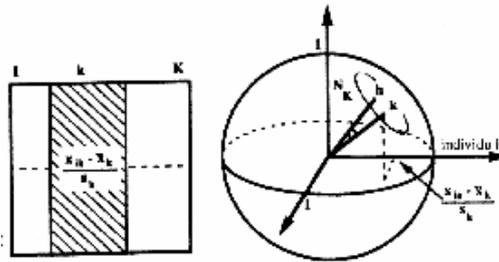


*Appunti per l'interpretazione dei dati dell'ACP (analisi delle componenti principali) e dell'AFC (analisi fattoriale delle corrispondenze).
(F. Spagnolo)*

ACP

- dalle variabili ai fattori: matrice di correlazione fra le variabili osservate e i “fattori” sottostanti.
- rappresentazione geometrica in uno spazio euclideo (tanti assi quanti sono i fattori): coefficiente di correlazione tra due fattori equivale al coseno dell'angolo fra i due assi corrispondenti ai fattori. (assi ortogonali $\cos 90^\circ=0$). Il valore del coseno è importante da controllare.
- L'ACP non è analisi fattoriale, serve per trovare tante componenti (fattori) quante sono le variabili, per cui tutta la varianza delle variabili è spiegata dalle componenti principali.
- Trasforma un insieme di variabili in altrettante non correlate.
- Non è un modello ma una trasformazione lineare di dati.
- 1^a componente quella che spiega più varianza, ecc...
- I fattori sono in numero minori delle variabili.
- 1° fattore spiega il massimo di varianza possibile, il 2° fattore spiega il massimo di varianza residua, ecc...
- I fattori sono tra loro ortogonali.
- Questioni a cui risponde l'ACP:
 1. Quali sono gli individui che si rassomigliano? Quali sono quelli che differiscono? Esistono dei gruppi omogenei di individui? Possiamo mettere in evidenza una tipologia di individui?
 2. Quali sono le variabili legate positivamente tra loro? Quali quelle che si oppongono (legate negativamente)? Possiamo mettere in evidenza una tipologia di variabili?
 3. Si possono riassumere un certo numero di insiemi di variabili chiamate componenti principali che rappresentano questo gruppo di variabili legate tra loro.
- Una “nuvola” è costituita dall'insieme delle distanze inter-personali:



- La forma della nuvola è n-dimensionale l'ACP fornisce delle immagini che si avvicinano il più possibile alla situazione reale. Gli assi fattoriali definiscono dei piani fattoriali sui quali si proietta la Nuvola N_j .
- L'ACP ci fornisce delle variabili sintetiche costituendo un riassunto dell'insieme delle variabili iniziali che sono la base di una rappresentazione piana delle variabili e dei loro angoli.
- Interpretazione grafica:
 1. Se due variabili sono correlate positivamente, sono situate dalla stessa parte su di un asse. Sull'asse corrispondente della nuvola degli individui questi hanno dei forti valori per queste due variabili e si situano dalla stessa parte delle variabili. Dalla parte opposta nel caso contrario. Gli individui che inducono delle correlazioni forti con le variabili sono individuati facilmente.
 2. Le variabili sono dei vettori e non dei punti. Non è la prossimità tra individuo e un insieme di punti rappresentante delle variabili che è importante ma l'allontanamento dell'individuo nella direzione di questo insieme di variabili.
- Il s.i. di elementi che contribuisce all'inerzia di un asse è la sua somma di contribuzioni degli elementi che lo compongono. Questo rapporto è prezioso per mettere in evidenza il s.i. di elementi che hanno contribuito principalmente alla costruzione dell'asse e sul quale si appoggerà in primo luogo l'interpretazione.
- La distanza tra due punti sarà interpretata come una somiglianza.
- Le coordinate di una variabile si interpretano come coefficienti di correlazione con i fattori sugli individui.
- I coefficienti di correlazione delle variabili attive con un fattore (generalmente il 1°) sono tutti positivi.

Le fasi cronologiche:

1. Bilancio sulle inerzie associate ai differenti fattori, che non si preoccupa del significato dei fattori, ma si fonda su indici numerici.

2. L'interpretazione dei fattori, difficilmente formalizzabile, che da un posto importante alle conoscenze sul problema studiato esterno alla tabella dei dati.

AFc

- Si analizzano variabili ed individui, studiandone il legame e non privilegiandone nessuna.
- La rassomiglianza tra due linee e due colonne è definita da una distanza tra i loro profili. Questa distanza è il χ^2 .
- Come nell'apc si chiama fattore l'insieme delle coordinate delle proiezioni dei punti di una nuvola su uno dei suoi assi fattoriali; i fattori sulle linee sono le proiezioni di N_i e i fattori sulle colonne le proiezioni di N_j .
- Gli elementi pesanti attirano il baricentro, una colonna j attira una linea i se il valore f_{ij} è elevato.
- Sui grafici di rappresentazione simultanea di linee e colonne, la posizione relativa di due punti di uno stesso insieme (linee o colonne) si interpreta come distanza quando la posizione di un punto di un insieme e di tutti i punti dell'altro insieme si interpreta come baricentro. Ogni associazione tra una linea e una colonna suggerita da una prossimità sul grafico deve essere controllata sulla tabella dei dati.
- Se i dati sono concentrati in un punto (il loro baricentro), l'inerzia totale è nulla e l'afc non da nessun fattore. Per ottenere fattori è necessaria una distribuzione non omogenea.
- Gli indici per interpretare l'afc sono analoghi all'apc.
- Variabili supplementari e individui supplementari non intervengono nel calcolo degli assi.
- La distanza tra due punti si interpreta come rassomiglianza tra i profili di questi punti. L'origine degli assi è confusa con il profilo medio.

ACM

(Analisi delle Corrispondenze Multiple)

- Si individuano classi di variabili e/o di individui.
- L'acm è una analisi fattoriale su una tabella disgiunta completa (vero/falso).
- Un piccolo numero di individui può essere significativo per un fattore. (sopprimere individui per vedere se c'è invarianza).

Considerazioni generali

- L'analisi fattoriale dà "senso" ai dati.
- L'interpretazione ha un ruolo fondamentale.
- La % d'inerzia estratta dai fattori va interpretata in funzione del numero di variabili: 10% è poca per 10 variabili ma è sufficientemente forte per 100 variabili.
- Se un fattore ha un contributo di individui <50% non è significativo.
- Se un fattore è individuato da pochi individui, è possibile rifare i calcoli eliminando questi individui (o osservazioni) e vedere l'effetto che fa...
- Variabili più numerose inducono un significato più forte.
- ACP → ACF:
 - La prima rappresentazione si riferisce all'apc: introduce gli aspetti generali comuni a tutti i metodi;
 - Le altre rappresentazioni si appoggiano su questo primo schema, sviluppando unicamente i punti sui quali il tentativo di interpretazione differisce, tra il metodo esaminato e l'apc.
- E' possibile immettere i dati non necessariamente binari:
 - Quando le unità di misura differiscono da una variabile all'altra, si ricorre alla riduzione di variabili (centratura). Riducendo si accorda a ogni variabile la stessa importanza a-priori nell'analisi. L'apc allora si dice normata. Se lo scarto-tipo varia poco tra le variabili (ad es. di un punto) le due analisi normate e non normate conducono a risultati vicini.