

12

L'analisi delle corrispondenze

L'analisi delle corrispondenze è una tecnica statistica multivariata per visualizzare e descrivere le associazioni fra due o più variabili qualitative le cui modalità sono state classificate in una tabella di contingenza a due o più vie. Quando la tabella di contingenza di partenza ha dimensioni superiori a 2×2 , risulta difficile una rappresentazione grafica diretta che ponga in luce le relazioni tra le diverse modalità che presentano le due variabili in esame. Una volta accertato che il valore del test χ^2 è significativo (v. sezione 7.3), si pone il problema di capire quali sono le combinazioni di righe e colonne nella tabella di contingenza che presentano la relazione più forte.

Il procedimento che sta alla base del metodo consiste nel “geometrizzare il problema” (Greenacre, 1984), nel senso che le righe e le colonne della tabella di contingenza, opportunamente ricodificate, vengono intese come punti geometrici in due diversi spazi multidimensionali, nei quali è definita una distanza, dando vita quindi a due “nuvole” di punti. Per poterne decifrare la struttura, ciascuna nuvola viene proiettata in un sottospazio a due dimensioni (un piano). Nella stessa maniera delle componenti principali, i sottospazi sono scelti in maniera ottimale, in modo tale che i punti proiettati diano una rappresentazione il più possibile fedele, della nube originaria. Grazie alle preventive trasformazioni operate simmetricamente sulle righe e sulle colonne della matrice dei dati, è possibile far coincidere i due piani, ottenendo così una rappresentazione grafica unica sulla quale le righe e le colonne della matrice vengono ad essere rappresentate dalle proiezioni dei loro punti rappresentativi. L'interpretazione delle prossimità tra proiezioni sulla mappa conduce l'analista a risalire alle prossimità tra punti delle nuvole nel loro spazio multidimensionale e perciò a riconoscere i legami tra le caratteristiche il cui l'insieme dei dati è ripartito.

La metodologia dell'analisi delle corrispondenze è stata introdotta dalla scuola francese guidata da J.P. Benzecri all'inizio degli anni sessanta (Benzecri, 1980). Quando l'analisi delle corrispondenze è relativa a due variabili si parla di analisi delle corrispondenze semplici. Qualora invece si considerano congiuntamente più di due variabili che danno luogo a tabelle di contingenza a tre o più vie, si parla di analisi delle corrispondenze multiple.

L'analisi delle corrispondenze è impiegata in contesti molto svariati. Ad esempio, dal punto di vista di marketing, l'analisi della corrispondenza risponde a domande come: