

più rilevanti nel determinare la scelta dell'alloggio.

Per problemi di questo genere dunque opportuno utilizzare una metodologia costruttiva che:

- tenga conto delle peculiarità del contesto d'analisi;
- rispetti la multidimensionalità delle valutazioni, considerando simultaneamente le diverse variabili (e tenendo conto implicitamente dei loro trade-offs);
- segmenti la popolazione considerata sulla base delle relazioni che intercorrono tra le variabili che la descrivono e quelle dell'altro insieme.

Per gli esempi citati, i segmenti di nuclei andrebbero costruiti a partire da quella parte di varianza delle variabili socioeconomiche che risulta correlata alle variabili che descrivono gli alloggi, trascurando o minimizzando gli aspetti residui.

Un'analisi canonica, seguita da una procedura di classificazione particolarmente accurata, sembra atta allo scopo. La figura 1 schematizza il metodo proposto. Com'è noto, la correlazione canonica è un metodo statistico di nobili ascendenze - a paritè da Hotelling, negli anni trenta - doverosamente trattato da tutti i manuali di statistica multivariata, ma del quale altrettanto concordemente vengono sottolineati i risultati empirici deludenti (si veda ad esempio Maxwell, 1981). L'uso che qui se ne propone sembra invece suggerire un filone dalle interessanti possibilità applicative.

I due insiemi di variabili considerate (nel nostro esempio, le variabili che descrivono i nuclei e quelle relative agli alloggi) definiscono due spazi vettoriali, nei quali si determinano opportunamente due insiemi di nuove variabili $\{z_j\}$ e $\{z'_j\}$ dette *canoniche*, combinazioni lineari delle variabili di partenza e tali che

$$\begin{aligned} \text{corr}(z_j, z'_j) &= \max \\ \text{corr}(z_r, z_t) &= 0, & \text{corr}(z'_r, z'_t) &= 0 & \text{per } r \text{ diverso da } t \end{aligned}$$

Risulta cioè massima la correlazione tra variabili canoniche corrispondenti nei due spazi, mentre risultano *non correlate* variabili canoniche diverse di uno stesso set. Ogni variabile canonica riassume dunque, in ciascuno spazio e con capacità esplicativa via via decrescente, la massima quota possibile della *varianza comune alle due descrizioni* non spiegata dalle variabili canoniche già determinate in precedenza.

Limitando opportunamente il numero dei fattori considerati si ot-