 1, creato allo stadio 5 attraverso la fusione degli elementi 1 e 20 (Australia e Nuova Zelanda).
- d) Il primo stadio in cui si fondono due gruppi, di cui almeno uno è formato da più di un elemento, è lo stadio 8, dove si fondono il cluster 2, formato allo stadio 3 dagli elementi 2 e 23 (Austria e Portogallo), e il gruppo 17, creato allo stadio 2 tra gli elementi 17 e 26 (Lussemburgo e Svezia). Fino a questo punto, cioè per i primi 7 stadi, si sono formati clusters che accorpano tra loro solo elementi singoli. Il cluster formato allo stadio 8 verrà identificato come 2 e verrà unito, allo stadio 14, con l'elemento 21 (Norvegia).
- e) L'ultimo stadio, il numero 28, vedrà l'accorparsi dei cluster 1 e 2. Questi si uniranno a un livello di distanza pari a 168,000 e deriveranno dall'unione avvenuta, rispettivamente, allo stadio 22 tra i cluster 1 e l'elemento 18 (Messico) e, allo stadio 27, tra i gruppi 2 e 5.

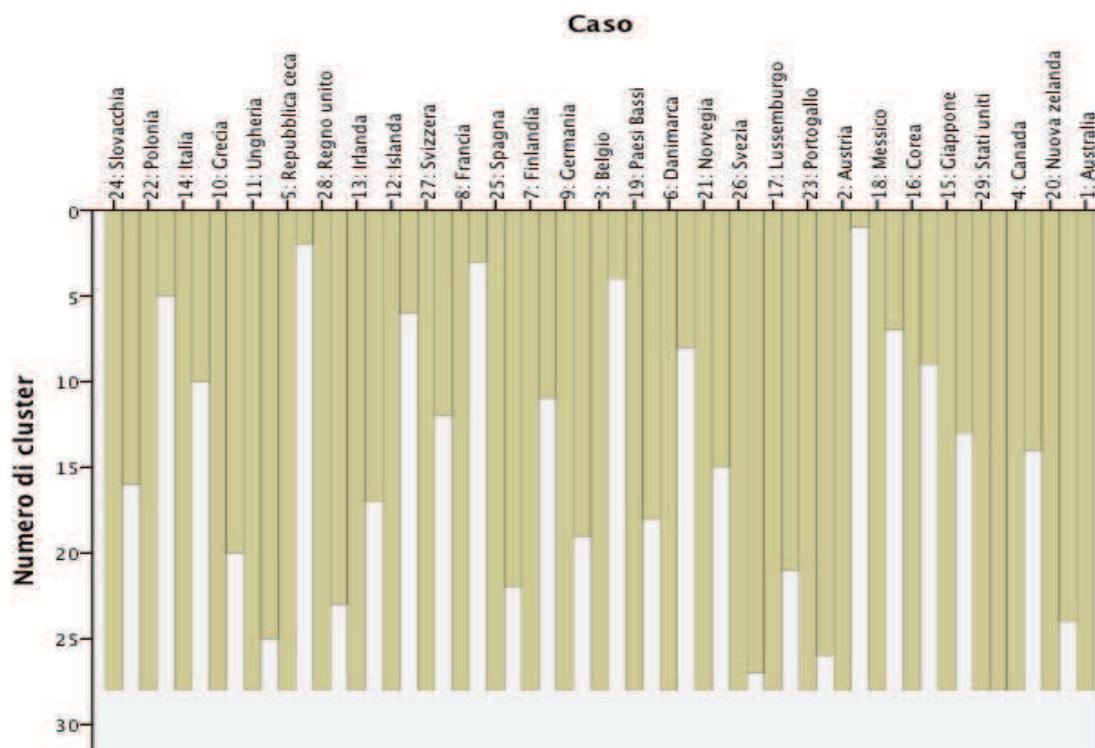
### 3.3.2 Icicle diagram

L'icicle diagram fornisce, graficamente, le stesse informazioni individuate nel programma di agglomerazione.

Il nome del grafico deriva dalla similitudine tra le colonne e le colonne di ghiaccio pendenti dalla grondaia di una casa.

Il diagramma va letto dal basso verso l'alto, dove le colonne indicano gli elementi che vengono via via accorpati e le righe rappresentano il numero di gruppi.

**Grafico 1:** icicle diagram – anno 2001



Esaminando il grafico, si possono fare le seguenti riflessioni:

- i primi due elementi che si fondono sono 4 e 29 (Canada e Stati Uniti) nella parte destra del grafico. Il numero di cluster rimanenti a seguito dell'unione è pari a 28.
- Successivamente agli elementi 4 e 29, vengono uniti gli elementi 17 e 26 (Lussemburgo e Svezia) riducendo il numero di gruppi da 28 a 27.
- Fino al momento in cui il gruppo formato dagli elementi 2 e 23 viene unito con il cluster 17, formato dai casi 17 e 26, diminuendo a 21 il numero di gruppi, non vi sono unioni tra gruppi costituiti da più di un elemento.

In sintesi, il grafico deve essere letto dal basso verso l'alto salendo di "un gradino" alla volta; in questo modo si identificheranno, di volta in volta, gli elementi che si fondono e il numero di cluster risultante dopo l'unione.

### 3.3.3 La scelta del numero di gruppi

Per identificare il numero di gruppi più sensato in cui raggruppare gli elementi del dataset analizzato si possono utilizzare svariati metodi.

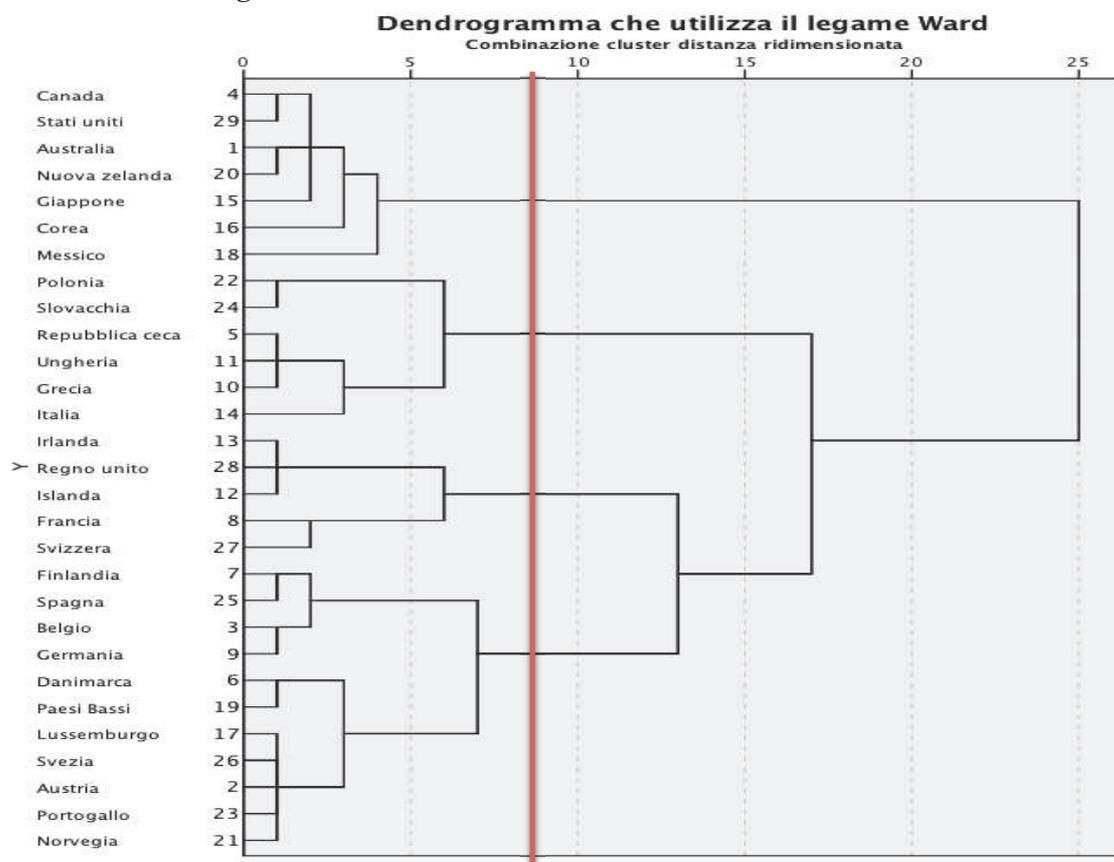
In questo contesto, si prenderanno in esame il dendrogramma, lo scree-plot della distanza di fusione e la tabella che riporta i coefficienti, la distanza di fusione riscalata e la variazione della stessa nel programma di agglomerazione.

#### 3.3.3.1 Il dendrogramma

Il dendrogramma rappresenta una sintesi grafica del risultato ottenuto dall'analisi gerarchica.

Nell'asse verticale di sinistra si leggono gli oggetti presenti nell'analisi, l'asse orizzontale mostra la distanza tra i cluster nel momento in cui questi vengono uniti. L'albero fornisce vari livelli di aggregazione: la scelta del livello a cui "tagliare" l'albero deve rappresentare un giusto compromesso tra numero di gruppi e omogeneità degli stessi. Generalmente, il taglio va fatto prima delle aggregazioni corrispondenti a salti molto grossi tra i valori dell'indice, considerato che un salto elevato nella distanza di aggregazione indicherebbe l'unione di due gruppi molto differenti tra loro.

Grafico 2: dendrogramma – anno 2001

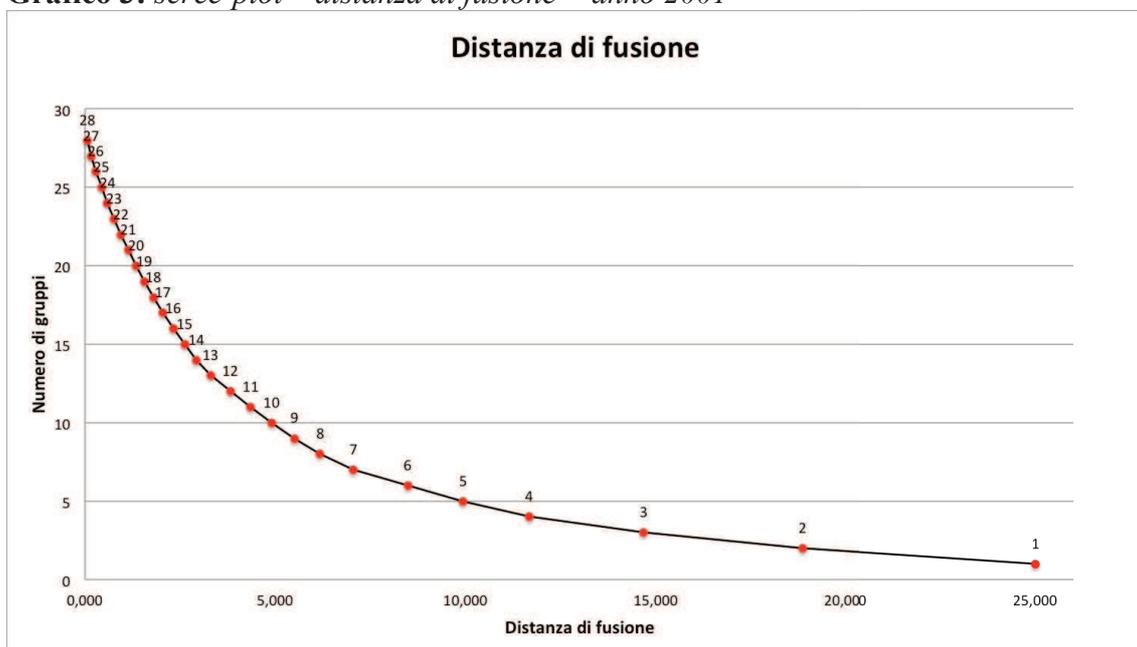


Il grafico presenta un salto consistente nella distanza di fusione nel passaggio da 4 a 3 clusters, quindi si ritiene opportuno “tagliare” in corrispondenza di 4 gruppi. Un taglio precedente comporterebbe la non unione di elementi simili, mentre la decisione di dividere il dataset in soli 3 cluster implicherebbe l’aggregazione in un unico gruppo di paesi molto diversi tra loro.

### 3.3.3.2 Lo scree-plot della distanza di fusione

Un ulteriore approccio grafico viene fornito dallo scree-plot della distanza di fusione, grafico in cui viene posto il numero di gruppi in ordinata ed in ascissa la distanza di fusione.

**Grafico 3:** scree-plot – distanza di fusione - anno 2001



Il grafico, in questo caso, suggerisce di operare una suddivisione in 4 gruppi. Infatti, nel passaggio da 4 a 3 clusters, si registra un consistente aumento della distanza di fusione.

### 3.3.3.3 L’incremento relativo della distanza di fusione

L’incremento relativo della distanza di fusione consente di operare una scelta ottimale del numero di gruppi attraverso un criterio numerico.

Si decide, infatti, di terminare il processo di agglomerazione nel momento in cui, raggiunto un numero di clusters abbastanza ridotto in proporzione alla numerosità della popolazione di analisi, l’unione di ulteriori due gruppi coincide con il valore massimo dell’incremento relativo della distanza di fusione. Si opera questa scelta in quanto la continuazione del processo di agglomerazione implicherebbe l’unione di due clusters molto lontani (e quindi diversi) tra loro.

**Tabella 5:** *distanza di fusione – anno 2001*

STADIO	NUMERO DI GRUPPI	COEFFICIENTI	DISTANZA DI FUSIONE	INCREMENTO RELATIVO
1	28	0,408	0,061	0,000
2	27	1,113	0,166	1,728
3	26	1,909	0,284	0,715
4	25	2,907	0,433	0,523
5	24	3,973	0,591	0,367
6	23	5,166	0,769	0,300
7	22	6,363	0,947	0,232
8	21	7,659	1,140	0,204
9	20	9,050	1,347	0,182
10	19	10,475	1,559	0,157
11	18	12,068	1,796	0,152
12	17	13,740	2,045	0,139
13	16	15,574	2,318	0,133
14	15	17,578	2,616	0,129
15	14	19,669	2,927	0,119
16	13	22,178	3,300	0,128
17	12	25,643	3,816	0,156
18	11	29,251	4,353	0,141
19	10	32,956	4,904	0,127
20	9	36,958	5,500	0,121
21	8	41,489	6,174	0,123
22	7	47,379	7,050	0,142
23	6	56,984	8,480	0,203
24	5	66,692	9,924	0,170
25	4	78,385	11,664	0,175
26	3	98,660	14,682	0,259
27	2	126,745	18,861	0,285
28	1	168,000	25,000	0,325

La tabella mostra che uno dei livelli maggiori di incremento relativo, escludendo i primi 6 stadi in cui il numero di gruppi rimane eccessivamente elevato, si ha allo stadio 26, nella riduzione da 4 a 3 clusters, con un valore pari a 0,259. Questo valore suggerisce, come già constatato attraverso i metodi grafici, che la divisione ottimale è quella in 4 gruppi.

I livelli di incremento relativo degli stadi 27 e 28 sono superiori a quelli del livello 26, ma, giunti a questi livelli, si sarebbero già uniti elementi troppo diversi tra loro. Anche in alcuni stadi precedenti il 26 si hanno incrementi abbastanza simili al valore 0,259, come ad esempio allo stadio 23, ma si ritiene che, dato il ridotto numero di paesi analizzati, una divisione in 6 gruppi risulterebbe troppo dispersiva.

### 3.3.4 Le caratteristiche dei gruppi

Dopo aver identificato il numero ottimale di gruppi in cui suddividere la popolazione, si procede ad una analisi degli elementi che compongono ciascun gruppo e delle caratteristiche peculiari di ciascuno dei clusters formati.

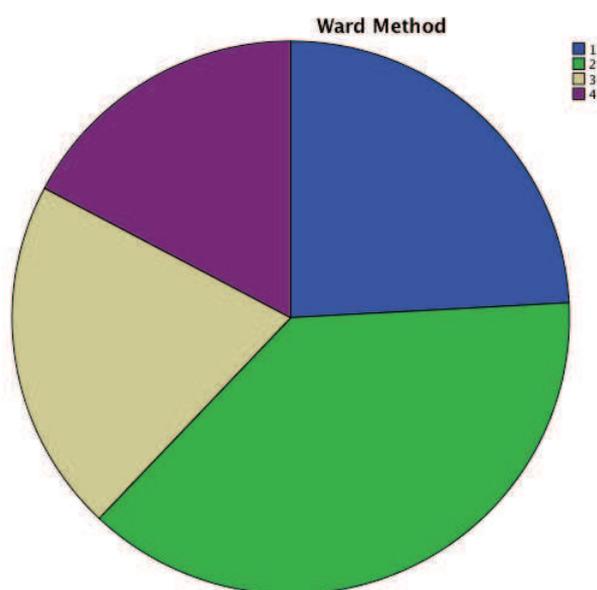
#### 3.3.4.1 La numerosità dei gruppi

Come primo step, si procede all'identificazione della numerosità dei 4 clusters, che devono essere equilibrati tra loro e avere una dimensione sufficientemente grande.

**Tabella 6:** numerosità dei cluster – anno 2001

Ward Method					
		Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Validi	1	7	24,1	24,1	24,1
	2	11	37,9	37,9	62,1
	3	6	20,7	20,7	82,8
	4	5	17,2	17,2	100,0
Totale		29	100,0	100,0	

**Grafico 4:** grafico a torta – numerosità dei cluster – anno 2001



Il grafico a torta e la tabella di numerosità individuano la presenza di un gruppo principale (gruppo 2), costituito da 11 elementi, e 3 gruppi di dimensioni simili, comprendenti rispettivamente 7, 6 e 5 paesi.

Il gruppo 2 include circa il 38% dei paesi analizzati, mentre i paesi restanti si suddividono in maniera abbastanza equa tra i 3 gruppi che comprendono, rispettivamente, circa il 24%, il 20% e il 17% dei casi.

### 3.3.4.2 Gli elementi nei gruppi

Successivamente all'identificazione della numerosità dei gruppi formati, si procede all'individuazione degli elementi che costituiscono ciascun cluster attraverso l'analisi della tabella che riporta il cluster di appartenenza per ciascun caso.

**Tabella 7:** cluster di appartenenza – anno 2001

Caso	4 cluster
1:Australia	1
2:Austria	2
3:Belgio	2
4:Canada	1
5:Repubblica ceca	3
6:Danimarca	2
7:Finlandia	2
8:Francia	4
9:Germania	2
10:Grecia	3
11:Ungheria	3
12:Islanda	4
13:Irlanda	4
14:Italia	3
15:Giappone	1
16:Corea	1
17:Lussemburgo	2
18:Messico	1
19:Paesi Bassi	2
20:Nuova zeland	1
21:Norvegia	2
22:Polonia	3
23:Portogallo	2
24:Slovacchia	3
25:Spagna	2
26:Svezia	2
27:Svizzera	4
28:Regno unito	4
29:Stati uniti	1

Dai dati riportati sulla tabella, si può evincere che i 4 cluster sono composti ciascuno dai seguenti paesi:

- 1) cluster 1: Australia, Canada, Giappone, Corea, Messico, Nuova Zelanda, Stati Uniti.
- 2) cluster 2: Austria, Belgio, Danimarca, Finlandia, Germania, Lussemburgo, Paesi Bassi, Norvegia, Portogallo, Spagna, Svezia.
- 3) cluster 3: Repubblica Ceca, Grecia, Ungheria, Italia, Polonia, Slovacchia,
- 4) cluster 4: Francia, Islanda, Irlanda, Svizzera, Regno Unito.

### 3.3.4.3 Le specifiche dei gruppi

A conclusione dell'analisi gerarchica, si confrontano i 4 cluster identificati per conoscerne le caratteristiche. A tal fine, si producono le statistiche descrittive delle variabili utilizzate nell'analisi.

**Tabella 8:** *caratteristiche dei gruppi – anno 2001*

		Statistiche descrittive					
Ward Method	N	Intervallo	Minimo	Massimo	Media	Deviazione std.	
1	Areageo	7	0	0	0	,00	,000
	Varpil	7	4,89	-,92	3,97	1,6086	1,64562
	Disoc	7	4,65	2,63	7,28	5,1886	1,57351
	dld	7	25,61	,97	26,58	12,0271	9,92404
	Sussidi	7	1	1	2	1,71	,488
	Vardom	7	4,17	,37	4,54	2,0686	1,53154
	Validi (listwise)	7					
2	Areageo	11	0	1	1	1,00	,000
	Varpil	11	2,97	,70	3,67	1,7973	,88038
	Disoc	11	8,72	1,81	10,53	5,3427	2,79163
	dld	11	48,20	3,45	51,65	28,6973	16,18039
	Sussidi	11	2	2	4	3,09	,831
	Vardom	11	4,84	-1,02	3,82	1,1827	1,62112
	Validi (listwise)	11					
3	Areageo	6	0	1	1	1,00	,000
	Varpil	6	2,92	1,28	4,20	2,9383	1,13311
	Disoc	6	13,60	5,73	19,33	11,9717	5,64613
	dld	6	20,30	43,06	63,36	52,0400	6,96711
	Sussidi	6	1	1	2	1,50	,548
	Vardom	6	4,60	-2,11	2,49	,8017	1,74158
	Validi (listwise)	6					
4	Areageo	5	0	1	1	1,00	,000
	Varpil	5	3,68	1,15	4,83	2,9680	1,50825
	Disoc	5	6,49	2,32	8,81	4,4340	2,64022
	dld	5	25,08	12,51	37,59	28,1780	9,51022
	Sussidi	5	2	1	3	1,60	,894
	Vardom	5	6,24	1,97	8,21	4,9920	2,51098
	Validi (listwise)	5					

Dalla lettura della tabella si possono trarre le seguenti conclusioni:

- a) il gruppo 1 è formato da paesi non europei che presentano, in media, un ridotto tasso di crescita del PIL, in alcuni casi addirittura negativo. Gli stessi mostrano un tasso di disoccupazione inferiore rispetto alla media dei paesi analizzati, condizionato in maniera ridotta dalla percentuale di disoccupati di lunga durata, il che indica che la maggiore incidenza sulla disoccupazione è data da soggetti che stanno cambiando lavoro. La spesa pubblica in sussidi di questi paesi è ridotta in rapporto al PIL.
- b) il gruppo 2 è costituito da paesi europei caratterizzati da un tasso di crescita del PIL discreto, comunque leggermente superiore ai paesi del gruppo 1 e sempre positivo, e da un tasso di disoccupazione ridotto, sul quale però incide in

maniera più rilevante la disoccupazione di lunga durata. La spesa pubblica in sussidi è abbastanza elevata rispetto ai paesi esaminati, ma il tasso di crescita della domanda interna di beni e servizi è abbastanza ridotto con alcune situazioni in cui lo stesso è negativo.

- c) I paesi appartenenti al gruppo 3 sono paesi europei con un elevato tasso di crescita del PIL nel 2001, ma un tasso di disoccupazione consistente, derivante anche da una forte incidenza della disoccupazione di lunga durata. La spesa pubblica in sussidi risulta ridotta rispetto al PIL e il tasso di crescita della domanda interna di beni e servizi è molto ridotto e in alcuni casi addirittura negativo.
- d) Il gruppo 4 è composto da paesi europei, aventi un consistente tasso di crescita del PIL e un tasso di disoccupazione ridotto, condizionato in discreta parte dalla disoccupazione di lunga durata. In questi paesi una percentuale non particolarmente rilevante della spesa pubblica viene destinata ai sussidi. Queste nazioni, infine, presentano un tasso di crescita della domanda di beni e servizi piuttosto consistente.

### 3.4 Un approccio non gerarchico: il metodo delle k-medie

Dopo aver individuato il numero di clusters ottimale (4 clusters) attraverso un metodo di agglomerazione gerarchico, si procede con l'implementazione di un processo non gerarchico, partendo dai risultati ottenuti, che consentirà di individuare l'esatta composizione dei gruppi e le caratteristiche principali degli stessi.

A tal fine, si sceglie di procedere con il metodo denominato "delle k-medie", utilizzando come dati di input, da cui sviluppare l'analisi, quelli relativi ai centroidi dei gruppi individuati attraverso il metodo di Ward, precedentemente analizzato.

#### 3.4.1 La convergenza del modello

Per verificare la convergenza del modello e, di conseguenza, la sua stabilità si prende in esame la cronologia delle iterazioni e le tabelle che riportano i centroidi dei clusters iniziali e finali, al fine di evidenziare eventuali differenze significative tra gli stessi.

**Tabella 9:** cronologia delle iterazioni – anno 2001

**Cronologia iterazioni<sup>a</sup>**

Iterazione	Modifiche ai centri dei cluster			
	1	2	3	4
1	5,057	1,437	1,861	1,957
2	,000	2,319	,000	3,246
3	1,431	1,697	,000	2,839
4	,000	,000	,000	,000

a. Convergenza ottenuta poiché nei centri dei cluster non sono presenti modifiche o sono presenti modifiche minime. La variazione massima assoluta delle coordinate per qualsiasi centro è ,000. L'iterazione corrente è 4. La distanza minima tra i centri iniziali è 4,382.

La tabella in esame mostra che il modello raggiunge la convergenza, si ottiene quindi una segmentazione significativa e stabile.

Una ulteriore conferma della stabilità del modello può essere ricercata prendendo in esame i centri dei cluster iniziali e finali.

**Tabella 10:** centri dei cluster iniziali – anno 2001

**Centri dei cluster iniziali**

	Cluster			
	1	2	3	4
Areageo	0	1	1	1
Varpil	1,61	1,80	2,94	2,97
Disoc	5,19	5,34	11,97	4,43
dld	12,03	28,70	52,04	28,18
Sussidi	2	3	2	2
Vardom	2,07	1,18	,80	4,99

Input dal sottocomando FILE

**Tabella 11:** centri dei cluster finali – anno 2001

**Centri dei cluster finali**

	Cluster			
	1	2	3	4
Areageo	0	1	1	1
Varpil	1,97	1,81	2,63	2,44
Disoc	3,89	5,12	10,72	4,84
dld	5,76	23,95	50,93	34,69
Sussidi	2	3	2	2
Vardom	1,73	2,34	,60	4,68

Il confronto tra le due tabelle suggerisce che sono presenti delle piccole differenze tra i centri dei cluster iniziali e finali. Questo indica che è stato possibile ridurre, anche se in maniera minima, la varianza all'interno dei gruppi, riassegnando alcuni oggetti.

La variazione tra i valori iniziali e finali è, in sostanza, poco considerevole, il che si pone a ulteriore conferma del fatto che il modello sia stabile e, di conseguenza, la segmentazione significativa.

### 3.4.2 La composizione dei clusters

Analizzando la tabella denominata “numero di casi in ogni cluster” si intuisce come i casi osservati siano stati suddivisi e si possono notare eventuali modifiche rispetto alla segmentazione ottenuta attraverso la procedura gerarchica.

**Tabella 12:** numero di casi in ogni cluster – anno 2001

**Numero di casi in ogni cluster**

Cluster	1	7,000
	2	9,000
	3	9,000
	4	4,000
Validi		29,000
Mancanti		,000

Dall'osservazione della tabella si può dedurre che:

- a) il gruppo 1, precedentemente formato da 7 elementi, ha mantenuto la stessa numerosità a seguito della nuova analisi.
- b) il gruppo 2, in precedenza composto da 11 casi, è ora formato da soli 9 casi.
- c) Il gruppo 3 ora è composto da 9 elementi, mentre prima era costituito da soli 6 casi.
- d) Il gruppo 4 comprende 4 stati, mentre prima al suo interno ne venivano rappresentati 5.

Bisogna tuttavia precisare che è molto probabile che i casi che compongono ogni gruppo siano significativamente differenti rispetto alla segmentazione precedentemente ottenuta.

Dopo aver identificato il numero di paesi presente in ogni gruppo, si procede con l'analisi della composizione degli stessi attraverso lo studio della tabella "cluster di appartenenza". Questa tabella, oltre a mostrare a quale gruppo è stato assegnato ciascun paese, nella colonna "distanza" riporta la distanza di ogni membro dal centro del proprio cluster, il che permette di identificare gli elementi più vicini ai centroidi e di conseguenza più rappresentativi del gruppo.

**Tabella 13:** *cluster di appartenenza – anno 2001*

Cluster di appartenenza			
Numero di caso	Nome	Cluster	Distanza
1	Australia	2	3,034
2	Austria	2	2,657
3	Belgio	3	5,364
4	Canada	1	5,088
5	Repubblica ceca	3	3,877
6	Danimarca	2	3,071
7	Finlandia	2	5,372
8	Francia	4	5,562
9	Germania	3	3,433
10	Grecia	3	2,701
11	Ungheria	3	6,866
12	Islanda	1	7,609
13	Irlanda	4	4,157
14	Italia	3	12,855
15	Giappone	2	3,525
16	Corea	1	4,105
17	Lussemburgo	2	5,727
18	Messico	1	5,856
19	Paesi Bassi	1	3,560
20	Nuova Zelanda	2	7,622
21	Norvegia	1	2,289
22	Polonia	3	11,337
23	Portogallo	4	4,638
24	Slovacchia	3	9,094
25	Spagna	3	7,372
26	Svezia	2	1,870
27	Svizzera	4	6,552
28	Regno unito	2	4,577
29	Stati Uniti	1	1,943

I dati riportati nella tabella forniscono le seguenti informazioni:

- a) il gruppo 1 è formato da Canada, Islanda, Corea, Messico, Paesi Bassi, Norvegia e Stati Uniti. Il paese più rappresentativo di questo gruppo sono gli Stati Uniti, con una distanza dal centro pari a 1,943. Lo stato più lontano e quindi meno rappresentativo, con una distanza di 7,609, è l'Islanda.

- b) il gruppo 2 è costituito da Australia, Austria, Danimarca, Finlandia, Giappone, Lussemburgo, Nuova Zelanda, Svezia e Regno Unito. Il paese più vicino al centro del cluster che meglio rappresenta il gruppo è, con una distanza di 1,870, la Svezia, mentre quello più lontano dal centro è la Nuova Zelanda distante 7,622.
- c) fanno parte del gruppo 3 Belgio, Repubblica Ceca, Germania, Grecia, Ungheria, Italia, Polonia, Slovacchia e Spagna. Il paese più vicino al centro è la Grecia, che dista 2,701, mentre il più distante è l'Italia con un valore pari a 12,855.
- d) il gruppo 4 comprende Francia, Irlanda, Portogallo e Svizzera. Tra questi, il più vicino è l'Irlanda (distanza pari a 4,157) mentre il più lontano, con una distanza di 6,552, è la Svizzera.

### 3.4.3 Le caratteristiche dei clusters

L'analisi delle caratteristiche dei cluster si focalizzerà principalmente su due aspetti: la vicinanza tra i gruppi creati e le caratteristiche peculiari di ciascuno di essi.

Si analizzerà d'apprima la tabella che riporta le distanze tra i centri dei cluster e, successivamente, quella indicante i centri finali dei cluster.

**Tabella 14:** *distanze tra i centri dei cluster finali – anno 2001*

**Distanze tra i centri dei cluster finali**

Cluster	1	2	3	4
1		18,261	45,707	29,107
2	18,261		27,623	11,044
3	45,707	27,623		17,751
4	29,107	11,044	17,751	

La tabella riportata sopra mostra la distanza euclidea tra i centroidi dei gruppi finali: chiaramente maggiore è tale distanza, maggiore sarà la dissomiglianza tra i 4 gruppi.

I dati suggeriscono che i gruppi sono tutti relativamente vicini, e quindi simili, tra loro. I due cluster più simili sono il numero 2 e 4, con una distanza tra i centri pari a 11,044, al contrario i più distanti, e quindi differenti, sono il 3 e 1, che presentano una distanza pari a 45,707.

Si procede ora allo studio delle caratteristiche specifiche di ciascun gruppo attraverso l'analisi della tabella che riporta i centri dei cluster finali.

**Tabella 15:** *centri dei cluster finali – anno 2001*

	Cluster			
	1	2	3	4
Areageo	0	1	1	1
Varpil	1,97	1,81	2,63	2,44
Disoc	3,89	5,12	10,72	4,84
dId	5,76	23,95	50,93	34,69
Sussidi	2	3	2	2
Vardom	1,73	2,34	,60	4,68

Dall'osservazione della tabella si può dedurre che:

- a) il gruppo 1 comprende paesi che presentano un tasso di crescita del PIL nel 2001 ridotto, ma un tasso di disoccupazione molto basso di cui una percentuale minima è prodotta da disoccupati di lunga durata. La percentuale di spesa pubblica in sussidi rispetto al PIL di questi paesi è bassa e il tasso di crescita della domanda interna è piuttosto ridotto.
- b) il gruppo 2 è formato da stati con un tasso di crescita del PIL molto ridotto, ma con un tasso di disoccupazione basso influenzato, in discreta parte, da disoccupati di lunga durata e una spesa pubblica in sussidi piuttosto rilevante. Il tasso di crescita della domanda interna si rivela discreto in questi paesi.
- c) nel gruppo 3 rientrano nazioni con un elevato tasso di crescita del PIL e un elevato tasso di disoccupazione, fortemente condizionato da casi di disoccupazione di lunga durata. In questi paesi la percentuale di spesa pubblica in sussidi è ridotta e il tasso di crescita della domanda interna di beni e servizi praticamente nullo.
- d) Gli elementi inclusi nel gruppo 4 sono caratterizzati da un discreto tasso di crescita del PIL, un basso tasso di disoccupazione, influenzato in parte da fenomeni di disoccupazione di lunga durata, e da una spesa pubblica in sussidi non molto rilevante. In questi paesi, contrariamente a quanto si rileva per gli altri gruppi, il tasso di crescita della domanda interna di beni e servizi è piuttosto elevato.

### 3.4.4 Le variabili che hanno determinato i clusters

A conclusione del procedimento di analisi, si propone la tabella ANOVA, che fornisce informazioni riguardanti la diversa influenza che hanno avuto le variabili nell'individuazione dei clusters. Le variabili che differenziano meglio i gruppi sono quelle che presentano un livello di significatività più basso ovvero un valore della F più elevato.

È opportuno ricordare che la procedura ANOVA di SPSS richiede i gruppi bilanciati e in questo caso non lo sono, quindi i risultati ottenuti avranno un'interpretazione puramente descrittiva.

**Tabella 16:** *anova - anno 2001*

ANOVA						
	Cluster		Errore		F	Sig.
	Media dei quadrati	df	Media dei quadrati	df		
Areageo	,532	3	,149	25	3,581	,028
Varpil	1,196	3	1,855	25	,644	,594
Disoc	78,781	3	10,858	25	7,255	,001
dld	2820,821	3	22,080	25	127,752	,000
Sussidi	,918	3	1,015	25	,904	,453
Vardom	15,980	3	3,741	25	4,271	,014

I test F devono essere utilizzati solo per motivi descrittivi poiché i cluster sono stati scelti per ottimizzare le differenze tra i casi in diversi cluster. I livelli di significatività osservati non sono perciò corretti e non possono quindi essere interpretati come test dell'ipotesi che le medie dei cluster siano uguali.

Analizzando la tabella sopra esposta, si può dedurre che **DLD** e **DISOC** siano le variabili significativamente associate ai clusters individuati, seguite da **VARDOM** e **AREAGEO**. **VARPIL** e **SUSSIDI** risultano invece le meno influenti nella divisione in gruppi così ottenuta.

## 4. Studio di un caso: i paesi membri dell'OECD – anno 2007

### 4.1 Fonte dei dati – anno 2007

I dati analizzati sono reperibili presso la sezione Statistics Portal del sito dell'OECD, Organization for Economic Cooperation and Development, ([www.oecd.org](http://www.oecd.org)). Essi fanno riferimento all'anno 2007.

**Tabella 17:** *dataset analizzato – anno 2007*

PAESE	AREAGEO	VARPIL	DISOC	D.L.D.	VARDIS	SUSSIDI	VARDOM
Australia	no	4,72	4,37	15,37	5,14	m. bassi	7,20
Austria	si	3,69	4,40	26,83	4,32	bassi	2,50
Belgio	si	2,83	7,45	50,36	7,95	alti	2,70
Canada	no	2,20	6,02	7,44	7,10	bassi	3,90
Repubblica Ceca	si	5,73	5,31	53,38	6,84	bassi	6,60
Danimarca	si	1,58	3,79	16,21	5,00	alti	2,30
Finlandia	si	5,33	6,82	22,97	8,24	alti	4,70
Francia	si	2,23	7,98	40,24	8,49	medi	3,10
Germania	si	3,39	8,63	56,59	7,95	medi	2,00
Grecia	si	2,99	8,27	49,98	9,86	m. bassi	6,10
Ungheria	si	,11	7,36	47,54	8,12	bassi	-1,30
Islanda	si	5,98	2,30	7,97	3,70	m. bassi	-,40
Irlanda	si	5,17	4,64	29,51	7,59	bassi	4,50
Italia	si	1,68	6,09	47,34	7,42	m. bassi	1,20
Giappone	no	2,36	3,88	31,98	4,15	m. bassi	1,30
Korea	no	5,10	3,23	,56	3,52	m. bassi	4,70
Lussemburgo	si	6,63	4,06	28,66	4,56	bassi	5,90
Messico	no	3,24	3,39	2,71	4,11	m. bassi	3,70
Olanda	si	3,90	3,57	39,37	3,77	medi	3,20
Nuova Zelanda	no	3,40	3,67	6,09	4,26	m. bassi	4,80
Norvegia	si	2,73	2,52	8,76	3,31	m. bassi	5,00
Polonia	si	6,83	9,60	45,86	11,31	m. bassi	8,70
Portogallo	si	2,38	7,98	47,14	8,09	bassi	2,00
Slovacchia	si	10,49	11,01	70,81	13,53	m. bassi	6,30
Spagna	si	3,47	8,26	27,61	12,59	alti	4,10
Svezia	si	3,42	6,06	12,95	7,30	bassi	4,70
Svizzera	si	3,64	3,64	40,78	3,74	bassi	1,40
Regno Unito	si	3,46	5,23	23,72	5,96	m. bassi	3,40
Stati Uniti	no	1,91	4,62	9,95	5,70	m. bassi	1,20

#### *Legenda:*

**PAESE:** nome degli stati appartenenti all'Oecd

**AREAGEO:** è uno stato europeo?

**VARPIL:** tasso di crescita del pil nel 2001

**DISOC:** tasso di disoccupazione nel 2001

**D.L.D.:** incidenza della disoccupazione di lunga durata (oltre 12 mesi) sul totale dei disoccupati (%)

**VARDIS:** tasso di crescita della disoccupazione nel 2001

**SUSSIDI:** spesa pubblica in sussidi per la disoccupazione in rapporto al pil (%)

(*molto bassi:* sussidi/pil <0,5 ; *bassi:* 0,5<= sussidi/pil <1; *medi:* 1<=sussidi/pil <1,5; *alti:* sussidi/pil =>1,5)

**VARDOM:** tasso di crescita della domanda interna di beni e servizi

Ai fini dell'analisi, la variabile **AREAGEO**, di tipo dicotomico, viene riclassificata ponendo la modalità NO pari a 0 e la modalità SI pari a 1.

Seguendo lo stesso principio, la modalità **SUSSIDI**, rilevata su scala ordinale, viene riclassificata sostituendo le modalità molto bassi, bassi, medi, alti con i valori 1, 2, 3, 4.

## 4.2 Analisi preliminare

Prima di iniziare la cluster analysis, si effettuano alcune analisi sulle variabili che verranno utilizzate: in particolare, si analizzerà la tabella di correlazione di Pearson, come è stato fatto per i dati riferiti all'anno 2001, al fine di individuare la presenza di eventuali correlazioni elevate tra le variabili.

Nel caso di elevata correlazione, le variabili in questione verranno escluse dall'analisi, con lo scopo di evitare che l'elevata interdipendenza distorca eccessivamente i risultati della stessa.

**Tabella 18:** correlazioni di Pearson – anno 2007

		Correlazioni						
		Areageo	Varpil	Disoc	dld	Vardis	Sussidi	Vardom
Areageo	Correlazione di Pearson	1	,151	,377*	,587**	,381*	,424*	-,048
	Sig. (2-code)		,434	,044	,001	,042	,022	,805
	N	29	29	29	29	29	29	29
Varpil	Correlazione di Pearson	,151	1	,192	,204	,272	-,186	,602**
	Sig. (2-code)	,434		,318	,288	,153	,335	,001
	N	29	29	29	29	29	29	29
Disoc	Correlazione di Pearson	,377*	,192	1	,717**	,940**	,235	,226
	Sig. (2-code)	,044	,318		,000	,000	,220	,239
	N	29	29	29	29	29	29	29
dld	Correlazione di Pearson	,587**	,204	,717**	1	,611**	,193	,027
	Sig. (2-code)	,001	,288	,000		,000	,315	,888
	N	29	29	29	29	29	29	29
Vardis	Correlazione di Pearson	,381*	,272	,940**	,611**	1	,233	,299
	Sig. (2-code)	,042	,153	,000	,000		,223	,115
	N	29	29	29	29	29	29	29
Sussidi	Correlazione di Pearson	,424*	-,186	,235	,193	,233	1	-,147
	Sig. (2-code)	,022	,335	,220	,315	,223		,447
	N	29	29	29	29	29	29	29
Vardom	Correlazione di Pearson	-,048	,602**	,226	,027	,299	-,147	1
	Sig. (2-code)	,805	,001	,239	,888	,115	,447	
	N	29	29	29	29	29	29	29

\*. La correlazione è significativa al livello 0,05 (2-code).

\*\*.. La correlazione è significativa al livello 0,01 (2-code).

Arbitrariamente, si decide di stabilire il valore massimo di Pearson accettato affinché le variabili vengano utilizzate nell'analisi pari a  $|0,900|$ , per corrispondenza con l'analisi dei dati riferiti all'anno 2001.

Un valore così notevole infatti sottolineerebbe un'elevata correlazione tra le variabili.

Poiché la variabile **VARDIS** è positivamente correlata con la variabile **DISOC** con un valore, significativo, pari a 0,940, la stessa verrà esclusa dall'analisi. Le restanti

variabili non presentano valori eccessivamente elevati di correlazione, pertanto verranno tutte impiegate nel procedimento.

È interessante notare come le variabili **VARDIS** e **DISOC** siano le medesime che sono state escluse dall'analisi facendo riferimento ai dati del 2001.

Infine, è importante precisare che, anche se le due variabili non avessero presentato una correlazione di tale entità, le stesse sarebbero state in ogni caso estromesse per permettere una comparazione tra i risultati riferiti ai diversi anni.

Dopo aver stabilito quali variabili utilizzare, mediante le statistiche descrittive delle stesse si possono fare alcune considerazioni preliminari sul modo in cui queste influenzeranno l'analisi e la creazione dei clusters.

**Tabella 19:** *statistiche descrittive – anno 2007*

**Statistiche descrittive**

	N	Intervallo	Minimo	Massimo	Media	Deviazione std.	Varianza
Areageo	29	1	0	1	,76	,435	,190
Varpil	29	10,38	,11	10,49	3,8134	2,04152	4,168
Disoc	29	8,71	2,30	11,01	5,6603	2,27361	5,169
dld	29	70,25	,56	70,81	29,9545	18,92780	358,262
Sussidi	29	3	1	4	1,93	1,067	1,138
Vardom	29	10,00	-1,30	8,70	3,6379	2,28261	5,210
Validi (listwise)	29						

La tabella mostra che la variabile **DLD** è quella che presenta una maggiore deviazione standard con un valore pari a 18,92780. Ciò suggerisce che è in questa variabile che si presenta una maggiore dispersione dei paesi esaminati e quindi è, in linea di principio, ipotizzabile che la stessa contribuisca in maniera rilevante alla formazione dei gruppi. Anche in questo caso, come è accaduto nell'analisi riferita ai dati del 2001, sarà la variabile **DLD** a incidere maggiormente sui risultati dell'analisi.

A seguire, si possono identificare la variabile **VARDOM**, con un valore di standard deviation pari a 2,28261, e **DISOC**, che presenta una deviazione standard uguale a 2,27361, con livelli inferiori di deviazione standard rispetto a **DLD**, ma comunque significativi.

### 4.3 Un apporccio gerarchico – il metodo di Ward

Dopo aver verificato che, a causa dell'elevata correlazione tra **DISOC** e **VARDIS**, la seconda verrà esclusa dal procedimento, si procede ora con un'analisi basata sul metodo di Ward, per simmetria rispetto all'analisi effettuata sui dati del 2001. La distanza utilizzata sarà la distanza euclidea quadratica e i valori iniziali, in quanto osservati su scale differenti, verranno standardizzati nei valori Z (sottraendo agli stessi la loro media e dividendo per lo scarto quadratico), in modo da eliminare eventuali distorsioni derivanti dalle diverse tipologie di misurazione.

Nello specifico, saranno oggetto di analisi il programma di agglomerazione, l'icicle diagram, il dendrogramma e, in ultimo, le caratteristiche peculiari dei clusters formati.

#### 4.3.1 Il programma di agglomerazione

Il programma di agglomerazione fornisce informazioni circa gli elementi e i gruppi combinati ad ogni stadio e la distanza a cui si verifica l'unione. Esso rappresenta una sintesi numerica del processo di agglomerazione.

Il cluster risultante dall'unione di due elementi prenderà il nome dell'elemento indicato nella seconda colonna nominata "cluster 1".

L'ultima colonna sulla destra, denominata "Stadio successivo", indica in quale stadio il cluster formato verrà nuovamente accorpato ad altri elementi.

Le due colonne riguardanti lo stadio di formazione del cluster indicano, rispettivamente, a quale stadio si è formato il primo cluster e a quale stadio si è formato il secondo cluster che vengono uniti allo stadio indicato nella prima colonna. Tali colonne riporteranno un valore pari a zero fino a quando non verranno accorpate due gruppi di cui almeno uno costituito da più di un elemento.

**Tabella 20:** *programma di agglomerazione – anno 2007*

**Programma di agglomerazione**

Stadio	Cluster accorpati		Coefficienti	Stadio di formazione del cluster		Stadio successivo
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	18	20	,143	0	0	4
2	2	27	,587	0	0	10
3	13	17	1,064	0	0	12
4	16	18	1,636	0	1	14
5	8	23	2,260	0	0	11
6	7	25	2,941	0	0	19
7	3	9	3,653	0	0	11
8	15	29	4,409	0	0	16
9	26	28	5,239	0	0	15
10	2	19	6,118	2	0	18
11	3	8	7,428	7	5	19
12	5	13	8,900	0	3	23
13	11	14	10,391	0	0	24
14	1	16	12,079	0	4	20
15	21	26	13,852	0	9	21
16	4	15	16,081	0	8	20
17	10	22	18,694	0	0	23
18	2	6	22,160	10	0	22
19	3	7	27,298	11	6	24
20	1	4	32,901	14	16	28
21	12	21	38,510	0	15	22
22	2	12	44,669	18	21	26
23	5	10	51,275	12	17	25
24	3	11	61,148	19	13	26
25	5	24	71,914	23	0	27
26	2	3	90,348	22	24	27
27	2	5	120,623	26	25	28
28	1	2	168,000	20	27	0

Osservando i dati della tabella si possono evincere alcune considerazioni:

- a) l'ordine di aggregazione dipende dalla distanza tra i casi, quindi si aggrediranno per primi i casi più simili tra loro. Nel caso esaminato, i casi più simili tra loro sono il 18 e 20, seguiti dagli elementi 2 e 27 e, successivamente, dal 13 e 17.
- b) i primi due elementi che si fondono in un cluster sono il 18 e il 20 (rispettivamente Messico e Nuova Zelanda) che si raggruppano ad un livello di distanza pari a 0,143; se la distanza di aggregazione fosse stata pari a 0,000, questo avrebbe significato che i due elementi presentavano lo stesso valore per tutte le variabili analizzate.

Il fatto che la distanza di aggregazione, pari a 0,143, sia inferiore rispetto a quella riscontrata nel 2001, pari a 0,408, indica che i due paesi più simili nel 2007 (Messico e Nuova Zelanda) siano più simili tra loro rispetto a quanto lo fossero i due più simili tra loro nel 2001 (Canada e Stati Uniti).

Il gruppo, così formato tra gli elementi 18 e 20, d'ora in avanti verrà identificato con il numero del primo elemento cioè 18.

- c) Il gruppo formato da questi due elementi verrà raggruppato nuovamente allo stadio 4 (colonna "stadio successivo") con l'elemento 16 (Korea). Questo stadio è quello in cui, per la prima volta, si fondono due gruppi di cui almeno uno comprende più di un elemento.
- d) Per i primi 3 stadi, si sono formati clusters che accorpano tra loro solo elementi singoli. Il cluster formato allo stadio 4 verrà identificato come 16 e verrà unito, allo stadio 14, con l'elemento 1 (Australia).

Rispetto al 2001, nel 2007 si nota come il processo di aggregazione porti più rapidamente alla creazione di un gruppo formato da più di 2 elementi infatti, nel 2001, si è dovuti giungere allo stadio 8 per assistere alla formazione di un gruppo composto da più di 2 elementi.

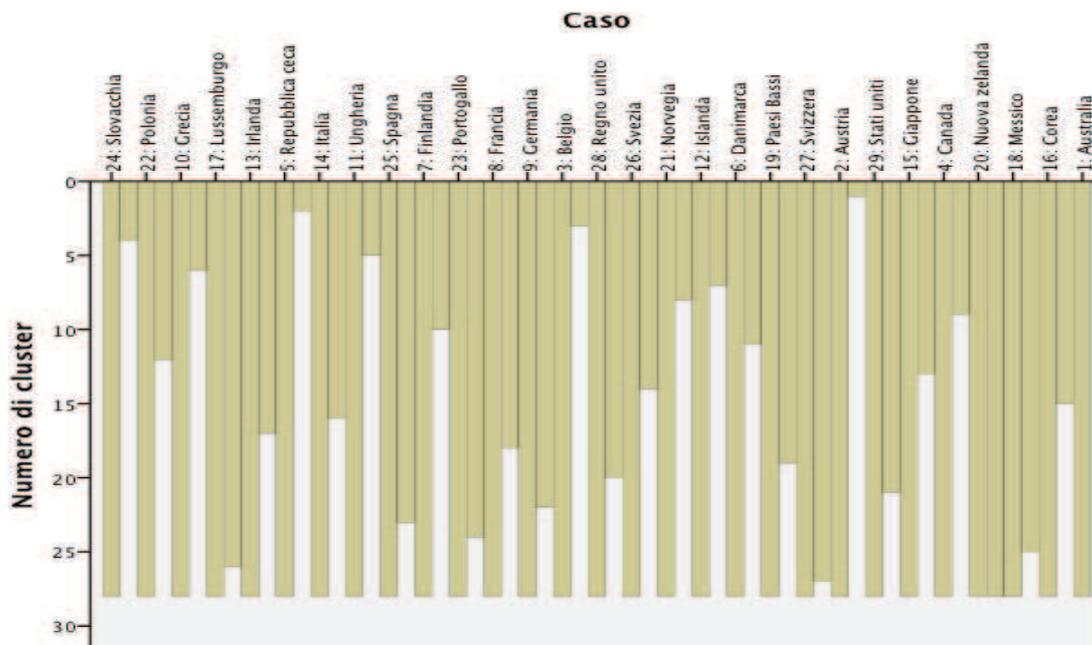
- e) L'ultimo stadio, il numero 28, vedrà l'accorparsi dei cluster 1 e 2. Questi si uniranno a un livello di distanza pari a 168,000 e deriveranno dall'unione avvenuta, rispettivamente, allo stadio 20 tra i cluster 1 e l'elemento 4 (Canada) e, allo stadio 27, tra i gruppi 2 e 5.

### 4.3.2 Icicle diagram

L'icicle diagram fornisce, graficamente, le stesse informazioni individuate nel programma di agglomerazione.

Il diagramma va letto dal basso verso l'alto, dove le colonne indicano gli elementi che vengono via via accorpati e le righe rappresentano il numero di gruppi.

**Grafico 5:** icicle diagram – anno 2007



Esaminando il grafico, si possono fare le seguenti riflessioni:

- i primi due elementi che si fondono sono 18 e 20 (Messico e Nuova Zelanda) nella parte destra del grafico. Il numero di cluster rimanenti a seguito dell'unione è pari a 28.
- Successivamente agli elementi 18 e 20, vengono uniti gli elementi 2 e 27 (Austria e Svizzera) riducendo il numero di gruppi da 28 a 27.
- Fino al momento in cui il gruppo formato dagli elementi 18 e 20 viene unito con l'elemento 16, diminuendo a 25 il numero di gruppi, non vi sono unioni tra gruppi costituiti da più di un elemento. Questo momento viene identificato graficamente dal fatto che sono presenti più di 3 colonne color bronzo adiacenti (nello specifico, le 5 colonne adiacenti indicano la presenza di 3 paesi nel gruppo).

In sintesi, il grafico deve essere letto dal basso verso l'alto salendo di "un gradino" alla volta; in questo modo si identificheranno, di volta in volta, gli elementi che si fondono e il numero di cluster risultante dopo l'unione.

### 4.3.3 La scelta del numero di gruppi

Per identificare il numero di gruppi più sensato in cui raggruppare gli elementi del dataset analizzato si possono utilizzare svariati metodi.

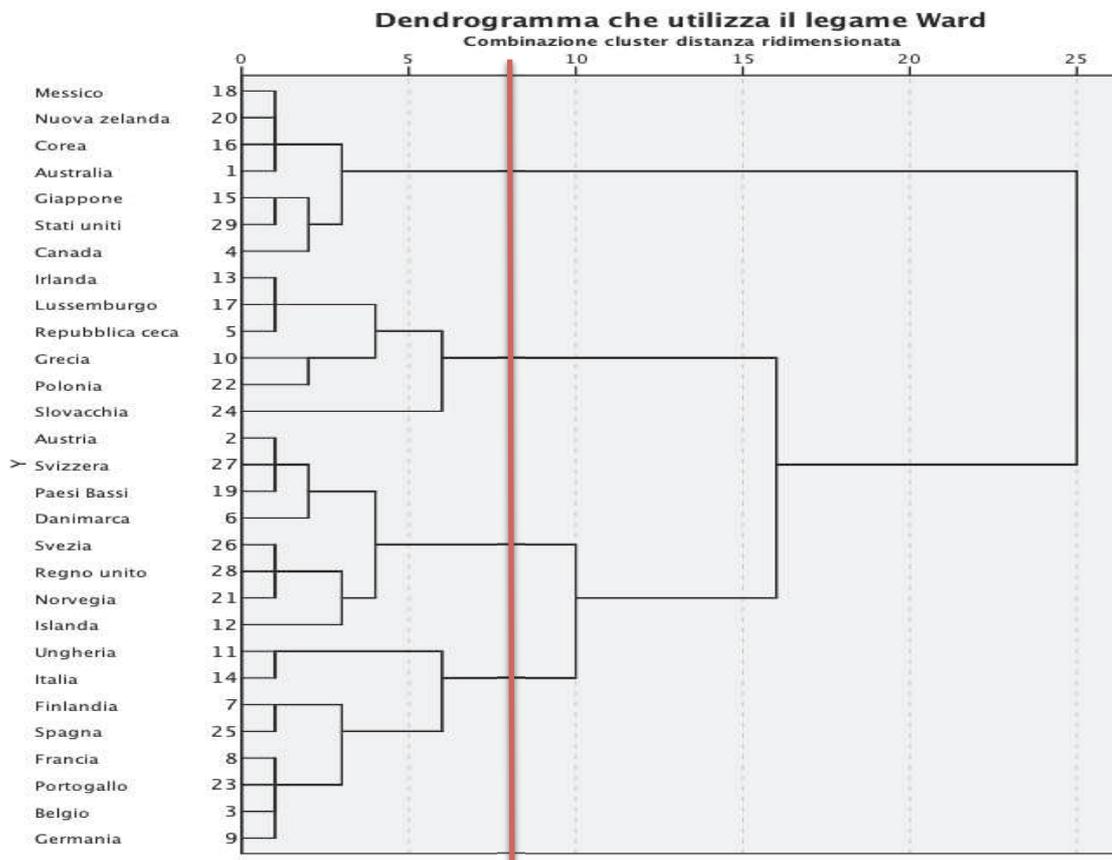
In questo contesto, per simmetria rispetto all'analisi effettuata sui dati riferiti all'anno 2001, si prenderanno in esame il dendrogramma, lo scree-plot della distanza di fusione e la tabella che riporta i coefficienti, la distanza di fusione riscalata e la variazione della stessa nel programma di agglomerazione.

#### 4.3.3.1 Il dendrogramma

Il dendrogramma rappresenta una sintesi grafica del risultato ottenuto dall'analisi gerarchica.

Nell'asse verticale di sinistra si leggono gli oggetti presenti nell'analisi, l'asse orizzontale mostra la distanza tra i cluster nel momento in cui questi vengono uniti. L'albero fornisce vari livelli di aggregazione: la scelta del livello a cui "tagliare" l'albero deve rappresentare un giusto compromesso tra numero di gruppi e omogeneità degli stessi. Generalmente, il taglio va fatto prima delle aggregazioni corrispondenti a salti molto grossi tra i valori dell'indice, considerato che un salto elevato nella distanza di aggregazione indicherebbe l'unione di due gruppi molto differenti tra loro.

**Grafico 6:** dendrogramma – anno 2007



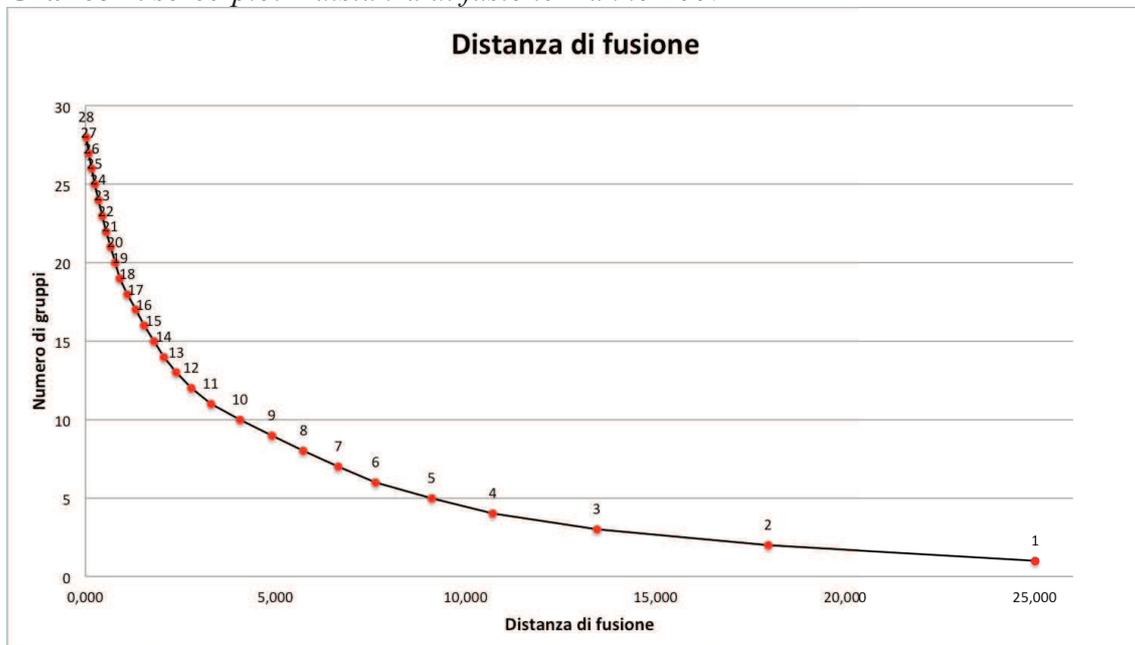
Il grafico presenta un salto consistente nella distanza di fusione nel passaggio da 4 a 3 clusters, quindi si ritiene opportuno “tagliare” in corrispondenza di 4 gruppi.

I risultati ottenuti in questo frangente sono uguali a quelli riscontrati nell’analisi relativa al 2001. Questo implica che, nel passaggio dal 2001 al 2007, nonostante possa modificarsi la composizione dei vari gruppi, la scelta ottimale appare quella di frazionare il dataset in 4 gruppi.

#### 4.3.3.2 Lo scree-plot della distanza di fusione

Un ulteriore metodo grafico è rappresentato dallo scree-plot della distanza di fusione, diagramma in cui viene posto il numero di gruppi in ordinata ed in ascissa la distanza di fusione.

**Grafico 7:** scree-plot – distanza di fusione – anno 2007



Il grafico, anche in questo caso, suggerisce di operare una suddivisione in 4 gruppi. Infatti, nel passaggio da 4 a 3 clusters, si registra un consistente aumento della distanza di fusione.

#### 4.3.3.3 L'incremento relativo della distanza di fusione

L'incremento relativo della distanza di fusione consente di operare una scelta ottimale del numero di gruppi attraverso un criterio numerico.

Si decide, infatti, di terminare il processo di agglomerazione nel momento in cui, raggiunto un numero di clusters abbastanza ridotto in proporzione alla numerosità della popolazione di analisi, l'unione di ulteriori due gruppi coincide con il valore massimo dell'incremento relativo della distanza di fusione. Si opera questa scelta in quanto la

continuazione del processo di agglomerazione implicherebbe l'unione di due clusters molto lontani (e quindi diversi) tra loro.

**Tabella 21:** *distanza di fusione – anno 2007*

STADIO	NUMERO DI GRUPPI	COEFFICIENTI	DISTANZA DI FUSIONE	INCREMENTO RELATIVO
1	28	0,143	0,021	0,000
2	27	0,587	0,087	3,105
3	26	1,064	0,158	0,813
4	25	1,636	0,243	0,538
5	24	2,260	0,336	0,381
6	23	2,941	0,438	0,301
7	22	3,653	0,544	0,242
8	21	4,409	0,656	0,207
9	20	5,239	0,780	0,188
10	19	6,118	0,910	0,168
11	18	7,428	1,105	0,214
12	17	8,900	1,324	0,198
13	16	10,391	1,546	0,168
14	15	12,079	1,797	0,162
15	14	13,852	2,061	0,147
16	13	16,081	2,393	0,161
17	12	18,694	2,782	0,162
18	11	22,160	3,298	0,185
19	10	27,298	4,062	0,232
20	9	32,901	4,896	0,205
21	8	38,510	5,731	0,170
22	7	44,669	6,647	0,160
23	6	51,275	7,630	0,148
24	5	61,148	9,099	0,193
25	4	71,914	10,701	0,176
26	3	90,348	13,445	0,256
27	2	120,623	17,950	0,335
28	1	168,000	25,000	0,393

La tabella mostra che uno dei livelli maggiori di incremento relativo, escludendo i primi stadi dove, ad esempio allo stadio 2, si registra un incremento molto elevato, ma il numero di gruppi derivanti risulta eccessivo e poco significativo (27 gruppi) si ha allo stadio 26, nella riduzione da 4 a 3 clusters, con un valore pari a 0,256. Questo valore suggerisce, come già constatato attraverso i metodi grafici, che la divisione ottimale è quella in 4 gruppi.

I livelli di incremento relativo degli stadi 27 e 28 sono superiori a quelli del livello 26, ma, giunti a questi livelli, si sarebbero già uniti elementi troppo diversi tra loro. Anche in alcuni stadi precedenti il 26 si hanno incrementi abbastanza simili al valore 0,259, come ad esempio agli stadi 19 e 20, ma si ritiene che, dato il ridotto numero di paesi analizzati, una divisione in 10 o 9 gruppi risulterebbe troppo dispersiva.

#### 4.3.4 Le caratteristiche dei gruppi

Dopo aver identificato il numero ottimale di gruppi in cui suddividere la popolazione, si procede ad una analisi degli elementi che compongono ciascun gruppo e delle caratteristiche peculiari di ciascuno dei clusters formati.

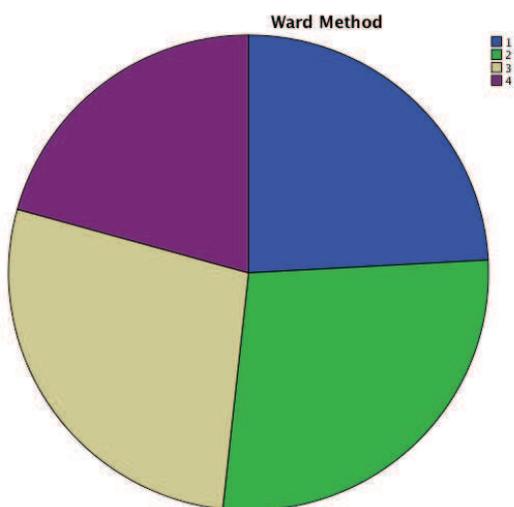
##### 4.3.4.1 La numerosità dei gruppi

Come prima operazione, si procede all'identificazione della numerosità dei 4 clusters, che devono essere equilibrati tra loro e avere una dimensione sufficientemente grande.

**Tabella 22:** numerosità dei cluster – anno 2007

Ward Method					
		Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Validi	1	7	24,1	24,1	24,1
	2	8	27,6	27,6	51,7
	3	8	27,6	27,6	79,3
	4	6	20,7	20,7	100,0
Totale		29	100,0	100,0	

**Grafico 8:** grafico a torta – numerosità dei cluster – anno 2007



Il grafico a torta e la tabella di numerosità individuano una sostanziale uniformità nella divisione dei paesi nei 4 gruppi. Si rilevano, infatti, 2 gruppi di identiche dimensioni, il gruppo 2 e il gruppo 3, costituiti entrambi da 8 elementi, e 2 gruppi di dimensioni leggermente inferiori, il gruppo 1 e il gruppo 4, comprendenti rispettivamente 7 e 6 paesi.

I gruppi 2 e 3 includono entrambi circa il 27% dei paesi analizzati, mentre i paesi restanti si suddividono in maniera abbastanza equa tra il gruppo 1 e 4, che comprendono, rispettivamente, circa il 24 e il 20% dei casi.

#### 4.3.4.2 Gli elementi nei gruppi

A seguito dell'identificazione della numerosità dei gruppi formati, si procede all'individuazione degli elementi che costituiscono ciascun cluster attraverso l'analisi della tabella che riporta il cluster di appartenenza per ciascun caso.

**Tabella 23:** cluster di appartenenza – anno 2007

Caso	4 cluster
1:Australia	1
2:Austria	2
3:Belgio	3
4:Canada	1
5:Repubblica ceca	4
6:Danimarca	2
7:Finlandia	3
8:Francia	3
9:Germania	3
10:Grecia	4
11:Ungheria	3
12:Islanda	2
13:Irlanda	4
14:Italia	3
15:Giappone	1
16:Corea	1
17:Lussemburgo	4
18:Messico	1
19:Paesi Bassi	2
20:Nuova zelandia	1
21:Norvegia	2
22:Polonia	4
23:Portogallo	3
24:Slovacchia	4
25:Spagna	3
26:Svezia	2
27:Svizzera	2
28:Regno unito	2
29:Stati uniti	1

Dai dati riportati sulla tabella, si può evincere che i 4 cluster sono composti ciascuno dai seguenti paesi:

- 1) cluster 1: Australia, Canada, Giappone, Corea, Messico, Nuova Zelanda, Stati Uniti.
- 2) cluster 2: Austria, Danimarca, Islanda, Paesi Bassi, Norvegia, Svezia, Svizzera, Regno Unito.
- 3) cluster 3: Belgio, Finlandia, Francia, Germania, Ungheria, Italia, Portogallo, Spagna.
- 4) cluster 4: Repubblica Ceca, Grecia, Irlanda, Lussemburgo, Polonia, Slovacchia.

Confrontando i dati ottenuti con quelli relativi all'anno 2001, si può rilevare che il cluster 1 rimane invariato nella sua composizione, mentre gli elementi che precedentemente costituivano il gruppo 2 si allocano, nel 2007, prevalentemente nei clusters 2 e 3.

Inoltre, va notato che il cluster 4 comprende paesi diversi rispetto al 2001 con la sola eccezione dell'Irlanda e che gli stessi si sono allocati, con l'eccezione della Francia, in blocco nel gruppo 2.

Infine, si rileva che tutti i paesi che precedentemente contribuivano a formare il gruppo 3 e che, nel 2007, hanno modificato la loro posizione sono adesso compresi nel gruppo 4.

#### 4.3.4.3 Le specifiche dei gruppi

A conclusione dell'analisi gerarchica, si confrontano i 4 cluster identificati per conoscerne le caratteristiche. Per questo scopo, si producono le statistiche descrittive delle variabili utilizzate nell'analisi.

**Tabella 24:** caratteristiche dei gruppi – anno 2007

Statistiche descrittive

Ward Method	N	Intervallo	Minimo	Massimo	Media	Deviazione std.	Varianza
1 Areageo	7	0	0	0	,00	,000	,000
Varpil	7	3,19	1,91	5,10	3,2757	1,24443	1,549
Disoc	7	2,79	3,23	6,02	4,1686	,95665	,915
dld	7	31,42	,56	31,98	10,5857	10,59605	112,276
Sussidi	7	1	1	2	1,14	,378	,143
Vardom	7	6,00	1,20	7,20	3,8286	2,09739	4,399
Validi (listwise)	7						
2 Areageo	8	0	1	1	1,00	,000	,000
Varpil	8	4,40	1,58	5,98	3,5500	1,23250	1,519
Disoc	8	3,76	2,30	6,06	3,9388	1,27198	1,618
dld	8	32,81	7,97	40,78	22,0737	12,91850	166,888
Sussidi	8	3	1	4	2,00	1,069	1,143
Vardom	8	5,40	-,40	5,00	2,7625	1,75250	3,071
Validi (listwise)	8						
3 Areageo	8	0	1	1	1,00	,000	,000
Varpil	8	5,22	,11	5,33	2,6775	1,51519	2,296
Disoc	8	2,54	6,09	8,63	7,5713	,82352	,678
dld	8	33,62	22,97	56,59	42,4738	11,58139	134,129
Sussidi	8	3	1	4	2,88	1,126	1,268
Vardom	8	6,00	-1,30	4,70	2,3125	1,85737	3,450
Validi (listwise)	8						
4 Areageo	6	0	1	1	1,00	,000	,000
Varpil	6	7,50	2,99	10,49	6,3067	2,46996	6,101
Disoc	6	6,95	4,06	11,01	7,1483	2,87716	8,278
dld	6	42,15	28,66	70,81	46,3667	15,85969	251,530
Sussidi	6	1	1	2	1,50	,548	,300
Vardom	6	4,20	4,50	8,70	6,3500	1,36198	1,855
Validi (listwise)	6						

Dalla lettura della tabella si possono trarre le seguenti conclusioni:

- a) il gruppo 1 è formato da paesi non europei che presentano, in media, un discreto tasso di crescita del PIL. Gli stessi mostrano un tasso di disoccupazione inferiore rispetto alla media dei paesi analizzati, condizionato in maniera ridotta dalla percentuale di disoccupati di lunga durata, il che indica che la maggiore incidenza sulla disoccupazione è data da soggetti che stanno cambiando lavoro. La spesa pubblica in sussidi di questi paesi è ridotta in rapporto al PIL mentre siamo in presenza di un discreto tasso di crescita della domanda interna di beni e servizi.
- b) il gruppo 2 è costituito da paesi europei caratterizzati da un tasso di crescita del PIL discreto, comunque leggermente superiore ai paesi del gruppo 1, e da un tasso di disoccupazione ridotto, sul quale però incide in maniera più rilevante la disoccupazione di lunga durata. La spesa pubblica in sussidi è abbastanza elevata rispetto ai paesi esaminati, ma il tasso di crescita della domanda interna di beni e servizi è abbastanza ridotto con alcune situazioni in cui lo stesso è negativo.
- c) I paesi appartenenti al gruppo 3 sono paesi europei con un basso tasso di crescita del PIL nel 2007 e un tasso di disoccupazione consistente, influenzato in maniera rilevante dalla disoccupazione di lunga durata. La spesa pubblica in sussidi risulta elevata rispetto al PIL e il tasso di crescita della domanda interna di beni e servizi è molto ridotto e in alcuni casi addirittura negativo.
- d) Il gruppo 4 è composto da paesi europei, aventi un tasso di crescita del PIL molto consistente a cui si accosta un tasso di disoccupazione elevato, condizionato fortemente dalla disoccupazione di lunga durata. In questi paesi una percentuale non particolarmente rilevante della spesa pubblica viene destinata ai sussidi. Queste nazioni, infine, presentano un tasso di crescita della domanda di beni e servizi piuttosto consistente.

#### 4.4 Un approccio non gerarchico: il metodo delle k-medie

Dopo aver individuato il numero di clusters ottimale (4 clusters) attraverso un metodo di agglomerazione gerarchico, come per i dati del 2011, si procede con l'implementazione di un processo non gerarchico, partendo dai risultati ottenuti, che consentirà di individuare l'esatta composizione dei gruppi e le caratteristiche principali degli stessi.

A tal fine, si procede con il metodo denominato "delle k-medie", utilizzando come dati di input, da cui sviluppare l'analisi, quelli relativi ai centroidi dei gruppi individuati attraverso il metodo di Ward, precedentemente analizzato.

##### 4.4.1 La convergenza del modello

Per verificare la convergenza del modello e, di conseguenza, la sua stabilità si analizza la cronologia delle iterazioni e le tabelle che riportano i centroidi dei clusters iniziali e finali, al fine di evidenziare eventuali differenze significative tra gli stessi.

**Tabella 25:** cronologia delle iterazioni – anno 2007

**Cronologia iterazioni<sup>a</sup>**

Iterazione	Modifiche ai centri dei cluster			
	1	2	3	4
1	2,659	4,179	2,209	8,344
2	,909	1,471	1,346	1,898
3	,000	,000	,836	1,720
4	,000	,000	,725	2,745
5	,000	,000	,000	,000

a. Convergenza ottenuta poiché nei centri dei cluster non sono presenti modifiche o sono presenti modifiche minime. La variazione massima assoluta delle coordinate per qualsiasi centro è ,000. L'iterazione corrente è 5. La distanza minima tra i centri iniziali è 6,833.

La tabella in esame mostra che, anche per i dati relativi all'anno 2007, il modello raggiunge la convergenza, si ottiene quindi una segmentazione significativa e stabile.

Una ulteriore conferma della stabilità del modello può essere ricercata prendendo in esame i centri dei cluster iniziali e finali.

**Tabella 26:** centri dei cluster iniziali – anno 2007

**Centri dei cluster iniziali**

	Cluster			
	1	2	3	4
Areageo	0	1	1	1
Varpil	3,28	3,55	2,68	6,31
Disoc	4,17	3,94	7,57	7,15
dld	10,59	22,07	42,47	46,37
Sussidi	1	2	3	2
Vardom	3,83	2,76	2,31	6,35

Input dal sottocomando FILE

**Tabella 27:** *centri dei cluster finali – anno 2007***Centri dei cluster finali**

	Cluster			
	1	2	3	4
Areageo	0	1	1	1
Varpil	3,43	4,30	2,95	6,54
Disoc	4,00	5,33	6,88	8,32
dld	8,80	27,33	45,40	60,26
Sussidi	2	2	2	2
Vardom	3,71	3,77	3,01	4,97

Il confronto tra le due tabelle suggerisce che sono presenti delle piccole differenze tra i centri dei cluster iniziali e finali. Questo indica che è stato possibile ridurre, anche se in maniera minima, la varianza all'interno dei gruppi, riassegnando alcuni oggetti.

La variazione tra i valori iniziali e finali è, in sostanza, poco considerevole, il che si pone a ulteriore conferma del fatto che il modello sia stabile e, di conseguenza, la segmentazione significativa.

#### 4.4.2 La composizione dei clusters

Analizzando la tabella denominata “numero di casi in ogni cluster” si intuisce come i casi osservati siano stati suddivisi e si possono notare eventuali modifiche rispetto alla segmentazione ottenuta attraverso la procedura gerarchica.

**Tabella 28:** *numero di casi in ogni cluster – anno 2007***Numero di casi in ogni cluster**

Cluster	1	10,000
	2	7,000
	3	9,000
	4	3,000
Validi		29,000
Mancanti		,000

Dall'osservazione della tabella si può dedurre che:

- a) il gruppo 1, precedentemente formato da 7 elementi, comprende ora al suo interno 10 elementi.
- b) il gruppo 2, in precedenza composto da 8 casi, è ora formato da soli 7 stati.
- c) Il gruppo 3 ora è composto da 9 elementi, mentre prima era costituito da 8 casi.
- d) Il gruppo 4 comprende 3 stati, mentre prima al suo interno ne venivano rappresentati 6. Questo cluster, per dimensioni rispetto agli altri 3, sembrerebbe quasi essere un gruppo residuale.

Bisogna tuttavia precisare che è molto probabile che i casi che compongono ogni gruppo siano significativamente differenti rispetto alla segmentazione precedentemente ottenuta.

Dopo aver identificato il numero di paesi presente in ogni gruppo, si procede con l'analisi della composizione degli stessi attraverso lo studio della tabella "cluster di appartenenza". Questa tabella, oltre a mostrare a quale gruppo è stato assegnato ciascun paese, nella colonna "distanza" riporta la distanza di ogni membro dal centro del proprio cluster, il che permette di identificare gli elementi più vicini ai centroidi e di conseguenza più rappresentativi del gruppo.

**Tabella 29:** *cluster di appartenenza – anno 2007*

**Cluster di appartenenza**

Numero di caso	Nome	Cluster	Distanza
1	Australia	1	7,586
2	Austria	2	1,788
3	Belgio	3	5,347
4	Canada	1	2,811
5	Repubblica ceca	4	7,726
6	Danimarca	1	8,182
7	Finlandia	2	5,107
8	Francia	3	5,400
9	Germania	4	5,768
10	Grecia	3	5,803
11	Ungheria	3	5,612
12	Islanda	1	5,252
13	Irlanda	2	2,575
14	Italia	3	3,244
15	Giappone	2	6,002
16	Corea	1	8,526
17	Lussemburgo	2	3,666
18	Messico	1	6,157
19	Paesi Bassi	3	7,005
20	Nuova Zelanda	1	3,009
21	Norvegia	1	2,224
22	Polonia	3	7,498
23	Portogallo	3	2,365
24	Slovacchia	4	11,703
25	Spagna	2	3,527
26	Svezia	1	4,802
27	Svizzera	3	5,911
28	Regno unito	2	3,941
29	Stati uniti	1	3,275

I dati riportati nella tabella forniscono le seguenti informazioni:

- a) il gruppo 1 è formato da Australia, Canada, Danimarca, Islanda, Corea, Messico, Nuova Zelanda, Norvegia, Svezia e Stati Uniti. Il paese più rappresentativo di questo gruppo è la Norvegia, con una distanza dal centro pari a 2,224. Lo stato più lontano e quindi meno rappresentativo, con una distanza di 8,526, è la Corea.

- b) il gruppo 2 è costituito da Austria, Finlandia, Irlanda, Giappone, Lussemburgo, Spagna e Regno Unito. Il paese più vicino al centro del cluster che meglio rappresenta il gruppo è, con una distanza di 1,788, l'Austria, mentre quello più lontano dal centro è il Giappone distante 6,002.
- c) fanno parte del gruppo 3 Belgio, Francia, Grecia, Ungheria, Italia, Paesi Bassi, Polonia, Portogallo e Svizzera. Il paese più vicino al centro è il Portogallo, che dista 2,365, mentre il più distante sono i Paesi Bassi con un valore pari a 7,005.
- d) il gruppo 4 comprende Repubblica Ceca, Germania e Slovacchia. Tra questi, il più vicino è la Germania (distanza pari a 5,768) mentre il più lontano, con una distanza di 11,703, è la Slovacchia.

#### 4.4.3 Le caratteristiche dei clusters

L'analisi delle caratteristiche dei cluster si focalizzerà principalmente su due aspetti: la vicinanza tra i gruppi creati e le caratteristiche peculiari di ciascuno di essi.

Si analizzerà d'apprima la tabella che riporta le distanze tra i centri dei cluster e, successivamente, quella indicante i centri finali dei cluster.

**Tabella 30:** *distanze tra i centri dei cluster finali – anno 2007*

**Distanze tra i centri dei cluster finali**

Cluster	1	2	3	4
1		18,615	36,733	51,755
2	18,615		18,209	33,168
3	36,733	18,209		15,476
4	51,755	33,168	15,476	

La tabella riportata sopra mostra la distanza euclidea tra i centroidi dei gruppi finali: chiaramente maggiore è tale distanza, maggiore sarà la dissomiglianza tra i 4 gruppi.

I dati suggeriscono che i gruppi sono tutti relativamente vicini, e quindi simili, tra loro. I due cluster più simili sono il numero 3 e 4, con una distanza tra i centri pari a 15,476, al contrario i più distanti, e quindi differenti, sono il 4 e 1, che presentano una distanza pari a 51,755.

Si procede ora allo studio delle caratteristiche specifiche di ciascun gruppo attraverso l'analisi della tabella che riporta i centri dei cluster finali.

**Tabella 31:** *centri dei cluster finali – anno 2007*

**Centri dei cluster finali**

	Cluster			
	1	2	3	4
Areageo	0	1	1	1
Varpil	3,43	4,30	2,95	6,54
Disoc	4,00	5,33	6,88	8,32
dId	8,80	27,33	45,40	60,26
Sussidi	2	2	2	2
Vardom	3,71	3,77	3,01	4,97

Dall'osservazione della tabella si può dedurre che:

- a) il gruppo 1 comprende paesi che presentano un tasso di crescita del PIL nel 2007 discreto, un tasso di disoccupazione molto basso di cui una percentuale minima è prodotta da disoccupati di lunga durata. La percentuale di spesa pubblica in sussidi rispetto al PIL di questi paesi è bassa e il tasso di crescita della domanda interna è ridotto.
- b) il gruppo 2 è formato da stati con un tasso di crescita del PIL importante, a cui si accosta però un tasso di disoccupazione più consistente rispetto al gruppo 1 condizionato, in buona parte, da disoccupati di lunga durata. In questi paesi la spesa pubblica in sussidi è comunque bassa. Il tasso di crescita della domanda interna si rivela ridotto, anche se superiore a quello dei paesi del gruppo 1.
- c) nel gruppo 3 rientrano nazioni con un basso tasso di crescita del PIL e un elevato tasso di disoccupazione fortemente condizionato da casi di disoccupazione di lunga durata. In questi paesi la percentuale di spesa pubblica in sussidi è ridotta e il tasso di crescita della domanda interna di beni e servizi è il più basso tra tutti i paesi analizzati.
- d) Gli elementi inclusi nel gruppo 4 sono caratterizzati da un forte tasso di crescita del PIL, accompagnato però da un elevato tasso di disoccupazione, influenzato in grossa parte da fenomeni di disoccupazione di lunga durata, e da una spesa pubblica in sussidi non molto rilevante. In questi paesi, contrariamente a quanto si rileva per gli altri gruppi, il tasso di crescita della domanda interna di beni e servizi è piuttosto elevato.

#### 4.4.4 Le variabili che hanno determinato i clusters

A completamento del procedimento di analisi, si propone la tabella ANOVA, che fornisce informazioni riguardanti la diversa influenza che hanno avuto le variabili nell'individuazione dei clusters. Le variabili che differenziano meglio i gruppi sono quelle che presentano un livello di significatività più basso ovvero un valore della F più elevato.

È opportuno ricordare che la procedura ANOVA di SPSS richiede i gruppi bilanciati e in questo caso non lo sono, quindi i risultati ottenuti avranno un'interpretazione puramente descrittiva.

**Tabella 32:** *anova – anno 2007*

**ANOVA**

	Cluster		Errore		F	Sig.
	Media dei quadrati	df	Media dei quadrati	df		
Areageo	,684	3	,130	25	5,253	,006
Varpil	10,680	3	3,386	25	3,154	,042
Disoc	21,016	3	3,268	25	6,432	,002
dld	3141,910	3	24,224	25	129,703	,000
Sussidi	1,015	3	1,153	25	,880	,465
Vardom	3,003	3	5,475	25	,549	,654

I test F devono essere utilizzati solo per motivi descrittivi poiché i cluster sono stati scelti per ottimizzare le differenze tra i casi in diversi cluster. I livelli di significatività osservati non sono perciò corretti e non possono quindi essere interpretati come test dell'ipotesi che le medie dei cluster siano uguali.

Analizzando la tabella sopra esposta, si può dedurre che **DLD**, **DISOC** e **AREAGEO** siano le variabili significativamente associate ai clusters individuati, seguite da **VARPIL**, mentre **SUSSIDI** e **VARDOM** risultano invece le meno influenti nella divisione in gruppi così ottenuta.

## 5. Conclusioni

L'analisi evidenzia che la segmentazione ottimale dei 29 paesi, di cui sono state analizzate grandezze economiche inerenti la produzione, la domanda interna di beni e servizi e la disoccupazione, è quella in 4 cluster, sia facendo riferimento all'anno 2001 che all'anno 2007.

Tuttavia, si possono constatare notevoli differenze nella composizione dei 4 gruppi, nonostante la proporzione tra gli elementi che compongono gli stessi risulti pressoché invariata passando dal 2001 al 2007.

Nello specifico:

- 1) il gruppo 1 presenta sostanzialmente le stesse caratteristiche nel 2001 e nel 2007, ad eccezione del tasso di crescita del PIL, che risulta essere piuttosto basso nella prima annata oggetto di analisi, mentre cresce discretamente nel 2007.

Nei paesi appartenenti a questo gruppo il mercato del lavoro appare abbastanza efficiente, come si può dedurre dai bassi livelli di disoccupazione e dalla minima percentuale di disoccupati di lunga durata. Tutti questi elementi fanno propendere verso la concezione di un mercato del lavoro in cui il ricircolo sia abbastanza elevato e la maggior parte degli inoccupati sia rappresentata da soggetti in transizione, temporaneamente senza impiego alla ricerca di un nuovo lavoro.

A questo gruppo appartengono 7 elementi nel 2001 e 10 nel 2007, rendendolo il gruppo di dimensioni maggiori nel secondo anno analizzato. Fanno parte di questo gruppo in entrambi gli anni Canada, Islanda, Corea, Messico, Norvegia, Stati Uniti quindi, per questi paesi, si può parlare di un sostanziale miglioramento della situazione economica, dovuto a un maggiore tasso di crescita del PIL a parità di un basso tasso di disoccupazione, su cui non influisce lo scarso tasso di crescita della domanda interna di beni e servizi, il che fa propendere verso l'idea che questi paesi traggano i maggiori benefici dalle esportazioni.

Nel 2007 vengono aggregati a questo cluster anche Australia, Danimarca, Nuova Zelanda e Svezia, tutti stati precedentemente accorpati al gruppo 2. Per questi paesi si può quindi evidenziare un miglioramento consistente nel tasso di crescita del PIL, accompagnato da una riduzione dell'incidenza della disoccupazione di lungo periodo sul totale degli inoccupati. Di conseguenza, in questi paesi, risulta conseguente la riduzione della percentuale di spesa pubblica dedicata ai sussidi per la disoccupazione. Tutti questi elementi fanno propendere per un miglioramento rilevante nella condizione economica di questi paesi.

- 2) Il gruppo 2 si caratterizza per alcuni elementi differenti negli anni 2001 e 2007. In particolare, si assiste a una crescita considerevole del tasso di crescita del PIL, accompagnato però da un aumento del tasso di disoccupazione e a una, seppur minima, riduzione della percentuale di spesa pubblica adibita a sussidi.

Il gruppo 2 vede ridotto leggermente il numero di paesi al suo interno, si passa infatti da 9 paesi nel 2001 a 7 nazioni nel 2007. I paesi che fanno parte di questo gruppo sia nel 2001 che nel 2007 sono Austria, Finlandia, Giappone, Lussemburgo, Regno Unito. A seguito delle considerazioni sopra elencate, si può parlare di una crescita economica per questi paesi, denotata principalmente da un maggior tasso di crescita del PIL, nonostante questo sia accompagnato da un aumento del tasso di disoccupazione, che queste nazioni sembrano cercare di fronteggiare riservando, rispetto agli altri paesi analizzati, una quota più importante di spesa pubblica ai sussidi.

L'Irlanda, che nel 2001 era stata inserita nel gruppo 4, fa parte nel 2007 di questo gruppo, quindi, in conseguenza delle caratteristiche precedentemente menzionate, si può osservare che la principale differenza sembra consistere in un aumento del tasso di crescita del PIL e in una riduzione, seppur con incidenza relativa, della percentuale di inoccupati di lungo periodo sul totale dei disoccupati.

La Spagna, inserita nel 2001 nel gruppo 3 e nel 2007 nel gruppo 2, presenta una riduzione sostanziale del tasso di disoccupazione, che viene influenzato in modo meno rilevante dagli inoccupati di lunga durata, il che lascia presumere un miglioramento consistente del mercato del lavoro in questo stato.

- 3) Nel gruppo 3 non si verificano modifiche per quanto riguarda il numero degli elementi nel passaggio da un'analisi all'altra, difatti in entrambi gli anni presi in considerazione appartengono a questo gruppo 9 stati, quindi circa il 30% del totale dei paesi analizzati.

Vengono inseriti in questo cluster sia nel 2001 che nel 2007 Belgio, Grecia, Ungheria, Italia, Polonia. Tutti questi paesi denotano evidenti problemi legati al mercato del lavoro, infatti hanno un tasso di disoccupazione elevato, condizionato in maniera rilevante da casi di disoccupazione di lunga durata. Nonostante ciò, la percentuale di spesa pubblica destinata a sussidi non è per nulla rilevante. Tutti questi elementi fanno comprendere come un crollo verticale del tasso di crescita del PIL non sia altro che una inevitabile conseguenza, passando dal 2001 al 2007, di un mercato del lavoro molto precario con conseguente impossibilità di stimolare la domanda interna e conseguentemente la produzione e la crescita.

Nel 2007 vengono inseriti in questo gruppo anche Francia, Portogallo e Svizzera, tutti inseriti nel gruppo 4 nell'analisi effettuata sui dati del 2001. Per questi paesi si può parlare di un vero e proprio crollo nell'economia, infatti non solo si assiste a una diminuzione consistente del tasso di crescita del PIL, ma anche a un aumento del tasso di disoccupazione, su cui incide maggiormente la percentuale di inoccupati di lungo periodo, e a una diminuzione drastica del tasso di crescita della domanda interna di beni e servizi.

- 4) Il gruppo 4, in entrambi gli anni analizzati, sembra essere un gruppo residuale, infatti vengono inseriti in questo gruppo solamente 4 paesi nel 2001 e 3 nel 2007. Tuttavia, i paesi che compongono questo cluster nel 2007 sono completamente differenti rispetto a quelli presenti nello stesso nel 2001 e, più precisamente, fanno parte di questo gruppo Repubblica Ceca, Germania e Slovacchia, tutte nazioni precedentemente inserite nel cluster 3. Appare quindi evidente la necessità di confrontare le caratteristiche di questo gruppo con quelle del cluster 3 derivante dall'analisi effettuata sui dati del 2001.

Come si può notare, tutti questi paesi presentano sia nel 2001 che nel 2007 elevati tassi di crescita del PIL, accompagnati però da un forte tasso di disoccupazione, in cui la parte di disoccupati "cronici" assume un'importanza rilevante. Ciò che differisce nel confronto tra i due anni è la variazione del tasso di crescita della domanda interna di beni e servizi, piuttosto ridotta nel 2001 e, al contrario, alquanto consistente nel 2007.

Questi paesi, in definitiva, sembrano aver trovato la soluzione per sostenere la propria economia nell'aumento della richiesta interna di beni e servizi. Infatti, attraverso una maggiore domanda interna, si stimola la produzione e la conseguente crescita dell'economia.

In conclusione, evidenziando nuovamente che la divisione ottimale dei paesi appartenenti all'OECD, sia per quanto riguarda il 2001 che il 2007, sia quella in 4 cluster, si può affermare che i 4 diversi gruppi di paesi abbiano subito evoluzioni differenti nell'economia che li porteranno, negli anni successivi a quelli oggetto di analisi, ad affrontare in maniera differente la crisi economica mondiale che si è sviluppata negli ultimi anni.

In particolare, i paesi appartenenti al gruppo 1 appaiono essere quelli con un'economia più solida e un mercato del lavoro maggiormente efficiente, tutte caratteristiche che permetteranno a queste nazioni di affrontare meglio la crisi, subendone gli effetti in misura minore e riuscendo a limitare, per quanto possibile, i danni da essa derivanti.

Anche i paesi inseriti nel gruppo 2 sembrano, negli anni antecedenti la crisi, avere un'economia relativamente forte, che permetterà loro di fronteggiare in maniera più adeguata lo shock dell'economia negli anni a venire, facendo leva anche sulle politiche di welfare a tutela della popolazione disoccupata che sembrano aver già messo in atto con costanza negli anni oggetto di analisi.

Al contrario, i paesi che compongono il gruppo 3 presentano numerosi problemi inerenti sia al mondo del lavoro che alla crescita dell'economia. Queste nazioni presentano delle economie in crisi, in cui i tassi di crescita sono molto ridotti e dove il mercato del lavoro è fortemente a rischio. Saranno questi i paesi che avranno le maggiori difficoltà in seguito agli avvenimenti economici che caratterizzeranno gli anni a venire e subiranno in maniera più consistente gli effetti negativi della crisi economica mondiale che si svilupperà.

## 6. Indice delle tabelle e dei grafici

1.	<b>Tabella 1:</b>	dataset analizzato – anno 2001	15
2.	<b>Tabella 2:</b>	correlazioni di Pearson – anno 2001	16
3.	<b>Tabella 3:</b>	statistiche descrittive – anno 2001	17
4.	<b>Tabella 4:</b>	programma di agglomerazione – anno 2001	18
5.	<b>Grafico 1:</b>	icicle diagram – anno 2001	20
6.	<b>Grafico 2:</b>	dendrogramma – anno 2001	21
7.	<b>Grafico 3:</b>	scree-plot – distanza di fusione – anno 2001	22
8.	<b>Tabella 5:</b>	distanza di fusione – anno 2001	23
9.	<b>Tabella 6:</b>	numerosità dei cluster – anno 2001	24
10.	<b>Grafico 4:</b>	grafico a torta – numerosità dei cluster – anno 2001	24
11.	<b>Tabella 7:</b>	cluster di appartenenza – anno 2001	25
12.	<b>Tabella 8:</b>	caratteristiche dei gruppi – anno 2001	26
13.	<b>Tabella 9:</b>	cronologia delle iterazioni – anno 2001	28
14.	<b>Tabella 10:</b>	centri dei cluster iniziali – anno 2001	28
15.	<b>Tabella 11:</b>	entri dei cluster finali – anno 2001	29
16.	<b>Tabella 12:</b>	numero di casi in ogni cluster – anno 2001	29
17.	<b>Tabella 13:</b>	cluster di appartenenza – anno 2001	30
18.	<b>Tabella 14:</b>	distanze tra i centri dei cluster finali – anno 2001	31
19.	<b>Tabella 15:</b>	centri dei cluster finali – anno 2001	32
20.	<b>Tabella 16:</b>	anova - anno 2001	33

21.	<b>Tabella 17:</b> dataset analizzato – anno 2007	34
22.	<b>Tabella 18:</b> correlazioni di Pearson – anno 2007	35
23.	<b>Tabella 19:</b> statistiche descrittive – anno 2007	36
24.	<b>Tabella 20:</b> programma di agglomerazione – anno 2007	37
25.	<b>Grafico 5:</b> icicle diagram – anno 2007	39
26.	<b>Grafico 6:</b> dendrogramma – anno 2007	40
27.	<b>Grafico 7:</b> scree-plot – distanza di fusione – anno 2007	41
28.	<b>Tabella 21:</b> distanza di fusione – anno 2007	42
29.	<b>Tabella 22:</b> numerosità dei cluster – anno 2007	43
30.	<b>Grafico 8:</b> grafico a torta – numerosità dei cluster – anno 2007	43
31.	<b>Tabella 23:</b> cluster di appartenenza – anno 2007	44
32.	<b>Tabella 24:</b> caratteristiche dei gruppi – anno 2007	45
33.	<b>Tabella 25:</b> cronologia delle iterazioni – anno 2007	47
34.	<b>Tabella 26:</b> centri dei cluster iniziali – anno 2007	47
35.	<b>Tabella 27:</b> centri dei cluster finali – anno 2007	48
36.	<b>Tabella 28:</b> numero di casi in ogni cluster – anno 2007	48
37.	<b>Tabella 29:</b> cluster di appartenenza – anno 2007	49
38.	<b>Tabella 30:</b> distanze tra i centri dei cluster finali – anno 2007	50
39.	<b>Tabella 31:</b> centri dei cluster finali – anno 2007	50
40.	<b>Tabella 32:</b> anova – anno 2007	52

## 6. Riferimenti bibliografici

- Borcio R. - *L'analisi dei gruppi* - FrancoAngeli editore
- Barbaranelli C. - *Analisi dei dati con SPSS II, le analisi multivariate* - LED edizioni
- Gates R. / McDaniel C. - *Marketing research with SPSS 8<sup>th</sup> edition* - Wiley
- Rencher A.C. - *Methods of multivariate analysis 2<sup>nd</sup> edition* – Wiley
- Everitt B.S. / Landau S. / Leese M. / Stahl D. - *Cluster analysis 5<sup>th</sup> edition* – Wiley
- Romesburg H.C. - *Cluster analysis for researchers* - Lulu press
- Mooi E. / Sarstedt M. - *A concise guide to market research, the process, data and methods using IBM SPSS statistics* – Springer
  
- Organization for Economic Cooperation and Development - [www.oecd.org](http://www.oecd.org)