*(Qualificazione del Territorio mediante un'applicazione dell'Analisi Multivariata)*

## 1. - PREMESSA

Quando si deve analizzare un fenomeno complesso è necessario disporre di un numero elevato di dati descrittivi del fenomeno. Tali dati esprimono "misure" parziali e non bilanciate del fenomeno, poiché ne descrivono i diversi elementi caratteristici secondo una scala di pesi non univoca e, soprattutto, non nota. Con il semplice assemblaggio dei dati non è possibile pervenire (se il fenomeno è complesso) ad una descrizione sintetica e quanto più possibile obiettiva: occorre, a tale fine, ricorrere a specifiche tecniche di analisi statistica.
L'analisi dei fattori (o multivariata) è la tecnica statistica più idonea, poiché essa si propone di esprimere un insieme di variabili attraverso un numero ridotto di fattori, allo scopo di realizzare un'economia ed una maggiore obiettività nella descrizione dei fenomeni cui le variabili si riferiscono.
Obiettivo di questo Rapporto è fornire un contributo di analisi per la qualificazione del territorio provinciale mediante un'applicazione dell'analisi multivariata, volta a dare una misura relativa (rispetto alla media provinciale) dello sviluppo socio-economico di ciascun comune.

Le statistiche ufficiali non offrono alcuna *misura diretta* dello sviluppo, mentre ne forniscono parecchie *indirette*:
- misure del *tenore di vita* (consumi elettrici, auto immatricolate, gettito ICI, ecc.);
- misure dello *sviluppo industriale* (addetti ai vari settori industriali, incidenza del settore manifatturiero sugli altri settori, ecc.), *agricolo* (superficie agricola utilizzata, tipo di seminativi, percentuale a boschi, ecc.) e dei *servizi* (addetti ad alberghi e ristoranti, numero di esercizi commerciali, ecc.);
- misure della *componente demografica* (indici di natalità, indice di vecchiaia, numero residenti, ecc.);
- misure della *condizione sanitaria* (numero di posti letto, cause di morte per gruppi di cause, numero di medici pro capite, ecc.);
- misure del *livello di istruzione* (numero di aule, numero di alunni elementari, medie, superiori, ecc.).

Tali variabili spesso misurano lo stesso fenomeno da diversi angoli visuali; in altri termini, esse non sono *indipendenti* le une dalle altre.

Né è possibile capire, con le tecniche convenzionali (aggregazione delle informazioni per categoria), quanto l'insieme dei dati sia bilanciato, o, più correttamente quanto esso esalti la visibilità di alcuni fenomeni a discapito di altri: il procedimento tradizionalmente seguito - quello cioè di considerare tutti gli aspetti possibili di un fenomeno aumentando il numero delle variabili - rischia, per tale motivo, di ridurre, anziché di aumentare, la globale efficacia delle informazioni.

L'analisi multivariata consente invece di:
- *costruire un numero limitato di indici, definiti fattori, assolutamente adimensionali, statisticamente indipendenti* tra loro, descrittivi ciascuno di *un'informazione completamente diversa* (al contrario delle variabili di partenza, che sono in buona parte "sovrapposte", ossia legate le une alle altre);
- ottenere, mediante un *numero minimo di fattori* (decisamente inferiore al numero di partenza), la sintesi, con una perdita accettabile, di tutte le informazioni contenute nelle variabili di partenza.

Il Rapporto è strutturato in due capitoli:
- nel primo è qualificata la metodologia adottata;
- nel secondo si esaminano i risultati forniti dall'applicazione sviluppata.

E' corredato di un'Appendice Statistica, in cui sono riportati i dati utilizzati per l'analisi.

# 2. - LA METODOLOGIA DI ANALISI

La metodologia adottata si struttura attraverso i seguenti passi:
1. *individuazione e reperimento dei dati elementari disponibili* su base comunale;
2. *selezione delle variabili esplicative*; tale passo è necessario poiché il numero di variabili acquisite è risultato maggiore del numero di comuni: il numero delle equazioni (una per comune) deve, infatti, essere necessariamente maggiore del numero di variabili;
3. *applicazione della tecnica statistica multivariata (in particolare dell'analisi in componenti principali) alle variabili esplicative* di cui al punto 2;
4. *interpretazione dei fattori*, mediante l'analisi delle correlazioni esistenti tra le variabili esplicative ed i fattori;
5. *analisi di raggruppamento*, ossia individuazione di gruppi "omogenei" di comuni, definiti sulla base dei suddetti fattori.

## 2.1. - I DATI ELEMENTARI

Sono stati acquisiti i dati comunali relativi a 251 variabili.
La fonte principale è costituita dal Censimento Intermedio delle Attività Produttive[1]; le altre fonti utilizzate sono le seguenti: ISTAT (Censimento generale della popolazione, Censimento Generale delle Attività Produttive, statistiche diverse), Prefettura di Imperia, Camera di Commercio, Industria Artigianato e Agricoltura di Imperia, Azienda di Promozione Turistica di Imperia, ENEL, ACI, CAIRE, Regione Liguria.

Nell'Appendice "Riepiloghi Statistici" sono riportate le schede comunali riportanti tali dati[2].
I dati non sono pienamente comparabili per disomogeneità inerenti le coordinate temporali, i metodi di rilevazione e la natura delle fonti.
Il problema dello sfasamento temporale si pone con particolare evidenza nel raffronto dei dati del 13° Censimento generale della popolazione e delle attività produttive (1991) e del Censimento intermedio del 1996. La soluzione adottata è quella - di fatto obbligata - di utilizzare i dati 1991 per quei settori (agricoltura e pubblica amministrazione) esclusi dal Censimento Intermedio del 1996.
Considerata la prevalenza numerica dei dati ISTAT 1996, in tutti i casi in cui ciò è stato possibile, si sono utilizzati i dati al 1996.

## 2.2. - LA SCELTA DELLE VARIABILI ESPLICATIVE

La base dati disponibile determina una *matrice elementare, o matrice dei dati di partenza*, costituita da 67 righe (comuni della provincia di Imperia) e da 251 colonne (variabili sopra definite).
L'applicabilità della tecnica statistica multivariata è subordinata al vincolo che la matrice abbia un numero di colonne non superiore al numero di righe: essendo 67 i comuni, le variabili esplicative non possono, pertanto, essere più di 66. Occorre pertanto preliminarmente selezionare non più di 66 variabili significative.
A tale fine la matrice elementare è stata sottoposta ad un'analisi esplorativa basata su statistiche descrittive di tipo univariato. E' stata costruita una matrice di correlazione lineare (avente lo stesso numero di righe e di colonne della matrice elementare), la quale condensa tutte le informazioni relative alle mutue relazioni a due a due tra le variabili, espresse mediante i *coefficienti di correlazione lineare semplice*[3].
Escludendo tutti i dati elementari che presentavano un coefficiente di correlazione lineare superiore ad una soglia stabilita[4], si sono individuate 58 variabili esplicative[5].Si precisa che, per una più diretta lettura dei dati, tutti sono stati normalizzati, rapportandoli alla popolazione residente al 31.12.1996 nel comune di riferimento[6].

---

[1] Il Censimento interedio ISTAT 1996 ha avuto per oggetto tutte le imprese attive al 31/12/1996 ed iscritte al Registro delle imprese delle Camere di Commercio, gli artigiani, i lavoratori autonomi (liberi professionisti iscritti agli albi professionali e lavoratori in proprio). Le imprese osservate sono state quelle operanti nel settore industriale e nei servizi**,** in organizzazioni profit, comprese le cooperative ed i consorzi. Sono state escluse dalla rilevazione le imprese operanti nel settore dell'agricoltura, della sanità, dell'istruzione e le organizzazioni no-profit.

[2] I dati tabellati sono quelli più recenti; nelle elaborazioni, come meglio nel seguito esposto, si è invece preferito utilizzare, quando possibile, i dati al 1996.

[3] *Il coefficiente di correlazione rij è* un numero compreso, in valore assoluto, tra 0 e 1; assume valore 0 quando la correlazione è nulla (i punti nel piano cartesiano sono disposti a nuvola); assume valore 1 nel caso di massima correlazione tra le variabili (i punti nel piano cartesiano si dispongono secondo una retta); assume valori intermedi negli altri casi. Esso assume, in segno, valori *positivi* se la relazione tra le due variabili è *diretta* (all'aumentare di una grandezza, tende ad aumentare anche l'altra); assume invece valori *negativi* quando, all'aumentare di una grandezza, l'altra tende a *diminuire*. Il quadrato del coefficiente di correlazione $R^2_{ij}$ è detto *Indice di determinazione* e rappresenta la frazione di varianza totale della variabile *i* spiegata dalla variabile *j* e viceversa.

[4] Si propongono alcuni grafici esemplificativi di diversi livelli di correlazione tra le variabili considerate.

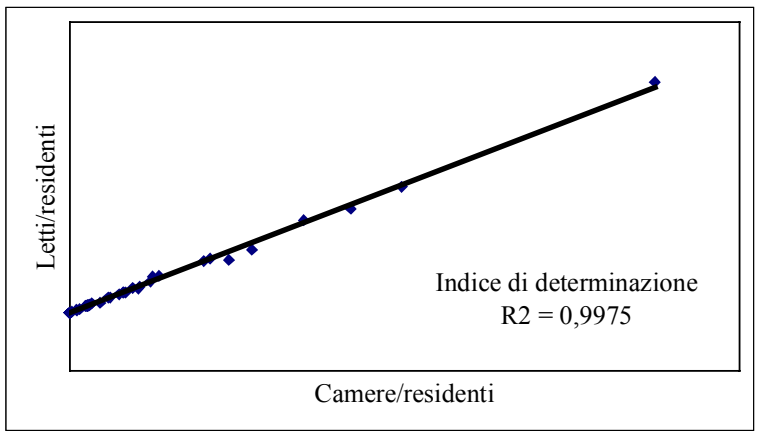
## 2.3. - APPLICAZIONE DELL'ANALISI IN COMPONENTI PRINCIPALI

La tecnica statistica multivariata utilizzata in questa sede, detta *analisi in componenti principali*, partendo dalle 58 variabili esplicative individua nuovi *fattori* in numero pari al numero delle variabili originarie, attraverso i quali è possibile spiegare tutta la varianza di queste ultime[7].
Si nota che alcuni dei fattori così individuati contribuiscono alla spiegazione di una quota minima di varianza: è pertanto possibile escluderli, così selezionando un numero limitato di componenti che spieghino la maggiore parte della varianza. Operando in questi termini si può descrivere il fenomeno in modo efficace con un numero di variabili inferiore a quello di partenza.

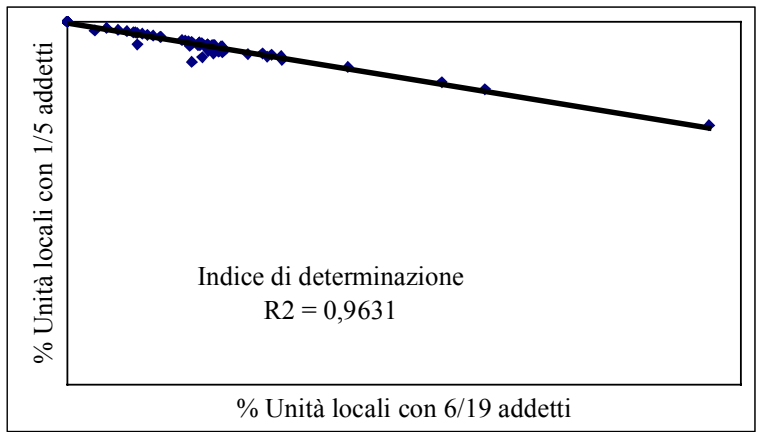
Per meglio esplicitare il fondamentale concetto espresso si produce un breve esempio.
Si vogliano suddividere in gruppi omogenei i comuni imperiesi sulla base del fenomeno "Struttura agricola". Le variabili esplicative scelte siano le seguenti 11, per ognuna delle quali sono disponibili i dati, a livello comunale, relativi a:

| | | | |
|---|---|---|---|
| *- densitàdemografica* | *-addetti agricoltura* | *-colture ortofloricole in pien'aria e vivai* | *-nucleo abitato a quota massima* |
| *-colture agrarie* | *-vigneti* | *-oliveti* | *-quota media.* |

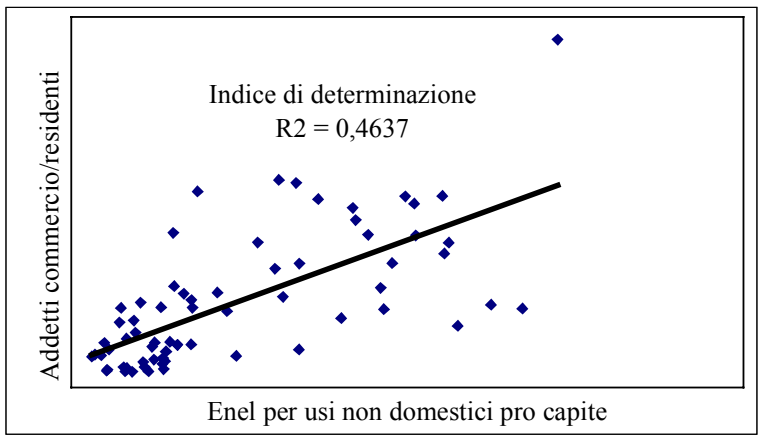
**Esempio di correlazione lineare diretta**



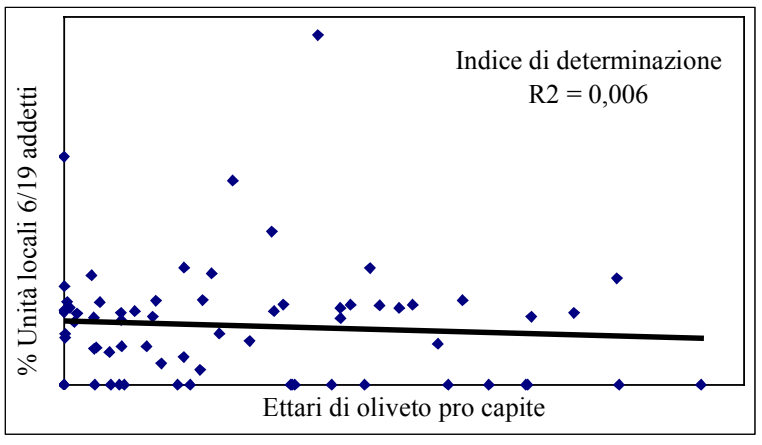

**Esempio di correlazione lineare indiretta**




**Esempio di correlazione lineare media**




**Esempio di variabili pressoché indipendenti**




[5] Si rileva la frequente presenza, soprattutto per i comuni non rivieraschi, di valori pari a zero o prossimi allo zero di molte variabili. Ciò determina una maggiore alea del procedimento di selezione delle variabili
significative, in quanto voci nulle o assolutamente insignificanti per taluni comuni, costituiscono invece punti di forza per altri.
[6] Ad eccezione dei dati mutuati dal 13° Censimento della popolazione e delle attività produttive, che sono stati rapportati ai residenti al 1991.
[7] L'idea centrale dell'analisi consiste nell'ipotizzare delle relazioni tra le *variabili* descrittive del fenomeno in esame $X_1, X_2, .., X_n$ (o, meglio, le loro trasformate standardizzate $Z_1, Z_2, .., Z_n$) ed i *fattori* $F_1, F_2, .., F_n$ considerati anch'essi come variabili in forma standardizzata ed ortogonali (cioè statisticamente indipendenti) tra loro. In questo modo, le diverse misure della variabilità osservabile ($Z_i$) sono messe in un sistema relativo a quelle variabili ($F_j$) che costituiscono le grandezze "vere" che governano il fenomeno. La relazione tra le variabili Zj ed i fattori Fj è espressa, nella forma più generale, dal sistema di relazioni:

$$Z_1 = a_{11}F_1 + a_{12}F_2 + ..... + a_{1m}F_m$$
$$.....$$
$$Z_n = a_{n1}F_1 + a_{n2}F_2 + ..... + a_{nm}F_m$$

in cui i coefficienti $a_{ij}$ dei fattori $F_j$ rappresentano i *coefficienti di correlazione tra il fattore F e la variabile Z*; questi coefficienti sono quindi in grado di fornire informazioni sulla frazione di varianza complessiva
(detta comunalità ed indicata con $h^2_j$) della variabile $Z_i$ *spiegata* dai fattori comuni $F_1, F_2,..., F_m$. In altri termini:

$$h_j^2 = \sum_{j=1}^{m} a_{ij}^2 \qquad \text{per j = 1, 2, ....., m.}$$

Genericamente, si distinguono diversi tipi di fattori:

- ***fattori comuni***, che partecipano alla spiegazione di una o più variabili, comprendenti:
  - *fattore generale* presente in tutte le variabili;
  - *fattori di gruppo* presenti in più di una variabile ma non in tutte;
- ***fattori unici o specifici***, che appaiono esclusivamente nella descrizione di una particolare variabile.

Proprietà fondamentale dell'analisi dei fattori è che qualsiasi coefficiente di relazione può essere ottenuto per mezzo della seguente espressione:

$$r_{ij} = a_{j1} \times a_{k1} + ..... + a_{jm} \times a_{km} \qquad (j, k = 1, 2, ....., m)$$

Questo aspetto del problema fattoriale consente di vedere fino a che punto un insieme di fattori *spieghi* le correlazioni esistenti tra le variabili date.

*-lavoratori agricoli* *-serre* *-nucleo abitatoa quota minima*

L'analisi in componenti principali, partendo dalle 11 variabili esplicative individua 11 nuovi fattori, adimensionali, attraverso i quali è possibile spiegare tutta la varianza delle variabili originarie.

*Tabella 1 Varianza spiegata dai valori estratti dall'analisi, in valore assoluto (autovalore) e percentuale.*

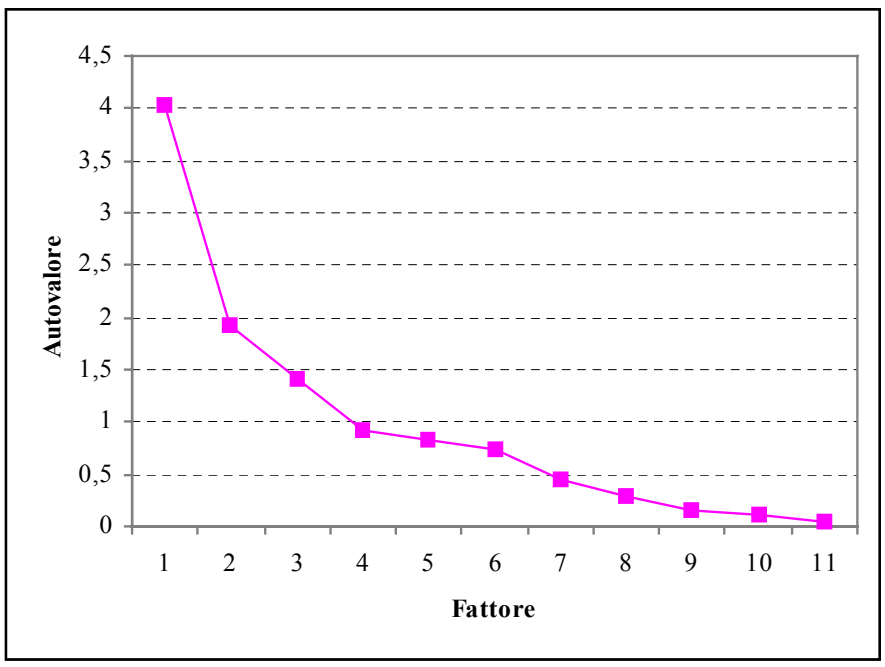
| Fattore | Autovalore | % di varianza | % varianza cumulata |
|---|---|---|---|
| **1** | 4,03153 | 36,650 | 36,650 |
| **2** | 1,93037 | 17,549 | 54,199 |
| **3** | 1,4169 | 12,881 | 67,080 |
| **4** | 0,912507 | 8,296 | 75,376 |
| **5** | 0,819229 | 7,448 | 82,824 |
| **6** | 0,732951 | 6,663 | 89,487 |
| **7** | 0,448066 | 4,073 | 93,560 |
| **8** | 0,285113 | 3,501 | 97,061 |
| **9** | 0,157206 | 1,429 | 98,490 |
| **10** | 0,111013 | 1,009 | 99,499 |
| **11** | 0,0551102 | 0,501 | 100,000 |

Si nota (riferimento tabella 1) che il primo fattore spiega, da solo, oltre 1/3 della varianza complessiva; il secondo fattore, unitamente al primo, giunge a spiegare oltre la metà della varianza spiegata dalle 11 variabili di partenza, e così via, con contributi decrescenti aumentando il numero d'ordine dei fattori.

Per procedere all'individuazione del numero di fattori che descrivano il fenomeno, ossia la struttura agricola, in modo efficace permettendo al contempo una sintesi, possono utilizzarsi diversi criteri.

1. Uno dei metodi più noti è quello di selezionare unicamente i fattori con autovalore maggiore di 1, ovvero quei fattori che spiegano una quota della varianza superiore a quella spiegata da una singola variabile. Nell'esempio in oggetto, dovrebbero essere estratti i primi 3 fattori.
2. Altro criterio di prevalente applicazione è lo *scree test*. In esso, vengono rappresentati graficamente (vedi grafico 1), in *ascissa,* i fattori estratti in ordine di estrazione e, in *ordinata*, gli autovalori. Si nota che, partendo dal primo fattore, la pendenza si riduce progressivamente, fino a che la curva diventa quasi orizzontale. Il punto in cui la curva inizia ad appiattirsi è quello che indica il massimo numero di fattori da estrarre. Nel nostro esempio, il numero massimo risulterebbe pari a 4.

*Grafico 1 Scree Plot. Varianza totale associata a ciascun fattore.*




Si nota che i fattori estratti, in numero variabile a seconda del metodo scelto e degli obiettivi dello studio, sono fattori *adimensionali*, ossia assolutamente *non* espressi in unità di misura, nonché assolutamente indipendenti gli uni dagli altri.
Essi sono "prodotti" di un metodo statistico e sta esclusivamente alla sensibilità ed all'esperienza del ricercatore attribuire loro significati economici e definizioni sintetiche che li descrivano in termini accettabilmente congruenti.

## 2.4. - L'INTERPRETAZIONE DEI FATTORI ESTRATTI

Una volta deciso il numero di fattori da estrarsi, essi vengono interpretati grazie alla matrice dei *factor loadings* che risulta così strutturata:
- in ogni riga della matrice è riportata la correlazione tra la variabile esplicativa ed ognuno dei fattori, ovvero il peso assunto dalla variabile all'interno del fattore;
- in ogni colonna si legge invece la correlazione di un fattore con ciascuna variabile esplicativa.

Con riferimento all'esempio sviluppato, possiamo costruire la matrice dei *factor loading*, (tabella 2) limitandoci ai primi 3 fattori estratti.
Si rileva che mentre la somma per colonne (dei quadrati dei valori) è pari a 1, la somma per riga (dei quadrati dei valori) sarebbe pari a 1 soltanto se si considerasse la matrice dei *factor loadings* riferita a tutti gli 11 fattori estratti.

*Tabella 2 Matrice dei factor loadings (a sinistra: valori assoluti; a destra: valori %).*

| | **Fattore 1** | **Fattore 2** | **Fattore 3** |
|---|---|---|---|
| *Densità demografica* | 0,20538 | -0,253735 | 0,00556588 |
| *Colture agrarie* | 0,315724 | -0,189299 | -0,321442 |
| *Lavoratori agricoltura* | -0,332996 | -0,3148 | -0,383435 |
| *Addetti agricoltura* | -0,0577294 | -0,234204 | 0,665317 |
| *Vigneti* | 0,42818 | -0,103243 | -0,153034 |
| *Serre* | 0,445213 | -0,0855381 | 0,0971809 |
| *Colture ortofloricole in piena aria e vivai* | -0,156204 | 0,55111 | 0,104114 |
| *Oliveti* | -0,147059 | -0,275718 | 0,416107 |
| *Nucleo abitato a quota minima* | 0,463778 | -0,0437511 | 0,0859216 |
| *Nucleo abitato a quota massima* | -0,3154433 | -0,390676 | -0,244909 |
| *Quota media* | 0,0418616 | 0,439969 | -0,150874 |

| | **Fattore 1** | **Fattore 2** | **Fattore 3** |
|---|---|---|---|
| *Densità demografica* | 4,22% | 6,44% | 0,00% |
| *Colture agrarie* | 9,97% | 3,58% | 10,33% |
| *Lavoratori agricoltura* | 11,09% | 9,91% | 14,70% |
| *Addetti agricoltura* | 0,33% | 5,49% | 44,26% |
| *Vigneti* | 18,33% | 1,07% | 2,34% |
| *Serre* | 19,82% | 0,73% | 0,94% |
| *Colture ortofloricole in piena aria e vivai* | 2,44% | 30,37% | 1,08% |
| *Oliveti* | 2,16% | 7,60% | 17,31% |
| *Nucleo abitato a quota minima* | 21,51% | 0,19% | 0,74% |
| *Nucleo abitato a quota massima* | 9,95% | 15,26% | 6,00% |
| *Quota media* | 0,18% | 19,36% | 2,28% |
| | 100% | 100% | 100% |

Per meglio chiarire il significato dei valori riportati nella matrice, si evidenziano le seguenti notazioni di commento riferite alla matrice *esempio*:
- i valori numerici che appaiono in tabella sono generalmente bassi: ciò vuol dire che all'individuazione di un singolo fattore concorrono, in diverse quote parti, più variabili esplicative; in casi come questi risulta estremamente difficile attribuire una definizione ed un significato al fattore estratto;
- alla definizione del fattore 1 - che spiega oltre 1/3 della varianza complessiva - concorrono principalmente - con segno positivo - le variabili "*vigneti*" e "*serre";*
- il fattore 2 risulta fortemente correlato alla variabile "*colture ortofloricole in piena aria e vivai*";
- il fattore 3 risulta fortemente correlato alle variabili "*addetti agricoltura"* e "*lavoratori agricoltura".*

## 2.5. - ANALISI DI RAGGRUPPAMENTO

L'ultima fase dell'analisi multivariata consente, partendo dalla matrice dei *factor loadings*, di calcolare la posizione che ogni soggetto - nel nostro esempio i comuni - assume rispetto ai fattori estratti.
Tali posizioni sono espresse mediante valori, denominati *factor scores*, che vengono composti in una matrice. Essa presenta, per riga, il peso che ogni fattore assume per il singolo comune.

*Tabella 3- Matrice dei factor scorse*

| | **Fattore 1** | **Fattore 2** | **Fattore 3** |
|---|---|---|---|
| Airole | -0,00140845 | -0,318944 | 0,681031 |
| Apricale | -0,63117 | -0,439781 | -0,511749 |
| Aquila di Arroscia | -1,27851 | 0,239447 | -0,759874 |
| Armo | -2,15601 | -0,189648 | -0,173978 |
| Aurigo | -1,25766 | -1,30468 | 0,646433 |
| Badalucco | -0,430044 | -0,143848 | 0,351099 |
| Baiardo | -3,99317 | 1,44725 | -0,218065 |
| Bordighera | 2,78816 | 2,11006 | 0,669472 |
| Borghetto d'Arroscia | -1,16298 | -0,190969 | 0,201817 |
| Borgomaro | -0,59 | -0,43562 | 0,472262 |
| Camporosso | 1,43641 | 0,132524 | -0,117811 |
| Caravonica | -0,786009 | -2,58751 | 0,654039 |
| Carpasio | -2,78093 | 0,710101 | -0,594359 |
| Castellaro | 0,464647 | 0,388823 | -1,43759 |
| Castel Vittorio | -1,19227 | 0,612004 | -1,34503 |
| Ceriana | -0,582907 | -0,999126 | -0,707033 |
| Cervo | 1,53995 | -0,65207 | 0,567645 |
| Cesio | -1,80192 | -0,993865 | 0,212327 |
| Chiusanico | -0,101519 | -0,958988 | -0,344793 |
| Chiusavecchia | 0,82008 | -1,35425 | -0,107396 |
| …….. | …… | …….. | …….. |

Con riferimento all'esempio sopra esplicitato, nella tabella 3, vengono presentate le prime venti righe, corrispondenti ai primi venti comuni in ordine alfabetico, della *matrice dei factor scores* e le posizioni assunte dai comuni con riferimento ai singoli fattori.
Ultimata l'analisi in componenti principali, si procede, infine, all'analisi di raggruppamento o classificazione automatica.
E' necessario rimarcare in questo contesto che ogni classificazione dipende:

- dalle variabili esplicative iniziali,
- dal metodo di aggregazione.

I metodi di classificazione automatica sono assai differenziati: i più usati sono quelli "ascendenti", quelli "discendenti" ed i metodi di partizionamento. Questi ultimi vengono sovente utilizzati per permettere la classificazione di popolazioni molto numerose (svariate migliaia di unità) e non si prestano quindi allo studio in oggetto.

La metodologia seguita nello studio fa riferimento ad una lettura di tipo "discendente" dei factor scores.

# 3. - L'APPLICAZIONE DELLA TECNICA STATISTICA MULTIVARIATA

Nel capitolo è sviluppata l'applicazione della tecnica statistica multivariata alle 58 variabili esplicative selezionate al § 2.2, passando attraverso le tre fasi descritte nei § 2.3 ÷ 2.5:

1. individuazione dei fattori,
2. interpretazione dei fattori estratti mediante la matrice dei *factor loadings*,
3. analisi di raggruppamento dei comuni imperiesi sulla base dei *factor scores*.

## 3.1. - L'INDIVIDUAZIONE DEI FATTORI DI ANALISI E LA LORO QUALIFICAZIONE

### 3.1.1. - Individuazione dei fattori di analisi

In tabella 4 sono riportati i primi (ossia i principali) 10 fattori estratti (rispetto ai 58 complessivi); di essi viene indicato l'autovalore, la percentuale di varianza spiegata e la varianza cumulata.

*Tabella 4 Varianza spiegata dai fattori estratti dall'analisi fattoriale, in valore assoluto (autovalore) e in percentuale.*

| Fattore | Autovalore | % di varianza | % varianza cumulata |
|---|---|---|---|
| 1 | 6,58982 | 32,949 | 32,949 |
| 2 | 3,62576 | 18,129 | 51,078 |
| 3 | 2,5815 | 12,907 | 63,985 |
| 4 | 1,50302 | 7,515 | 71,500 |
| 5 | 1,25787 | 6,289 | 77,789 |
| 6 | 0,944408 | 4,722 | 82,511 |
| 7 | 0,621633 | 3,108 | 85,619 |
| 8 | 0,586032 | 2,930 | 88,549 |
| 9 | 0,517839 | 2,589 | 91,138 |
| 10 | 0,356015 | 1,780 | 92,918 |
| 11 | ……. | ……. | …… |

I fattori sono numerati in ordine di percentuale di varianza spiegata:

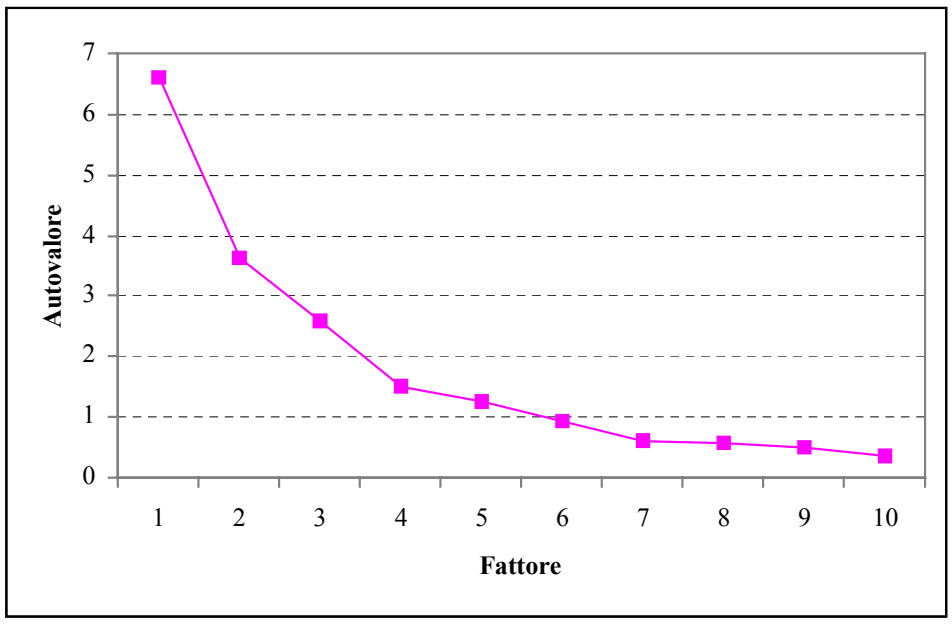
1. il primo fattore spiega circa 1/3 della varianza complessiva delle variabili considerate;
2. il secondo, unitamente al primo, spiega oltre la metà della varianza complessiva;
3. i primi tre complessivamente spiegano i 2/3 circa.

La scelta del numero di fattori da estrarre può essere definita:

- con il metodo dell'autovalore almeno pari all'unità: in questo caso, dovrebbero estrarsi i primi 5 fattori (ossia quelli con autovalore maggiore di 1);
- con il metodo dello *scree plot*: in questo caso dovrebbero essere estratti i primi 4 fattori; la curva, infatti, inizia ad appiattirsi dopo il 4° valore (grafico 2).

Trattasi, in entrambi i casi di un numero troppo alto di fattori rispetto al numero di osservazioni di nostro interesse (i 67 comuni). Si è, pertanto, preferito limitarsi ai primi 3 fattori. Si nota che l'elevato numero di fattori che le metodologie di estrazione indicano come significativi costituisce un primo importante segnale di una reale difficoltà di lettura del territorio.

*Grafico 2 Metodo dello Scree Plot. Varianza totale associata a ciascun fattore.*



### 3.1.2. - Qualificazione del primo fattore

Il primo fattore estratto dall'analisi fattoriale (fattore F1) spiega circa un terzo della varianza complessiva delle variabili originali.
Le variabili di partenza che lo definiscono maggiormente (percentuale di varianza spiegata compresa l'11.51 % ed il 6.04 %), sono indicate nel seguente prospetto[8].

| | ***Fattore 1*** | ***%*** |
|---|---|---|
| *Totale addetti* | 0,33923 | 11,51% |
| *Addetti alberghi e ristoranti* | 0,32064 | 10,28% |
| *Gettito ICI pro - capite* | 0,30768 | 9,47% |
| *Numero letti* | 0,29315 | 8,59% |
| *Addetti attività finaziarie e amministrative* | 0,28585 | 8,17% |
| *Alberghi con 3 stelle* | 0,27561 | 7,60% |
| *Consumi pro - capite forniture elettriche in Locali e Luoghi diversi dalle abitazioni(kWh $10^3$)* | 0,26140 | 6,83% |
| *Alberghi con 4 - 5 stelle* | 0,25493 | 6,50% |
| *Addetti commercio* | 0,25035 | 6,27% |
| *Numero unità locali nel commercio* | 0,24572 | 6,04% |
| …… | ……. | …… |

I legami tra le variabili ed il fattore esplicativo risultano piuttosto bassi; non si delinea, inoltre, una correlazione dominante rispetto alle altre. Ciò, evidentemente, non agevola la qualificazione del fattore.
Si rileva tuttavia la sostanziale omogeneità delle variabili principali, tutte descrittive dell'economia terziaria, con particolare accentuazione della componente turistica.
Appare dunque proponibile per il fattore F1 la definizione di *"Economia del terziario ed in particolare del turismo più qualificato"*.
In relazione a questa definizione, il fattore F1 individuerebbe:

- Quando positivo (ossia quando assume valori maggiori di 0): comuni caratterizzati da un elevato grado di terziarizzazione dell'economia, con un (relativamente[9]) elevato numero di unità produttive e di addetti; in questi casi si è in presenza di elevati valori sia delle imposte comunali (tasse sui rifiuti, gettito ICI, ICIAP) sia dei consumi di energia elettrica (a livello di consumi domestici e non domestici); l'incidenza più elevata della media degli addetti all'edilizia, nonché dei volumi di costruito confermerebbero l'uso turistico del territorio e lo sfruttamento della rendita di posizione dovuta, nella maggior parte dei casi, alla vicinanza del mare.
- Quando negativo: comuni caratterizzati da una relativamente bassa attività turistica e/o da arretratezza nel processo di terziarizzazione dell'economia.

### 3.1.3. - Qualificazione del secondo fattore

Il secondo fattore estratto (fattore F2) spiega il 18,1 % della varianza complessiva, e, unitamente al primo, perviene a spiegare la metà circa della varianza totale.
Anche per esso si evidenziano coefficienti di correlazione piuttosto bassi, ma accentrati su un numero minore di variabili.

| | ***Fattore 2*** | ***%*** |
|---|---|---|
| *Alberghi con meno di 3 stelle* | 0,39805 | 15,84% |
| *Aree residenziali sature e servizi* | 0,39248 | 15,40% |
| *Quota media* | 0,38670 | 14,95% |
| *Numero unità locali nel commercio* | 0,33738 | 11,38% |
| …… | ……. | ……. |

Il fattore appare meno decifrabile del precedente, poiché le variabili principali forniscono indicazioni contrastanti: ad esempio male si coniuga la presenza della variabile *Aree residenziali sature e servizi* (che assume valori elevati nei comuni costieri) con quella della variabile *Quota media* (che

[8] Nel prospetto, così come nei successivi riferiti agli altri due fattori, sono indicate soltanto le variabili esplicative che presentano i coefficienti di correlazione maggiori (positivi o negativi) nei confronti del fattore estratto.

[9] Si rammenta che l'analisi assume sempre a riferimento i valori assoluti delle variabili rapportati alla numerosità della popolazione residente (ad esempio: non *totale - addetti*, bensì *rapporto totale – addetti / totale – popolazione - residente*).
Inoltre i "giudizi" vanno sempre rapportati alla situazione media provinciale.

assume valori elevati nelle aree interne).
Una definizione proponibile, ma non del tutto convincente, è quella di *Livello di urbanizzazione*, giustificata essenzialmente dalle presenza delle variabili *Aree residenziali sature e servizi* e *N° unità locali nel commercio*.

Si rileva peraltro che:
- la presenza, come variabile principale, dell'offerta alberghiera più modesta (meno di tre stelle) potrebbe spiegare un turismo tipico di alcune aree interne, che hanno flussi turistici esigui in termini assoluti, ma probabilmente significativi in relazione alla consistenza demografica ed agli altri settori economici statisticamente qualificati;
- un'altra interpretazione per il fattore F2, positivamente correlato con il numero di unità produttive (unità locali nel commercio, alberghi con meno di tre stelle) ma al contrario negativamente correlato con il numero degli addetti alle stesse attività (- 0,23), lo proporrebbe in negativo come una possibile misura della sotto utilizzazione delle strutture turistiche - ricettive.

### 3.1.4. - Qualificazione del terzo fattore

Il terzo fattore estratto (fattore F3) spiega il 13 % della varianza complessiva, e, unitamente agli altri due, ne spiega quasi i 2/3.

| | ***Fattore 3*** | ***%*** |
|---|---|---|
| *Addetti attività manifatturiere* | - 0,47762 | 22,81% |
| *Numero unità locali con 1-5 addetti* | 0,42731 | 18,26% |
| *Numero unità locali con 6- 19 addetti* | - 0,39120 | 15,30% |
| *Floricoltura* | 0,28865 | 8,33% |
| *Arrivi turisti italiani* | 0,25252 | 6,38% |
| *Serre* | 0,24472 | 5,99% |
| *Oliveti* | - 0,21883 | 4,79% |
| ……. | ……. | ……. |

Si rileva per esso:
- la correlazione negativa con il numero di addetti nelle attività manifatturiere, coerente con l'altra negativa correlazione con il numero di imprese a maggiore intensità di manodopera;
- la correlazione positiva con le attività agricole di tipo intensivo (serre e fiori), coerente con l'altra positiva correlazione con il numero di imprese a bassissima intensità di manodopera.

La natura delle variabili significative giustifica per il fattore F3 la definizione di *Struttura produttiva e delle caratteristiche del mercato del lavoro, soprattutto di tipo agricolo.*

### 3.1.5. - I fattori estratti: valori tipici e istogrammi di frequenza

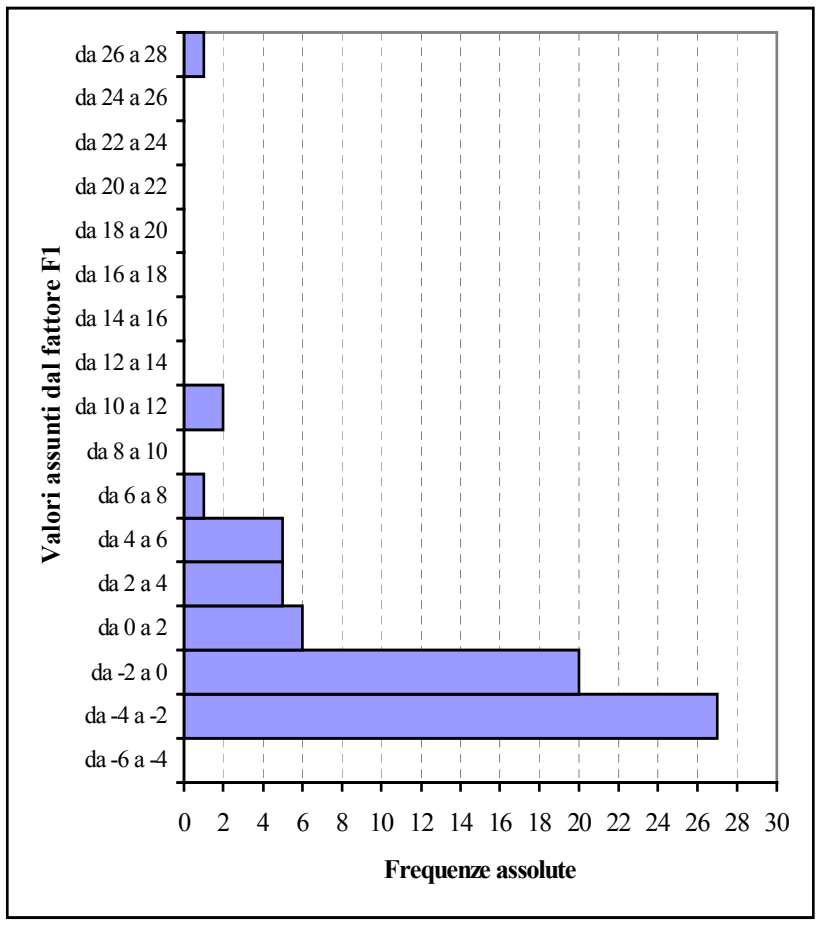
Nella tabella 5 sono indicati i valori caratteristici dei fattori estratti (minimo, massimo, valore medio e deviazione standard).
Si nota l'elevata dispersione dei dati attorno al valore medio (che per definizione è pari a 0), sintomatico della disomogeneità che caratterizza il territorio provinciale, probabilmente riconducibile al processo di crescita dualistico, che ha concentrato risorse ed attività nella fascia costiera e aperto processi di abbandono delle attività e degli insediamenti nell'entroterra.
Coerentemente con questa considerazione, si nota che nessun comune, con riferimento a tutti e tre i fattori estratti, ottiene un punteggio pari alla media matematica.

*Tabella 5 I fattori estratti. Valori tipici.*

| | **F1** | **F2** | **F3** |
|---|---|---|---|
| **Media** | 0,0 | 0,0 | 0,0 |
| **Dev. Standard** | 4,6 | 4,6 | 3,2 |
| **Minimo** | -3,8 | -5,7 | -9,4 |
| **Massimo** | 27,8 | 16,5 | 4,6 |

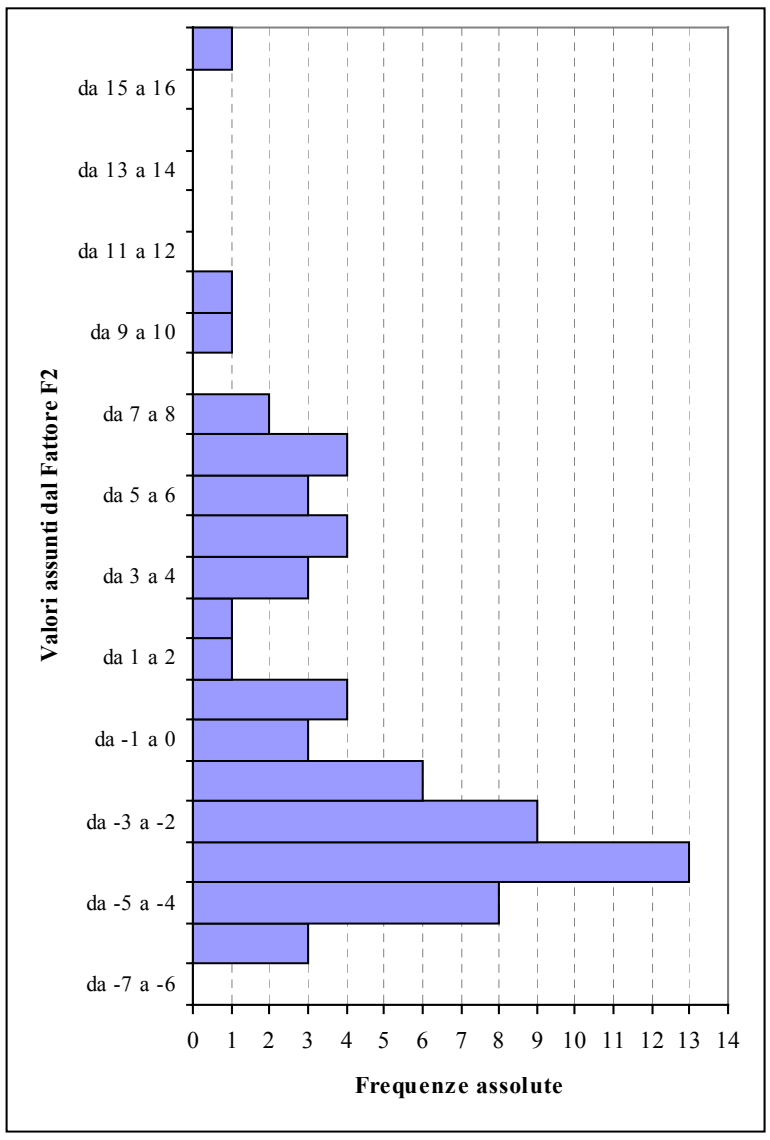
Il fenomeno risulta avvalorato dalle configurazioni grafiche dei successivi istogrammi di frequenza (Grafici 3, 4, 5), costruiti su diagrammi cartesiani che presentano, *in ascissa,* il numero di osservazioni (comuni) compresi nelle classi di valori assunti dal fattore estratto e, *in ordinata*, le classi di valori assunti dal fattore estratto. L'istogramma di frequenza relativo al fattore F1 (*Economia del terziario e del turismo in particolare*) appare nettamente asimmetrico (del tipo esponenziale negativo), con una lunga coda attribuibile ai comuni con il relativamente più elevato livello di terziarizzazione: Diano Marina, Bordighera, San Bartolomeo al Mare, Sanremo, Ospedaletti, Cervo). Per la maggiore parte dei comuni (circa il 70 %) il fattore F1 risulta negativo.

*Grafico 3 Istogramma di frequenza del Fattore F1 (Livello di terziarizzazione / sviluppo turistico).*



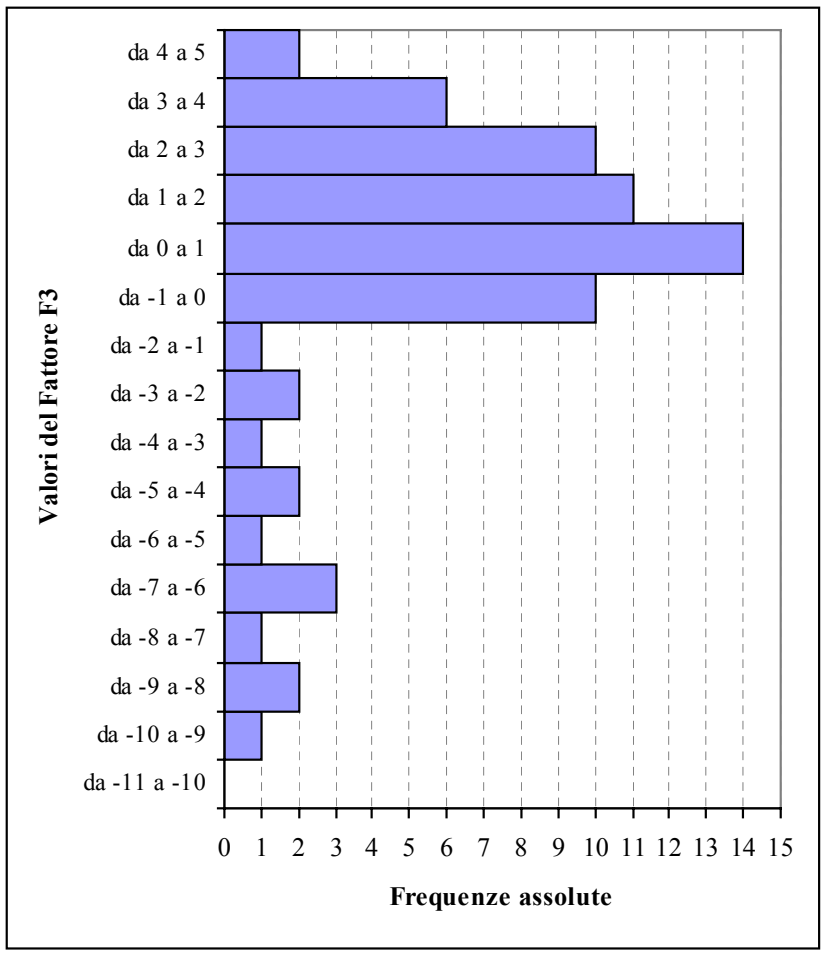

L'istogramma di frequenza relativo al fattore F2 mostra un andamento bimodale, con una coda, ancora una volta coincidente con i comuni di Diano Marina, Bordighera, Sanremo, San Bartolomeo al Mare, Ospedaletti. La sottopopolazione più numerosa è ancora la prima, cui afferisce circa il 70 % dei comuni.

*Grafico 4 Istogramma di frequenza del fattore F2 (Livello di urbanizzazione).*




L'istogramma di frequenza del fattore F3 mostra, anche esso, una tipica distribuzione bimodale, ascrivibile alla presenza di due sottopopolazioni. La prima e più numerosa (79 % dei comuni) individua aree a forte vocazione agricola (valori di F3 compresi tra - 1 e + 5).

*Grafico 5 Istogramma di frequenza del Fattore F3 (Struttura produttiva e delle caratteristiche del mercato del lavoro, soprattutto di tipo agricolo.)*




## 3.2. - CONSIDERAZIONI CONCLUSIVE: L'ANALISI DI RAGGRUPPAMENTO

L'analisi di raggruppamento è finalizzata a valutare come ciascun comune si ponga rispetto ai fattori estratti. A tale fine si utilizza il metodo della classificazione "discendente".Per ricondurre l'analisi ad una tipica chiave di lettura territoriale, questa metodologia di classificazione è incrociata con la classica ripartizione della provincia in *riviera, zona intermedia ed entroterra*, assumendosi così l'obiettivo (parallelo a quello della qualificazione socio-economica dei comuni) di individuare gruppi di comuni "omogenei" all'interno degli ambiti sub provinciali sopra definiti[10].
L'esposizione dei risultati finali è dunque sviluppata per ambiti; per ciascuno di essi:
- sono specificati i punteggi più significativi assunti dai comuni in relazione ai fattori estratti;
- sono prodotti sintetici commenti atti a supportare le risultanze statistiche;
- è definita una tabella riassuntiva, costruita in base ai punteggi ottenuti da ciascun comune.

Si rileva che l'obiettivo di definire *gruppi* "omogenei", significativamente ampi, costituiti di comuni territorialmente contigui, non è stato generalmente conseguito: tale "insuccesso" è ascrivibile ad un elevata frammentazione del tessuto insediativo ed economico imperiese, che è risultata, per questo motivo, difficile da "fotografare" mediante tecniche statistiche, ancorché evolute come quelle multivariate.

### 3.2.1. - "Riviera"

All'interno dell'ambito "*Riviera*" si individuano soltanto comuni caratterizzati da alti valori del fattore F2 (livello di urbanizzazione): si va dal 16,5 di Diano Marina, massimo provinciale, al 2,3 di Riva Ligure. Si tratta di valori sensibilmente al di sopra della media provinciale, a motivo della concentrazione, nella fascia rivierasca, delle aree residenziali sature e servizi, ossia di spazi strutturati dagli edifici e dalla viabilità, che occupano oltre l'80 % della superficie comunale[11].
Coerentemente con l'elevato indice di urbanizzazione, si registra, per gli stessi comuni, un elevato indice di terziarizzazione (fattore F1).
I valori più elevati del fattore F1 individuano i due principali poli turistici della provincia:
- il primo costituito da Diano Marina (27,8, massimo provinciale), San Bartolomeo al Mare (10,3) e Cervo (5,1);
- l'altro da Bordighera (10,9), Sanremo (6,3) e Ospedaletti (5,2).

Un "sottogruppo" è rappresentato dal ventimigliese e dall'imperiese, aree che, a fronte di punteggi molto elevati nel fattore F2 (livello di urbanizzazione), presentano valori medio-alti di F1: in entrambi i casi lo sviluppo urbano è stato indotto da motivi diversi dal turismo (la frontiera, le funzioni commerciali, industriali e amministrative).

---

[10] Tale soluzione media due diverse esigenze: da un lato, la tecnica classificatoria necessita di non ridurre eccessivamente la numerosità delle osservazioni (i comuni) all'interno dei gruppi omogenei; dall'altro, le risultanze dell'analisi consiglierebbero, in molti casi, di definire (stante il "vincolo" della contiguità territoriale) gruppi costituiti da un numero esiguo di comuni, talora potendosi configurare singoli comuni facenti gruppo a sé.

[11] Confronta C.A.I.R.E., Luglio 1995 *"…. Lungo l'arco costiero che va da Cervo a Ventimiglia – con l'unica soluzione di continuità rappresentata da Cipressa – si collocano tutti i comuni che , per densità abitativa o anche soltanto per livelli di accessibilità, possono essere in qualche modo ascritti al sistema urbano. Al di fuori di questa ristretta fascia terra (eccezione fatta per i comuni di Diano Castello, Civezza, Terzorio, Castellaro, Pompeiana, Vallebona, San Biagio della Cima, Soldano), la densità abitativa scende al di sotto dei 100 abitanti/Kmq e la popolazione accessibile entro un raggio di spostamento di 30 minuti primi è sempre inferiore a 10.000 abitanti delineando i contorni di un contesto marcatamente rurale."*

Più debole, ma sempre al di sopra della media provinciale, il livello di terziarizzazione di Vallecrosia, Riva Ligure e Camporosso, comuni per i quali si hanno punteggi più bassi (all'interno dell'ambito) anche con riferimento al fattore F2.
Nei valori (globalmente maggiori della media provinciale) del fattore F3 e nella loro differenziazione su base comunale, si riconoscono, infine, i differenti livelli di concentrazione dell'agricoltura specializzata.

*Tabella 6 Fascia territoriale "Riviera".Tabella riassuntiva.*

| Fascia Territoriale | Livello Urbanizzazione | | Indice di terziarizzazione | | Struttura agricola | |
|---|---|---|---|---|---|---|
| **Riviera** | **Alto** | Diano Marina<br>S. Bartolomeo al Mare<br>Bordighera<br>San Remo<br>Imperia<br>Ospedaletti<br>S Lorenzo al Mare<br>Taggia<br>Cervo<br>S. Stefano al Mare<br>Ventimiglia<br>Camporosso<br>Vallecrosia<br>Riva Ligure | **Alto** | Diano Marina<br>S. Bartolomeo al Mare<br>Bordighera<br>San Remo<br>Ospedaletti<br>Cervo<br>S. Stefano al Mare<br>Taggia<br>S Lorenzo al Mare<br>Imperia | **Minima** | Diano Marina<br>S. Bartolomeo al Mare<br>Cervo |
| | | | | | **Media** | Bordighera<br>Ospedaletti<br>S Lorenzo al Mare |
| | | | | | **Buona** | San Remo<br>S. Stefano al Mare<br>Imperia<br>Ventimiglia<br>Vallecrosia<br>Riva Ligure<br>Camporosso<br>Taggia |
| | | | **Medio** | Ventimiglia<br>Vallecrosia<br>Riva Ligure<br>Camporosso | | |

### 3.2.2. - "Zona intermedia"

All'interno della fascia territoriale "*Zona intermedia*" si registrano situazioni differenziate:

| Fascia Territoriale | Livello Urbanizzazione | | Indice di terziarizzazione | | Struttura agricola | |
|---|---|---|---|---|---|---|
| **Zona intermedia** | **Alto** | Chiusanico<br>Chiusavecchia<br>Pontedassio<br>Diano Castello | **Medio** | Airole<br>Chiusavecchia<br>Diano Castello<br>Dolceacqua<br>Isolabona<br>Olivetta S. Michele<br>Pontedassio<br>Villa Faraldi | **Media** | Airole<br>Apricale<br>Badalucco<br>Borgomaro<br>Caravonica<br>Carpasio<br>Cesio<br>Chiusavecchia<br>Civezza<br>Diano Arentino<br>Diano S. Pietro<br>Isolabona<br>Lucinasco<br>Olivetta S. Michele<br>Perinaldo<br>Prelà<br>Seborga<br>Vasia |
| | **Medio** | Terzorio<br>Dolceacqua<br>Castellaro<br>Olivetta S. Michele | | | | |
| | **Basso** | Airole<br>Apricale<br>Badalucco<br>Borgomaro<br>Caravonica<br>Carpasio<br>Ceriana<br>Cesio<br>Civezza<br>Diano Arentino<br>Diano S. Pietro<br>Dolcedo<br>Isolabona<br>Lucinasco<br>Montalto Ligure<br>Perinaldo<br>Pietrabruna<br>Pompeiana<br>Prelà<br>S. Biagio della Cima<br>Seborga<br>Soldano<br>Vallebona<br>Vasia<br>Villa Faraldi | **Basso** | Apricale<br>Badalucco<br>Borgomaro<br>Caravonica<br>Carpasio<br>Castellaro<br>Ceriana<br>Cesio<br>Chiusanico<br>Civezza<br>Diano Arentino<br>Diano S. Pietro<br>Dolcedo<br>Lucinasco<br>Montalto Ligure<br>Perinaldo<br>Pietrabruna<br>Pompeiana<br>Prelà<br>S. Biagio della Cima<br>Seborga<br>Soldano<br>Terzorio<br>Vallebona<br>Vasia | **Buona** | Villa Faraldi<br>Castellaro<br>Ceriana<br>Chiusanico<br>Diano Castello<br>Dolceacqua<br>Dolcedo<br>Montalto Ligure<br>Pietrabruna<br>Pompeiana<br>Pontedassio<br>S. Biagio della Cima<br>Soldano<br>Terzorio<br>Vallebona |

1. i comuni posti immediatamente alle spalle di quelli costieri presentano caratteristiche a questi assimilabili per uno o più dei tre fattori;

2. negli altri casi, si osservano valori al di sotto della media per i fattori F1 e F2 e valori medi per il fattore F3: si tratta di comuni per i quali non si configura una struttura agricola in grado di sopperire alle carenze del secondario e terziario, cosicché l'occupazione risulta tendenzialmente carente.

Nel primo gruppo si collocano zone che risentono dell'effetto attrattivo di aree omogenee costiere di relativamente elevato potenziale (cui tendono ad assomigliare); queste zone presentano valori decrescenti dei fattori F1 e F2 all'allontanarsi dal centro di gravitazione:
- Olivetta San Michele ed Airole nei confronti di Ventimiglia (polo di gravitazione),
- Pontedassio, Chiusavecchia, Chiusanico nei confronti di Imperia,
- Isolabona e Dolceacqua nei confronti di Camporosso,
- Diano Castello e Villa Faraldi nei confronti del dianese.

Sempre nel primo gruppo ricadono comuni, che, con valori bassi di terziarizzazione e di urbanizzazione, sembrano risentire dell'effetto gravitazionale della costa soprattutto per il settore agricolo (F3):
- nell'area della bassa valle Argentina: San Biagio della Cima, Vallebona, Soldano,
- nell'immediato entroterra di Riva Ligure e San Lorenzo al Mare: Terzorio, Pompeiana, Castellaro, Dolcedo, Civezza.

In detti comuni si ripropongono infatti le stesse colture agricole in cui sono specializzati i limitrofi – e costieri – comuni confinanti, in particolare la floricoltura.

Al tipo *sub 2* appartengono i restanti comuni, che presentano connotati più assimilabili all'interno, con rarefazione degli insediamenti produttivi e urbani. L'analisi indica pertanto una tendenziale "scomparsa" della fascia intermedia il cui territorio
- in alcune zone, tende ad un'assimilazione / integrazione con i comuni costieri su cui esse gravitano;
- in altre, non ha connotati propri che la differenzino significativamente dalle zone più interne della provincia.

### 3.2.3. - "Entroterra"

Se i risultati dell'analisi multivariata riescono a fornire chiavi di lettura diversamente soddisfacenti per i comuni della riviera e per un gruppo di comuni della fascia intermedia (quelli definiti *sub 1*), non riescono a cogliere invece le specificità dei comuni della fascia territoriale *"entroterra*", o, meglio, restituiscono la fotografia di un contesto territoriale il cui unico elemento omogeneo risulta essere l'elevata frammentarietà e despecializzazione. Per tutti fattori si registra un'ampia prevalenza di punteggi bassi.

*Tabella 8 Fascia territoriale "Entroterra".Tabella riassuntiva.*

| Fascia Territoriale | Livello Urbanizzazione | | Indice di terziarizzazione | | Struttura agricola | |
|---|---|---|---|---|---|---|
| **Entroterra** | **Basso** | Montegrosso Pian Latte | **Basso** | Montegrosso Pian Latte | **Bassa** | Mendatica |
| | | Castel Vittorio | | Aurigo | | Pornassio |
| | | Aurigo | | Vessalico | | Triora |
| | | Armo | | Ranzo | | Aquila di Arroscia |
| | | Vessalico | **Medio** | Castel Vittorio | | Armo |
| | | Bajardo | | Pigna | | Cosio di Arroscia |
| | | Rezzo | | Borghetto d'Arroscia | | Rezzo |
| | | Aquila di Arroscia | | Rezzo | | Bajardo |
| | | Ranzo | | Bajardo | | Borghetto d'Arroscia |
| | | Pigna | | Armo | | Molini di Triora |
| | | Cosio di Arroscia | | Molini di Triora | | Castel Vittorio |
| | | Borghetto d'Arroscia | | Pieve di Teco | **Media** | Rocchetta Nervina |
| | **Medio** | Rocchetta Nervina | | Cosio di Arroscia | | Pieve di Teco |
| | | Mendatica | **Alto** | Aquila di Arroscia | | Montegrosso Pian Latte |
| | | Molini di Triora | | Mendatica | | Aurigo |
| | **Alto** | Pieve di Teco | | Triora | | Vessalico |
| | | Pornassio | | Pornassio | | Ranzo |
| | | Triora | | Rocchetta Nervina | | Pigna |

Nella fascia "alta" si collocano i comuni di Pieve di Teco, Pornassio, Triora, caratterizzati da un livello di urbanizzazione (fattore F2) relativamente alto. I restanti comuni si propongono come un ideale *continuum* di quelli appartenenti alla fascia intermedia del tipo *sub 2*: nonostante la presenza in loco di emergenze di valore storico, paesaggistico e naturalistico, non si rilevano situazioni di significativo sviluppo connesse all'uno o all'altro fattore.
Neppure la struttura agricola (fattore F3), salvo circoscritte eccezioni (ad esempio Rocchetta Nervina posizionata nella zona di produzione del Rossese), appare particolarmente significativa.