

Análisis Factorial de Correspondencias

Dr. Oldemar Rodríguez Rojas

11 de febrero de 2008

Contents

1. Análisis Factorial de Correspondencias	v
1. Los datos	VII
2. El AFC como un ACP de las dos tablas de perfiles	VII
3. ACP de la tabla de perfiles de la primera variable cualitativa	VII

Análisis Factorial de Correspondencias

Theorem 1 An CA is the PCA of the profiles table of the first qualitative variable, taking the rows like individuals and the columns like variables. This PCA is made using the distance χ^2 . Those main components are the eigenvectors a_1^k .

- We use the distance χ^2 :

$$d^2(i, i') = \sum_{j=1}^{m_2} \frac{1}{f_{\cdot j}} \left(\frac{f_{ij}}{f_{i \cdot}} - \frac{f_{i'j}}{f_{i' \cdot}} \right)^2.$$

- The first eigenvalue of this PCA is always 1.
- Row profile: $pf_i = \left(\frac{k_{i1}}{f_i}, \dots, \frac{k_{ip}}{f_i} \right)$, where $f_i = \sum_{j=1}^p k_{ij}$
- Column profile: $pc_s = \left(\frac{k_{1s}}{c_s}, \dots, \frac{k_{ns}}{c_s} \right)$ where $c_s = \sum_{i=1}^n k_{is}$

Fue propuesto en 1960 por el francés J.P. Benzecri.

1. Los datos

- Dos tablas X_1 y X_2 de códigos disyuntivos completos de dos variables cualitativas, o bien:
- La tabla de contingencia o tabla cruzada $K = X_1 X_2$.
- $X_1 \in M_{n \times m_1}$, es decir el número de modalidades de la primera variable cualitativa es m_1 .
- $X_2 \in M_{n \times m_2}$, es decir el número de modalidades de la segunda variable cualitativa es m_2 .
- $K = X_1 X_2 \in M_{m_1 \times m_2}$.
- Supóngamos S.P.G. que $m_2 > m_1$.

2. El AFC como un ACP de las dos tablas de perfiles

- AFC = Caso particular del AC (Análisis Canónico).
- AFC = La sobreposición de dos ACPs, uno de la tabla perfiles fila y otro de la tabla perfiles columna.

3. ACP de la tabla de perfiles de la primera variable cualitativa

- Se utilizará la distancia χ^2 para medir la distancia entre dos puntos (filas) i e i' en \mathbb{R}^{m_1} de la tabla de perfiles fila, es decir:

$$d^2(i, i') = \sum_{j=1}^{m_2} \frac{1}{f_{\cdot j}} \left(\frac{f_{ij}}{f_{i \cdot}} - \frac{f_{i'j}}{f_{i' \cdot}} \right)^2.$$

- Así, dos modalidades de la primera variable son cercanas si los perfiles fila son cercanos.

Theorem 2 Un AFC es el ACP de la tabla de perfiles fila de la primera variable cualitativa, tomando las filas como individuos y las columnas como variables. Este ACP se efectúa usando la distancia χ^2 . Las componentes principales son los vectores propios a_1^k .

Observación 1 ■ El ACP anterior no fue centrado.

- El primer valor propio de este ACP siempre es 1.

Theorem 3 Hacer ACP de la tabla de perfiles fila centrada es equivalente a hacer un ACP de la tabla de perfiles fila NO centrada e ignorar el primer valor y vector propio.

Observación 2 Para representar las modalidades de la segunda variable es suficiente cambiar el rol de i y de j .