

Come già notato, nella tabella complessa l'inerzia totale è il valore medio delle 35 sottotabelle. La tabella 1.5 mostra l'inerzia di tutte le 35 tabelle nonché l'inerzia media delle caratteristiche sociodemografiche e dei sette indicatori di identità nazionale. Si può notare che le inerzie più elevate appartengono alle tabelle incrociate con i paesi, ovvero la variazione maggiore nei dati è causata dalle differenze tra paesi. Inoltre, non ci sono quasi differenze di sesso per i sette item sull'identità nazionale, anche se ci sono alcuni risultati che potrebbe valere la pena riportare. Ad esempio, l'associazione tra "sesso" e "sport internazionale" è molto più piccola dell'associazione tra "sesso" e "vergognarsi del paese".

1.6 Analisi delle corrispondenze multiple

Negli esempi precedenti abbiamo analizzato la relazione tra due variabili o tra due diversi insiemi di variabili. In questa sezione, siamo interessati alle relazioni all'interno di un insieme di variabili, ad esempio, le interrelazioni tra le affermazioni sull'identità nazionale. Così, ad esempio, potremmo scoprire se esiste un'associazione tra un "forte accordo verso lo sport internazionale" e un "forte accordo verso le persone che dovrebbero sostenere il proprio Paese". Nella precedente analisi delle tabelle impilate, abbiamo potuto vedere solo se queste categorie avevano la stessa associazione con le variabili sociodemografiche.

Questo nuovo caso, che ricorda l'analisi delle componenti principali, coinvolge tutte le tavole incrociate di un insieme di variabili, come gli indicatori di identità nazionale, con se stesse. Assemblando tutte queste tabelle incrociate in una supermatrice quadrata di tabelle incrociate, otteniamo quella che nella letteratura CA è conosciuta come *matrice di Burt*, che indicheremo con C . In alternativa, è possibile costruire una struttura dati nota come *matrice indicatore* basata sulla dati originali. La matrice degli indicatori, indicata con Z , è una tabella intervistati per categorie con tante righe quanti intervistati (6066 nel nostro esempio) e tante colonne quante sono le categorie di risposta (35 per i sette indicatori di identità nazionale). Gli elementi di Z sono zeri tranne gli uno nelle posizioni per indicare le categorie di risposta di ciascun intervistato (Z è spesso chiamata matrice di variabili dummy). La matrice di Burt è correlata semplicemente alla matrice dell'indicatore come segue: $C = Z^T Z$. Se il consueto algoritmo CA viene applicato a una matrice di indicatori o a una matrice di Burt, il metodo viene chiamato *analisi delle corrispondenze multiple* (MCA). Nell'MC

tra variabili che descrivono e variabili da descrivere, come nel caso delle CA semplici di tabelle singole o impilate. In MCA tutte le variabili hanno lo stesso stato. La relazione tra le analisi di C e Z nell'MCA è discussa in modo approfondito nel Capitolo 2. Di seguito, illustriamo il metodo analizzando la matrice indicatore 6066×35 Z che codifica le risposte alle sette domande sull'identità nazionale. La soluzione grafica è riportata nella Figura 1.6.

Esaminando la Figura 1.6, vediamo che la prima dimensione contrasta i forti accordi e i forti disaccordi (parte positiva) delle categorie intermedie (parte negativa). Con due eccezioni (affermazioni b e g), la seconda dimensione contrappone le affermazioni positive a quelle negative. Pertanto, può essere vista come una dimensione complessiva verso l'identità nazionale, con un'identità nazionale relativamente alta nella parte positiva e un'identità nazionale relativamente bassa nella parte negativa. Le variabili a, c, d, e ed f formano un ferro di cavallo, una struttura tipica che di solito troviamo nei dati categorici ordinati (per maggiori dettagli, vedere i Capitoli 2 e 4). Ci sono però due punti da menzionare. Innanzitutto, né l'elemento b, "ci sono alcune cose...", né l'elemento g, "Sono spesso meno orgoglioso...", soddisfano questa struttura, e forse anche peggio, le categorie più opposte "b1" e "b5" come così come "g1" e "g5" sono vicini tra loro. Una ragione di questo risultato potrebbe essere che un numero significativo di intervistati ha frainteso la direzione della domanda, il che porterebbe a una tale struttura (Blasius e Thiessen 2001b). Un altro motivo è che due dimensioni non sono sufficienti per rispecchiare adeguatamente la struttura di queste due variabili. In una soluzione di dimensione superiore "b1" e "b5" così come g1 e g5 potrebbero essere lontani l'uno dall'altro. In secondo luogo, il ferro di cavallo appartiene alla prima dimensione, cioè la seconda dimensione è più importante per l'interpretazione sostanziale. Ciò potrebbe essere causato dall'analisi congiunta dei dati di cinque paesi, in cui gli intervistati di paesi diversi potrebbero comprendere le singole domande (leggermente) in modo diverso (vedere Capitolo 20).

Si noti che non abbiamo indicato le percentuali di inerzia spiegate nella Figura 1.6. Esistono diversi metodi per scalare la soluzione in MCA che modificano l'adattamento (vedere Capitoli 2 e 3), ma la struttura complessiva delle categorie variabili nello spazio rimane la stessa e, quindi, anche l'interpretazione sostanziale. Utilizzando la matrice degli indicatori come input, la misura dell'adattamento in termini di inerzia spiegata è fortemente sottostimata. Varie soluzioni a questo problema sono proposte nel Capitolo 2 (vedi Sezione 2.3.4).

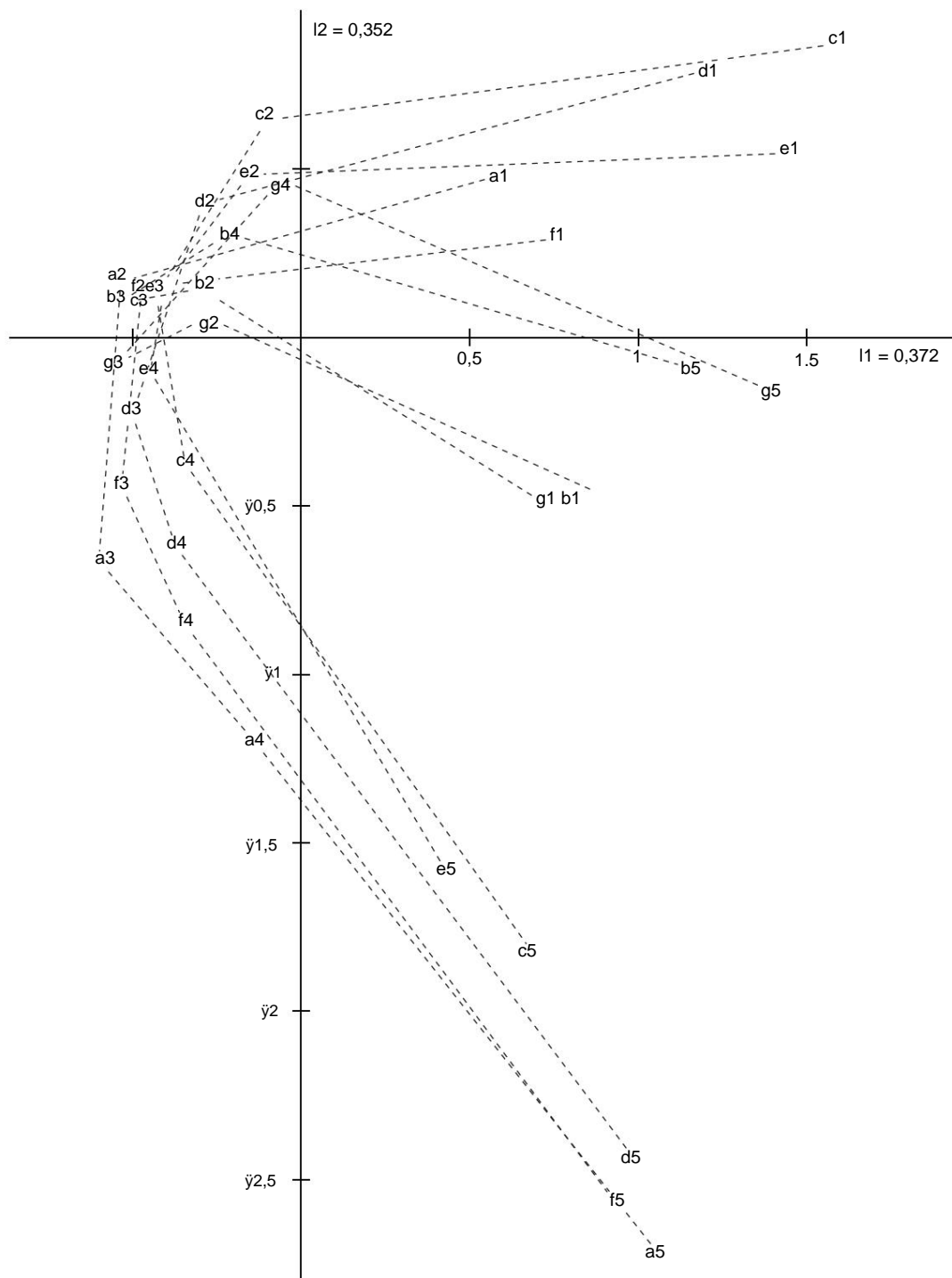


Figura 1.6 Mappa MCA degli indicatori sull'identità nazionale.