Il metodo delle componenti principali

Se, abbandonando l'esigenza della semplificazione, ci si propone di ricavare con il metodo centroide un numero di fattori comuni pari al numero dei tests, il problema non è più indeterminato, poiché il numero delle equazioni è pari al numero delle incognite.

Consideriamo, per semplicità, il caso di tre tests (1, 2, 3) e tre fattori (A, B, C) [estensibile a n tests e n fattori]; possiamo impostare le tre seguenti equazioni di specificazione(1):

$$z_{1p} = a_1 A_p + b_1 B_p + c_1 C_p$$
$$z_{2p} = a_2 A_p + b_2 B_p + c_2 C_p$$
$$z_{3p} = a_3 A_p + b_3 B_p + c_3 C_p$$

in cui z_{1p}, z_{2p}, z_{3p} sono i risultati, espressi in unità standard, di una persona p nei tre tests 1, 2, 3

A_p, B_p, C_p sono le misure, espresse in unità standard, dei 3 fattori del soggetto p

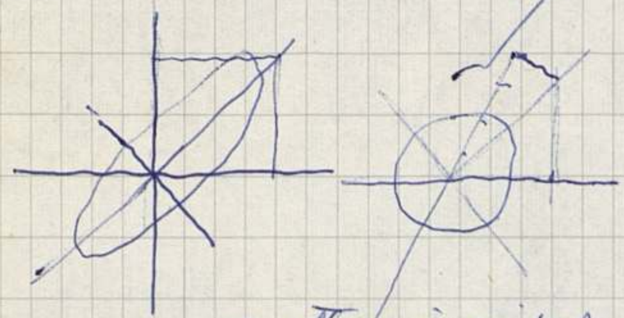
$a_1 a_2 a_3$ $b_1 b_2 b_3$ $c_1 c_2 c_3$ sono i coefficienti di saturazione dei 3 tests 1, 2, 3 nei 3 fattori A, B, C

(Salvo la restrizione relativa al numero dei fattori (uguale al numero dei tests))

(1) Si noterà che mancano i fattori specifici. ~~Questa è una~~ Quella data è una soluzione generale in cui ogni fattore può essere presente in ogni test. ~~La presenza di un~~ Il fattore specifico ~~proprio di un test~~ (cioè un fattore che ~~abbia~~ ha saturazione zero in tutti i tests tranne in uno) rappresenta un caso particolare che deriva dal caso generale in seguito alla introduzione della restrizione suddetta (per ogni test c'è un fattore che ha ~~saturaz. soltanto~~ saturazione non nulla soltanto in quel test, e nulla in tutti gli altri).

Del resto è ovvio che estraendo un numero di fattori uguale al numero dei tests non vi è posto per i fattori specifici, il cui numero è uguale al numero dei tests

~~Il metodo di~~ Il metodo di Hotelling, cioè la determinazione delle componenti principali (espressione che equivale a quella di fattori) presenta certe analogie con il metodo centroide.



Per ~~presentazione~~ ~~osservare~~ mettere in evidenza le caratteristiche conviene di ~~considerare~~ esaminare, sotto un angolo diverso, il metodo centroide.

Consideriamo a tale scopo una batteria di 3 tests, 1, 2, 3 ~~ed ammettiamo~~ dalla quale ci proponiamo di ricavare 3 fattori (A, B, C). In questo caso abbiamo cioè rinunciato al vantaggio della descrizione più economica (numero di fattori minore del numero di tests) ma il numero delle equazioni ~~sarà~~ è uguale a quello delle incognite, e ~~il numero di~~ il problema non è più indeterminato. Si avrà infatti

$$z_{1p} = a_1 A_p + b_1 B_p + c_1 C_p$$
$$z_{2p} = a_2 A_p + b_2 B_p + c_2 C_p$$
$$z_{3p} = a_3 A_p + b_3 B_p + c_3 C_p$$

Risolvendo il sistema delle 3 equazioni per A_p, B_p, C_p avremo 3 equazioni che ci daranno i 3 valori esatti (non le loro stime) di A, B, C in termini di z_1, z_2, z_3, e dei 9 coefficienti.

Occorre dunque disporre anche dei valori numerici dei 9 coefficienti, che a differenza dei valori di z e A, B, C sono costanti per i 3 tests e che si ottengono appunto mediante la tecnica dell'analisi centroide.

Procedendo al calcolo, col metodo centroide, si constata che in questo caso c'è un importante cambiamento, per quanto riguarda le communanze. Infatti, ~~mentre~~ l'equazione della varianza per 3 fattori comuni

$1 = a_t^2 + b_t^2 + c_t^2 + u_t^2$ non essendovi il fattore specifico (essendo $u_t = 0$) si riduce a $1 = a_t^2 + b_t^2 + c_t^2$; e la communanza $h^2 = 1 - u_t^2$ diventa uguale a 1.

~~Avremo allora~~ La matrice delle correlazioni non richiede dunque la stima delle communanze, che sono tutte uguali a 1

$$\begin{array}{ccc} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{array}$$

tradotta in termini fattoriali, ~~con 1 al posto delle communanze, cioè~~

$$\begin{array}{lll} a_1^2+b_1^2+c_1^2=1 & a_1a_2+b_1b_2+c_1c_2 & a_1a_3+b_1b_3+c_1c_3 \\ a_2a_1+b_2b_1+c_2c_1 & a_2^2+b_2^2+c_2^2=1 & a_2a_3+b_2b_3+c_2c_3 \\ a_3a_1+b_3b_1+c_3c_1 & a_3a_2+b_3b_2+c_3c_2 & a_3^2+b_3^2+c_3^2=1 \end{array}$$

la matrice che non pone nessun nuovo problema e consente l'applicazione ~~diretta~~ del metodo centroide. ~~senza la previa stima delle communanze.~~

Il procedimento di calcolo ~~a sua volta~~ equivale alla soluzione di 9 equazioni a 9 incognite, ~~in quanto dà come risultati i valori dei 9 coefficienti, utilizzando dei dati che non consentono l'ostacolo di impostazione di un minor numero di equazioni.~~

È opportuno precisare quali sono queste equazioni. Sei sono presenti nella matrice delle correlazioni, ~~cioè~~ tre dedotti dall'equazione della varianza e tre dall'equazione della correlazione in termini fattoriali e cioè

$$\begin{array}{ll} a_1^2+b_1^2+c_1^2=1 & a_1a_2+b_1b_2+c_1c_2=r_{12} \\ a_2^2+b_2^2+c_2^2=1 & a_1a_3+b_1b_3+c_1c_3=r_{13} \\ a_3^2+b_3^2+c_3^2=1 & a_2a_3+b_2b_3+c_2c_3=r_{23} \end{array}$$

Altre ~~tre~~ due equazioni si deducono dalla condizione che viene presupposta per rendere possibile l'applicazione del procedimento delle sommatorie, ~~che~~ cioè la somma algebrica delle saturazioni sia uguale a zero; cioè, nel caso qui considerato

$$\cancel{a_1 + a_2 + a_3 = 0}$$
$$b_1 + b_2 + b_3 = 0$$
$$c_1 + c_2 + c_3 = 0$$

A queste 8 equazioni se ne aggiunge una 9ª ricavata dalla condizione che viene posta implicitamente in seguito alla necessità di cambiare i segni per rirendere ~~fa~~ calcolare ~~il 2º~~ le saturazioni nel 2º e nel 3º fattore. < generalizzare>

Abbiamo in tal modo definito una forma di analisi fattoriale che differisce da quelle finora esaminate non tanto per il procedimento (che, tranne per la scelta dei ~~fattori~~ numeri da inserire nei vuoti della matrice delle correlazioni è identico a quello dell'analisi centroide di Thurstone) ma per la finalità e il risultato. In questo caso infatti non si approfitta della possibilità di ridurre il rango della matrice delle correlazioni inserendo numeri adatti a tale scopo ~~al posto della [illegible]~~ al posto degli elementi mancanti della diagonale principale; il rango è pari all'ordine della matrice, e quindi il numero dei fattori comuni è pari al numero dei tests. ~~Tali fattori~~ Ma ~~come~~ con ~~ogni~~ le successive estrazioni dei ~~fatt~~ coefficienti di saturazione fattoriale si ~~ottengono~~ estraggono quote decrescenti di varianza. ~~Ha~~ La quota massima di varianza si estrae col primo fattore, col secondo fattore si estrae una quota di varianza inferiore a quella estratta col primo fattore ma superiore a quella che sarà estratta col terzo, e così di seguito. ~~A~~ Ad un certo punto la quota residua è tanto piccola da poter essere trascurata, e perciò anche con questo metodo si ~~possono~~ può giungere alla soluzione economica di avere un numero di fattori comuni minore del numero di tests. In