Introducción

En el tema del Análisis de Componentes Principales hemos introducido el problema de llegar a comprender de una forma sencilla y simplificada a qué se debe la diversidad de calificaciones que se observa en nueve asignaturas de los 29 alumnos de un curso, según se indica en el cuadro siguiente (lista detallada de datos en el ejemplo al final del tema dedicado al ACP):

Calificaciones de los alumnos de un curso

Caso	ST1	ST2	GES	ST3	IOP	INF	MAT	ECO	ING
1	,3	,3	1,0	,0	1,7	,6	,6	,6	,3
...	...	...	...	...	...	...	...	...	...
15	6,2	6,5	4,1	4,2	2,4	3,4	5,5	6,2	5,8
...	...	...	...	...	...	...	...	...	...
29	10,0	10,0	9,6	8,7	9,6	9,3	10,0	7,2	7,5

variables: Estadística 1 (ST1), Estadística 2 (ST2), Estadística 3 (ST3), Investigacion Operativa (IOP), Informática (INF), Matemáticas (MAT), Economía (ECO), Gestión (GES) e Inglés (ING).

Allí, el Análisis de Componentes Principales se preguntaba por cuántas y cuáles serían unas nuevas variables (componentes principales) que nos permitieran resumir la diversidad de calificaciones observada en un espacio de menor dimensión y con la menor pérdida de información posible.

Intuitivamente, cuando hemos tratado de explicar el significado de las dos primeras componentes principales allí obtenidas hemos llegado a relacionarlas con la cierta capacidad o habilidad para sacar buenas notas en general (¿inteligencia? ¿preparación? ¿esfuerzo?) y con la capacidad o habilidad para sacar buenas notas en materias que requieren mayor habilidad de tipo lingüístico que de cálculo numérico.(¿inteligencia, capacidad o habilidad en lenguaje versus cálculo numérico?). Si esta interpretación se repitiese en estudios similares, cabría pensar que probablemente estas dos componentes podrían ser las causas principales (salvo otros efectos menores) de las calificaciones observadas. Este es el punto de partida del Análisis Factorial.

El Análisis Factorial presupone la existencia de un número pequeño de variables no observables o latentes (que llamaremos factores), que serían las causas reales (aunque no observadas) de las calificaciones obtenidas por los alumnos, y tiene por objeto su identificación.

Como técnica multivariante, podrá emplearse tanto para investigar cuantos y cuáles podrían ser tentativamente estos factores (enfoque exploratorio), como para tratar de confirmar alguna teoría que propugna o defiende la existencia de ciertos factores causantes de los efectos observados (enfoque confirmatorio).

Trataremos de encontrar pues un número de factores que informen de la realidad observada de la forma más sencilla y clara.

Para que los factores informen de la forma más sencilla posible, trataremos de encontrar el menor número de ellos que sea suficiente para reproducir la realidad observada lo más fiablemente posible. Pero, si cada una de las p variables observadas aportara información neta sustancial, independiente de las demás, entonces difícilmente podremos obviarla ya que estaríamos perdiendo información neta. La consideración de un menor número m de factores ($m<p$) sólo será posible si las variables contienen información redundante que puede ser eliminada, por lo que será preciso que las variables originales presenten altas dosis de correlación entre ellas.

Para que los factores informen de forma clara, trataremos además, en una primera aproximación, que cada uno facilite la máxima información posible y lo haga de forma desligada o independiente de los demás. Exigiremos por tanto, en esta primera aproximación, que los factores estén incorrelacionados. Así garantizaremos que cada uno aportará una información que no facilita ningún otro y por tanto, con el menor número de factores vamos a poder cubrir la máxima información que los datos facilitan.

El modelo general del Análisis Factorial

Nuestro objetivo pues es encontrar ese número reducido de variables latentes (en adelante, factores) que expliquen de la forma más simple y clara posible (número de factores reducido e incorrelacionados), a qué se debe la variedad de comportamientos observado en un conjunto de individuos sobre los que hemos observado p variables.

Así, procedamos a continuación a formular el modelo general del Análisis Factorial.

Formulación del modelo general del Análisis Factorial:

Sean $X_1,X_2,...,X_p$ las variables observadas que supondremos centradas (sus medias son 0) sin pérdida de generalidad.

(alternativamente puede suponerse que las variables observadas $X_1,X_2,...,X_p$ se encuentran tipificadas; es decir, sus varianzas son iguales a 1 y sus medias son 0)

Llamemos factores comunes $F_1,F_2,...,F_m$ a las variables latentes que pretendemos identificar y que son causas ocultas de los comportamientos observados. Debemos además considerar un cierto margen de error para recoger el efecto de la posible aleatoriedad o simplemente de la incompleta especificación del modelo. Por ello, consideraremos además unos residuos, comportamientos de cada variable original no explicados por los factores comunes, y que llamaremos factores específicos $e_1,e_2,...,e_p$.

El problema será identificar esos m Factores Comunes F_1, $F_2,...,F_m$ que expliquen la mayor parte del comportamiento de todas las variables observadas y los p factores específicos e_1, $e_2,....,e_p$ residuales para las variables originales, tales que cada variable original pueda ser expresada como una combinación lineal de esos m factores (por ello reciben el nombre de comunes) más una parte residual propia de cada variable (por ello reciben el nombre de factor específico) no explicada por los factores comunes y que se pretende sea pequeña con relación a lo explicado por éstos.

De esta forma, el modelo general puede expresarse como sigue:

$$X_1 = l_{11} \cdot F_1 + \cdots + l_{1m} F_m + e_1$$
$$\ldots\ldots \quad \ldots\ldots \quad \ldots\ldots \quad \ldots\ldots \quad \ldots$$
$$X_p = l_{p1} \cdot F_1 + \cdots + l_{pm} F_m + e_p$$

Este sistema lineal lo podemos escribir matricialmente como:

$$\begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix} = \begin{pmatrix} l_{11} & \cdots & l_{1m} \\ \vdots & \ddots & \vdots \\ l_{p1} & \cdots & l_{pm} \end{pmatrix} \begin{pmatrix} F_1 \\ \vdots \\ F_m \end{pmatrix} + \begin{pmatrix} e_1 \\ \vdots \\ e_p \end{pmatrix}$$

o simplificadamente,

$$x = Lf + e$$

donde estamos representando por x el vector columna de las variables originales; por L, la matriz de coeficientes del sistema lineal; por f, el vector columna de factores comunes; y por e, el vector columna de factores específicos:

A esta formulación tenemos que añadirle ciertas condiciones que son necesarias para que el modelo pueda ser resuelto convenientemente, bien por necesidades de espeificación (normalización en forma de tipificación de los factores comunes y medias nulas para los específicos), o bien por conveniencia del modelo (incorrelación entre todos los factores, sean comunes o específicos, para que todos aporten información neta y puedan reducirse al menor número posible). El modelo finalmente así formulado sería:

Formulación:

$$x = Lf + e$$

$$\text{Hipótesis}: \quad \begin{aligned} &\mathrm{E}[\mathrm{f}] = 0 \\ &\mathrm{E}[\mathrm{f} \cdot \mathrm{f}'] = I \\ &\mathrm{E}[\mathrm{e}] = 0 \\ &\mathrm{E}[\mathrm{e} \cdot \mathrm{e}'] = \Omega \ diagonal = ((w_i^2)) \\ &\mathrm{E}[\mathrm{f} \cdot \mathrm{e}'] = 0 \end{aligned} \qquad [1]$$

Las condiciones $E[f] = 0$, $E[e] = 0$ aseguran que todos los factores tendrán medias cero; $E[f \cdot f'] = I$ asegura que las varianzas de los factores comunes valdrán todas 1 y que además los factores comunes estarán incorrelacionados (tienen covarianzas nulas) entre sí; $E[f \cdot e'] = 0$ asegura la incorrelación entre factores comunes y específicos (tienen covarianzas nulas); y finalmente, la condición $E[e \cdot e'] = \Omega \text{ diagonal} = ((w_i^2))$ asegura que los factores específicos estarán igualmente incorrelacionados (tienen covarianzas nulas) entre sí, presentando cada uno una varianza w_i^2 que llamaremos a partir de ahora *especificidad*.

Consecuencia 1: Obsérvese que, como consecuencia inmediata de esta formulación, los coeficientes l_{ij} de la matriz L, son justamente las covarianzas entre la variable original X_i y el Factor Común F_j.

$$l_{ij} = Cov(X_i, F_j)$$

ya que la covarianza de una combinación lineal de variables, como es el caso de $X_i = l_{i1} \cdot F_1 + l_{i2} \cdot F_2 + \cdots + l_{im} \cdot F_m + e_i$, y otra variable, como es F_j, será la combinación lineal de las covarianzas correspondientes; por lo que:

$$\begin{aligned} Cov(X_i, F_j) &= Cov(l_{i1} \cdot F_1 + \cdots + l_{im} \cdot F_m + e_i \ , \ F_j) = \\ &= l_{i1} \cdot Cov(F_1, F_j) + \cdots + l_{im} \cdot Cov(F_m, F_j) + Cov(e_i, F_j) = l_{ij} Cov(F_j, F_j) = l_{ij} \end{aligned}$$

al ser, por las hipótesis del modelo,

$$Cov(F_k, F_j) = \begin{cases} 0 & si\ k \neq j \\ 1 & si\ k = j \end{cases} \qquad \text{y} \qquad Cov(e_i, F_j) = 0$$

Consecuencia 2: Si las variables originales estuvieran tipificadas, los coeficientes l_{ij} de la matriz L son justamente los coeficientes de correlación entre la variable original X_i y el Factor Común F_j.

$$\text{Si } X_i \text{ está tipificada } \Rightarrow l_{ij} = corr(X_i, F_j)$$

ya que la correlación entre dos variables es su covarianza dividida por la raíz cuadrada del producto de sus varianzas, y ambas varianzas de X_i y de F_j serían 1, al estar tipificadas.

A los coeficientes l_{ij} de la matriz L se les conoce como *saturaciones factoriales, cargas factoriales* o *pesos factoriales*.

Por tanto, la covarianza de cualquier variable inicial, X_i, (correlación en el caso de estar tipificada) con cualquier factor común F_j, es siempre el coeficiente que acompaña al factor común F_j en la explicación de la variable X_i en el modelo factorial. Por ello, dado el carácter informativo de la covarianza y el coeficiente de correlación sobre la dependencia, estos coeficientes l_{ij} nos serán tremendamente útiles para tratar de explicar el significado de los factores F_j. Así, valores absolutos grandes de l_{ij} indican una dependencia alta entre la variable X_i y Factor común F_j; y recíprocamente, valores absolutos pequeños de l_{ij} implican una dependencia baja entre la variable X_i y Factor común F_j.

Resolución del Modelo:

Las incógnitas indeterminadas del modelo [1] a resolver son las cargas factoriales, l_{ij}, y las especificidades w_i^2. Vamos, pues, a tratar de estimar L, cuyos coeficientes (cargas factoriales) nos informarán de qué y cuales son los Factores Comunes, y Ω, cuyos coeficientes (especificidades) nos informarán de cuánta dispersión no puede ser explicada por los factores comunes; y, por tanto, nos informarán de la bondad del modelo..

Para ello, observemos que, al ser *L* y *L'* matrices escalares, la matriz de varianzas y covarianzas de las variables observadas es

$$\begin{aligned}\Sigma = E[xx'] &= E[(Lf+e)\cdot(Lf+e)'] = E[(Lf+e)\cdot(f'\,L'+e')] = \\ &= E[L\cdot f\cdot f'\cdot L'] + E[L\cdot f\cdot e'] + E[e\cdot f'\cdot L'] + E[e\cdot e'] = \\ &= LE[f\cdot f']L' + LE[f\cdot e'] + E[e\cdot f']L' + E[e\cdot e']\end{aligned}$$

y teniendo en cuenta las hipótesis del modelo factorial [1], y que también $E[e\cdot f'] = 0$ (por ser $e\cdot f'$ la traspuesta de la matriz $f\cdot e'$ anterior),

$$\Sigma = L\cdot I\cdot L' + L\cdot 0 + 0\cdot L' + \Omega = L\cdot L' + \Omega$$

Observemos que si las variables observadas ($X_1, \ldots, X_p$) estuvieran tipificadas, entonces su matriz de varianzas y covarianzas Σ coincidiría con su matriz de correlaciones ρ, en cuyo caso el sistema anterior sería equivalente al siguiente:

$$\Sigma = L\cdot L' + \Omega \Leftrightarrow \rho = L\cdot L' + \Omega$$

y que alternativamente puede escribirse de la forma:

a) caso general

$$\begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_2^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \cdots & \cdots & \sigma_p^2 \end{pmatrix} = \begin{pmatrix} l_{11} & l_{12} & \cdots & l_{1m} \\ l_{21} & l_{22} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ l_{p1} & \cdots & \cdots & l_{pm} \end{pmatrix} \cdot \begin{pmatrix} l_{11} & l_{21} & \cdots & l_{p1} \\ l_{12} & l_{22} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ l_{1m} & \cdots & \cdots & l_{pm} \end{pmatrix} + \begin{pmatrix} w_1^2 & 0 & \cdots & 0 \\ 0 & w_2^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & w_p^2 \end{pmatrix}$$

b) caso de las variables observadas tipificadas

$$\begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{12} & 1 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1p} & \cdots & \cdots & 1 \end{pmatrix} = \begin{pmatrix} l_{11} & l_{12} & \cdots & l_{1m} \\ l_{21} & l_{22} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ l_{p1} & \cdots & \cdots & l_{pm} \end{pmatrix} \cdot \begin{pmatrix} l_{11} & l_{21} & \cdots & l_{p1} \\ l_{12} & l_{22} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ l_{1m} & \cdots & \cdots & l_{pm} \end{pmatrix} + \begin{pmatrix} w_1^2 & 0 & \cdots & 0 \\ 0 & w_2^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & w_p^2 \end{pmatrix}$$

En lo que sigue, nos referiremos al caso general y lo particularizaremos cuando sea conveniente para cuando las variables observadas estén tipificadas.

Nuestro problema es encontrar los l_{ij} y los w_i^2, para tener realmente especificado nuestro problema y poder expresar las relaciones que existen entre los factores y las variables iniciales.

Observemos que tanto las varianzas como las covarianzas de las variables observadas pueden ser calculadas (enfoque descriptivo) o estimadas (enfoque inferencial) a partir de los datos iniciales; por lo que tenemos planteado un sistema de ecuaciones, no lineal en las cargas factoriales y lineal en las especificidades, que debemos resolver.

Propiedad-1: No siempre hay solución; depende del número de variables y del número de factores comunes que queramos obtener.

Para analizar la solubilidad del sistema, veamos cuántas ecuaciones presenta y cuántas incógnitas involucra realmente.

Con relación al número de ecuaciones, inicialmente tenemos tantas como elementos tienen las matrices de ambos lados del sistema, Σ y LL'+Ω, y que deben ser igualados; es decir, inicialmente podemos plantear $p{\cdot}p$ ecuaciones. Sin embargo, observamos que tanto le lado izquierdo del sistema, Σ, como el lado derecho, LL'+Ω son matrices simétricas, por lo que las ecuaciones que corresponden igualar los coeficientes de la semi-matriz triangular superior se repiten al igualar los de la semi-matriz inferior. Luego realmente sólo podremos plantear tantas ecuaciones como tengamos en uno de los triángulos, inferior o superior, de la matriz incluyendo la diagonal principal; por lo que tendremos un número de ecuaciones igual a:

$$1+2+3+\cdots+p=\frac{p(p+1)}{2}$$

Con respecto a las incógnitas, éstas son las $m{\cdot}p$ cargas factoriales $l_{11},\ldots,l_{1m},\ldots,l_{p1},\ldots,l_{pm}$ y las p especificidades $w_1,\ldots,w_p$. Es decir, en definitiva hay $p{\cdot}m+p$ variables:

$$p{\cdot}m+p=p(m+1)$$

Así, el problema del análisis factorial que hemos planteado tendrá solución si y sólo si el número de ecuaciones, por lo menos, es igual al número de incógnitas:

$$\frac{p(p+1)}{2}\geq p(m+1)\Leftrightarrow\frac{p+1}{2}\geq m+1\Leftrightarrow m\leq\frac{p-1}{2}$$

lo que limita el número máximo de factores comunes que podemos obtener para garantizar el cumplimiento de las condiciones (básicamente incorrelación entre todos los factores) establecidas en el modelo general. Por ejemplo, si observamos p=10 variables, solamente podemos plantear modelos con m=4 factores comunes a lo sumo y p=10 factores específicos, todos ellos incorrelacionados entre sí.

Propiedad-2: Si hay solución, no es única. Si tenemos una solución del problema, (L,Ω), entonces, dada una transformación ortogonal cualquiera H de la dimensión adecuada, entonces (LH,Ω) también es solución.

Recordemos que una transformación ortogonal es aquélla en la que su inversa coincide con su traspuesta, $H^{-1} = H'$. Y para ver que (LH,Ω) es solución, debemos comprobar que se verifica que:

$$\Sigma = (LH)\cdot(LH') + \Omega$$

y sin más que operar, vemos que esto será cierto si y solo si

$$\Sigma = (LH)\cdot(LH') + \Omega \Leftrightarrow \Sigma = LHH'L' + \Omega = LIL' + \Omega = LL' + \Omega \Leftrightarrow \Sigma = LL' + \Omega$$

lo que efectivamente se verifica, al ser (L,Ω) solución.

Propiedad-3: $Var(X_i) = \sigma_i^2 = l_{i1}^2 + l_{i2}^2 + \cdots + l_{im}^2 + w_i^2 = h_i^2 + w_i^2$

Se deduce de la igualdad fundamental $\Sigma = LL' + \Omega$, fijándonos en las ecuaciones provenientes de comparar los elementos de la diagonal, y teniendo en cuenta que Σ es la matriz de varianzas y covarianzas de las variables observadas.

En el producto *LL'*, el elemento (*i*, *i*) de la diagonal principal sería la fila *i*-ésima de *L* por la columna *i*-ésima de *L'* , siendo ésta la misma fila *i*-ésima de *L*. A este producto, parte de la varianza σ_i^2 que depende exclusivamente de las pesos de los factores comunes que intervienen en la expresión de la variable original *Xi* recibe el nombre de *comunalidad* y explícitamente es la suma de los cuadrados de los coeficientes de la fila *i* de la matriz de cargas factoriales.

$$h_i^2 = l_{i1}^2 + l_{i2}^2 + \cdots + l_{im}^2 = \sum_{j=1}^{m} l_{ij}^2 = comunalidad$$

Así pues, una vez resuelto el modelo, la varianza de cada variable observada X_i, σ_i^2 , puede descomponerse como la suma de una *comunalidad* h_i^2 (o parte de la varianza debida a los factores comunes) más una *especificidad* w_i^2 (o parte de la varianza que no pueden explicar los factores comunes y que recoge el correspondiente factor específico e_i), como asevera el enunciado de esta propiedad.

Obsérvese que cuando las variables originales vengan tipificadas, la propiedad se reduce a

$$Var(X_i) = \rho_{ii} = 1 = l_{i1}^2 + l_{i2}^2 + \cdots + l_{im}^2 + w_i^2 = h_i^2 + w_i^2$$

tomando entonces la comunalidad h_i^2 el significado de la proporción de la varianza de *Xi*, σ_i^2 , que explican los factores comunes, y la especificidad w_i^2 el significado de la proporción de la varianza de *Xi*, σ_i^2 , que no pueden explicar los factores comunes.

Propiedad-4: $Cov(X_i, X_j) = \sigma_{ij} = l_{i1}l_{j1} + \cdots + l_{im}l_{jm}$

Fijándonos en la misma igualdad básica, $\Sigma = LL' + \Omega$, pero en las ecuaciones procedentes de igualar elementos extradiagonales (i, j) con $i \neq$j, entonces el la covarianza σ_{ij} será igual al producto de la fila i-ésima de L por la columna j-ésima de L' (que es la fila j-ésima de L) más el elemento (i, j) de Ω que vale cero, pues Ω es diagonal. En definitiva, el producto escalar de la fila i-ésima por la fila j-ésima de la matriz de cargas factoriales L.

Propiedad-5: La varianza explicada por el Factor j-ésimo es:

$$Var(F_j) = \lambda_j = l_{1j}^2 + l_{2j}^2 + \cdots + l_{pj}^2 = \sum_{i=1}^{p} l_{ij}^2$$

Técnicas de Resolución

Hasta ahora hemos visto que el Análisis Factorial tiene solución bajo las condiciones de dimensionalidad que hemos especificado, y que la solución no va a ser única, porque podemos obtener otras mediante transformaciones ortogonales. Procedamos ahora a abordar el problema obtener una solución concreta.

Todos tratan de resolver el sistema de ecuaciones para estimar las cargas factoriales (L) y las especificidades (Ω).

$$\Sigma = L \cdot L' + \Omega \quad \text{o bien}: \ \rho = L \cdot L' + \Omega$$

Diferentes métodos de obtener las estimaciones de las matrices de covarianzas o de correlación a partir de los datos observados y diferentes algoritmos de cálculo iterativos particularizan los diversos métodos.

Entre los métodos clásicos empleados por excelencia para la estimación de los parámetros de L y Ω a partir de la matriz de varianzas y covarianzas de los datos, podemos citar

–Máxima verosimilitud

Si la Población es $N_p(\mu,\Sigma)$, entonces la función soporte de verosimilitud (excluida constante) es:

$$\ln L(\mu, \Sigma \mid x) = -\frac{n}{2}\ln|\Sigma| - \frac{n}{2} traza(\Sigma^{-1}S) - \frac{n}{2}(\bar{x}-\mu)'\Sigma^{-1}(\bar{x}-\mu)$$

y maximizarla es equivalente, para $\hat{\mu} = \bar{x}$ (estimador de m.v. para la media), a maximizar

$$\ln L(\mu, \Sigma \mid x) = -\frac{n}{2}\ln|\Sigma| - \frac{n}{2} traza(\Sigma^{-1}S) = -\frac{n}{2}\ln|LL'+\Omega| - \frac{n}{2} traza((LL'+\Omega)^{-1}S)$$

lo que se resuelve iterativamente mediante un algoritmo tipo Newton Rapsom, por ejemplo.

Si existieran estimadores eficientes, los estimadores de máxima verosimilitud los proporcionan. Además, los estimadores de máxima verosimilitud son asintóticamente óptimos y Normales (OAN), y proporcionan estimaciones equivalentes ante transformaciones lineales de las variables. Así, con la estimación obtenida a partir de la matriz de correlaciones, se puede obtener equivalentemente la estimación correspondiente para cuando se parte de la matriz de covarianzas.

–Mínimos cuadrados generalizados

Maximizar el soporte de la función de verosimilitud es equivalente a minimizar

$$(-\frac{n}{2}\ln|S| - \frac{n}{2}p) - \ln L(\mu, \Sigma \mid x) = \frac{n}{2}\left(traza(\Sigma^{-1}S) - p - \ln\left|\Sigma^{-1}S\right|\right)$$

siendo la expresión $(-\frac{n}{2}\ln|S| - \frac{n}{2}p)$ el valor que tomaría el soporte de la función de verosimilitud en el óptimo para la solución de máxima verosimilitud sin restricciones, $\hat{\Sigma} = S$, que infraestima Σ ya que sabemos que el estimador insesgado de Σ divide por n-1, en lugar de dividir por n como hace S.

Con la restricción de ser $\left|\Sigma^{-1}S\right| \leq 1$, sería equivalente a minimizar

$$traza(\Sigma^{-1}S) - p = traza(\Sigma^{-1}S - I) = traza(\Sigma^{-1}(S - \Sigma))$$

ya que el logaritmo sería negativo o nulo. Así, en base a esto, el método de mínimos cuadrados generalizados minimiza la expresión

$$traza\left[\Sigma^{-1}(S - \Sigma)\right]^2$$

que evalúa las discrepancias entre S y Σ, ponderándolas por un peso dependiente de Σ^{-1}; lo que conduce a estimadores asintóticamente eficientes.

–Factorización de Ejes principales (o método del factor principal)

Éste y otros métodos suele recurrirse a estimar a la llamada *Matriz Reducida de varianzas y covarianzas*,

$$\Sigma^* = \Sigma - \Omega \quad \textit{o equivalentemente} \quad \rho^* = \rho - \Omega$$

cuya diagonal contiene solamente la comunalidad o parte de varianza (o proporción de varianza) que explican los factores comunes; y en este caso, el sistema a resolver se reduce a:

$$\Sigma^* = L \cdot L' \quad \textit{o equivalentemente} \quad \rho^* = L \cdot L'$$

Como son análogos, para simplificar la exposición de los procedimientos de estimación de estas matrices Σ^* o ρ^*, centrémonos en el caso particular de la matriz reducida de correlaciones ρ^*. Ésta es la misma de correlaciones salvo en su diagonal, donde se resta a los 1 las especificidades correspondientes; es decir, en su diagonal aparecen las correspondientes comunalidades de las variables observadas.

La comunalidad de una variable X_j cualquiera sería el coeficiente de determinación de ella con respecto a los factores comunes considerados como variables explicativas (proporción de su varianza que explican los factores comunes. Y como se espera que los factores comunes contengan prácticamente toda la información sobre las variables originales y que a su vez todas las variables observadas estén bastante correlacionadas entre sí, se suele estimar inicialmente dicha comunalidad a partir del coeficiente de determinación múltiple entre la correspondiente variable observada y el resto de ellas.

$$\hat{h}_i^2 = R^2_{X_i;F_1\ldots F_p} \approx R^2_{X_i;X_1\ldots X_p}$$

Análogamente se estimará inicialmente la especificidad por su diferencia a 1.

$$\hat{e}_i^2 = 1 - R^2_{X_i;F_1\ldots F_p} \approx 1 - R^2_{X_i;X_1\ldots X_p}$$

Con estas estimaciones se obtendrían unas primeras soluciones para L a partir de las cuales se obtendría una primera aproximación de la varianza explicada por cada factor j

$$\lambda_j^2 = l_{1j}^2 + l_{2j}^2 + \cdots + l_{mj}^2 = \sum_{i=1}^{p} l_{ij}^2$$

y una estimación de la comunalidad que esta solución para la variable X_i proporcionaría

$$h_i^2 = l_{i1}^2 + l_{i2}^2 + \cdots + l_{im}^2 = \sum_{j=1}^{p} l_{ij}^2$$

para reiterar el procedimienteo hasta conseguir la convergencia en términos de un error tolerable.

Los estimadores del método del factor principal son consistentes, pero no eficientes en general. Además, tampoco son invariantes ante transformaciones lineales de las variables como lo era la estimación de Máxima verosimilitud.

Sin embargo, el procedimiento de resolución más comúnmente empleado, al menos desde una perspectiva exploratoria, se basa en el cálculo y transformación convenientes de las componentes principales, como vamos a ver a continuación.

Relación con el Análisis de Componentes Principales:

Según vimos en el capítulo anterior, las Componentes Principales estaban incorrelacionadas entre sí y podían expresarse en función de las variables observadas como:

$$Z_i = X_1 u_{1i} + \cdots + X_p u_{pi} \;, \qquad i=1,2,...,p$$

o, matricialmente,

$$Z = X \cdot U$$

siendo U la matriz cuyas columnas eran los autovectores ortonormalizados de la matriz de varianzas y covarianzas de las variables originales y Z la matriz de puntuaciones o valores de las componentes calculadas para todos los casos.

Sabemos también que la matriz U de los autovectores ortonormalizados es una matriz ortogonal y, por tanto, su inversa coincide con su transpuesta, $U^{-1} = U'$; por lo que podemos despejar la matriz X de la siguiente forma:

$$Z = X \cdot U \Leftrightarrow X = Z \cdot U^{-1} = Z \cdot U' \Leftrightarrow X = Z \cdot U'$$

siendo cada columna de la matriz X la correspondiente variable X_i, cada columna de la matriz Z la correspondiente variable Z_i, y las filas de la matriz U' los autovectores que U tenía en sus columnas, y que habíamos notado por $u_{11}, u_{21}, \cdots, u_{i1}, \cdots u_{p1}$ para la primera columna y análogamente para las demás; lo que más explícitamente podemos poner como:

$$\begin{pmatrix} X_1 & \cdots & X_i & \cdots & X_p \end{pmatrix} = \begin{pmatrix} Z_1 & \cdots & Z_i & \cdots & Z_p \end{pmatrix} \begin{pmatrix} u_{11} & u_{21} & \cdots & u_{i1} & \cdots & u_{p1} \\ \vdots & \vdots & & \vdots & & \vdots \\ u_{1p} & u_{2p} & \cdots & u_{ip} & \cdots & u_{pp} \end{pmatrix}$$

Por tanto, podemos expresar la variable X_i como una combinación lineal de las Componentes Principales Z_i sin más que realizar el producto correspondiente por la columna i-ésima de U', $u_{i1}, \ldots\ldots, u_{ip}$; lo que conduce a:

$$X_i = Z_i \cdot u_{i1} + \cdots + Z_p \cdot u_{ip} \,, \qquad i=1,....,p$$

Obsérvese que las Componentes Principales cumplen rigurosamente con la principal propiedad exigida a los factores comunes en el Análisis Factorial que no es otra que la de estar incorrelacionados entre sí.

Sin embargo, también hemos exigido a éstos que estén tipificados, lo que en principio no cumplían las componentes principales que obtuvimos, ya que sabemos que cada componente principal, presentando media nula, Z_h explicaba una varianza igual a su correspondiente autovalor asociado λ_h. Por ello, para ver si las componentes principales pudieran considerarse soluciones del problema del Análisis Factorial, necesitamos tipificarlas dividiéndolas por la raíz cuadrada del autovalor (su desviación típica). Así, las componentes principales tipificadas pueden expresarse como:

$$Y_h = \frac{Z_h}{\sqrt{\lambda_h}}$$

en cuyo caso, la anterior expresión de las variables observadas en función de las componente, ahora tipificadas y ortogonales entre sí, sin más que sustituir las Z_h por su expresión en función de las Y_h quedaría como:

$$X_i = u_{i1}\sqrt{\lambda_1}\cdot Y_1 + \cdots + u_{im}\sqrt{\lambda_m}\cdot Y_m + \cdots + u_{ip}\sqrt{\lambda_p}\cdot Y_p \quad , \qquad i=1,\ldots,p$$

De esta expresión podemos deducir la solución que el Análisis de Componentes Principales aporta a la búsqueda de m factores mediante el Análisis Factorial y que consiste justamente tomar como factores comunes F_h las m primeras componentes principales tipificadas Y_h y como factores específicos e_h los restos que aparecen en cada una de las expresiones de X_1 , X_2 ,.....,X_p , combinaciones lineales de las últimas p-m componentes principales, que también son ortogonales con las anteriores puesto que todas las Componentes Principales eran ortogonales entre sí.

$$\left.\begin{array}{l} F_h = Y_h \\ e_h = u_{h,m+1}\sqrt{\lambda_{m+1}}Y_{m+1} + \cdots + u_{hp}\sqrt{\lambda_p}\cdot Y_p \end{array}\right\} \ , h = 1,2,\ldots,m$$

Observemos que los coeficientes $u_{ih}\sqrt{\lambda_h}$ para $h=1,2,\ldots,m$ serían, según el enfoque factorial, las cargas factoriales, pesos factoriales o saturaciones factoriales de cada factor común y deben representar la correlación existente entre la variable observada X_i y el factor F_h , lo que puede verse coincide con las propiedades estudiadas en el análisis de Componentes principales.

Un problema que presenta este método es que incumple la ortogonalidad de los factores específicos entre sí, ya que todos ellos (los p factores específicos) dependen linealmente de las últimas p-m componentes principales y , obviamente, p-m<p. En cualquier caso, desde un punto de vista práctico, esto no se considera en general un gran inconveniente ya que si el modelo es medianamente bueno cabe esperar que los factores específicos se comporten como residuales y generalmente despreciables.

Puntuaciones Factoriales

Se llaman *puntuaciones factoriales* a las coordenadas de los casos en el espacio de los factores.

En el caso del Análisis de Componentes Principales, la obtención de la matriz de puntuaciones factoriales F es especialmente sencilla, ya que allí se verificaba que:

$$F_h = \frac{Z_h}{\sqrt{\lambda_h}} \quad y \quad Z = X\cdot U \Leftrightarrow X = Z\cdot U^{-1} = Z\cdot U' \Leftrightarrow X = Z\cdot U'$$

por lo que, para obtener las puntuaciones factoriales $F=Z\cdot\Lambda^{-1/2}$, bastará con obtener primero Z, para lo que basta con multiplicar las matrices de datos observados X y de autovectores ortonormalizados U.

En el modelo general $x = Lf + e$, será algo más complicado. Aquí podemos expresar la relación entre coordenadas de los casos en los espacios de variables observadas y de los factores como $X \approx L \cdot F$, dado el esperado carácter residual de los factores específicos, siendo X la matriz de datos observada y F la matriz de puntuaciones factoriales correspondiente. Sin embargo, "despejar" F no resulta tan evidente, ya que la matriz de cargas factoriales L no es invertible (ni siquiera tiene que ser una matriz cuadrada).

Por tanto, para pasar de $X=L \cdot F$ a $F=X \cdot L^*$, se necesita estimar, por algún método, la matriz de paso L* que nos permita poder calcular la matriz de puntuaciones factoriales F.

Los métodos más comúnmente empleados para ello son:

- Método de Regresión,
  aunque su solución no garantiza el cumplimiento de hipótesis del modelo general
- Anderson-Rubin,
  que garantiza incorrelación y varianzas 1 de los vectores de puntuaciones
- Barlett,
  Especialmente desarrollado para estimaciones de Máxima Verosimilitud bajo el supuesto de que los factores se distribuyan normalmente.

Rotaciones

Hemos visto al resolver el problema del Análisis Factorial que, por un lado, la solución factorial, cuando existe, no es única; y por otro, que, estamos buscando simplificar la dimensión del problema a un número menor de dimensiones que expresen claramente el comportamiento observado. Aunque también sabemos que la simplificación a una menor dimensionalidad, tanto en el Análisis de Componentes Principales, como en el Análisis Factorial, la realizábamos a través de un número reducido de nuevas variables (componentes o factores) combinaciones lineales de las variables originales, y que en principio no tenían por qué ser fáciles de interpretar.

Así por ejemplo, ¿qué significaría que un factor fuera 0.5*Nota en Matemáticas+0.3*Nota en Investigación Operativa+ ···, en nuestro ejemplo de las 9 asignaturas?.

Lo que se pretende con las rotaciones, aprovechando la propiedad de que la solución no es única y que cualquier transformación ortogonal (las rotaciones lo son) de una solución también lo es, es justamente obtener, de entre todas las posibles soluciones que podríamos encontrar aplicando transformaciones ortogonales, aquélla que nos haga lo más fácilmente interpretable los factores resultantes.

Y para poder interpretar los factores, tendremos que recurrir a la relación existente entre éstos y las variables cuyo contenido conceptual conocemos. Pero esta relación, tal como se había planteado el modelo, viene dada por los l_{ij} —cargas, saturaciones o pesos factoriales— que expresaban la correlación entre la variable observada X_i y el Factor común F_j.. Por ello, lo que se pretende será encontrar rotaciones (que son transformaciones ortogonales) que produzcan una nueva solución del Análisis Factorial de manera que los nuevos factores comunes estén lo más altamente correlacionados posible con unas variables y lo menos correlacionados posible con otras, para así poder interpretarlos mejor en función de su más estrecha relación con unas variables y su baja o nula dependencia de otras.

Cuanto más claramente dependa un factor de unas variables y más claramente no dependa de otras variables, más claramente interpretable será éste. Por ello, pretendemos que la rotación consiga que la mayoría de los l_{ij} estén cercanos a cero y sólo haya otros cuantos cercanos a +1 ó -1; pero que no haya demasiados valores intermedios que nos complicarían mucho más la interpretación:

Rotaciones Ortogonales

Con esta lógica por guía, las llamadas *Rotaciones Ortogonales* se obtienen estrictamente empleando una transformación ortogonal T, siendo $T^{-1}=T'$. Así se obtiene una nueva solución del problema general del Análisis Factorial planteado que conserva la incorrelación entre los factores, indicando las nuevas cargas factoriales el grado de correlación lineal existente entre variables y factores, así como las comunalidades de las variables observadas (los nuevos factores rotados consiguen explicar exactamente la misma proporción de diversidad de las variables observadas, que los factores originales); lo que asegura que mantienen conjuntamente la máxima capacidad informativa y que lo hacen de forma prácticamente independientemente de las demás, aportando cada factor información perfectamente diferenciada.

En consecuencia, tras una rotación ortogonal, se conserva la incorrelación de los factores y las comunalidades de las variables observadas, siendo la *matriz de componentes* (matriz de las cargas factoriales) coincidente siempre con la *matriz de estructura* (matriz de correlaciones entre variables observadas y factores).

El método de rotación ortogonal más utilizado es el de la R*otación Varimax* y su versión normalizada (*Rotación Varimax Normalizada*).

La idea fundamental del método de *Rotación Varimax* es maximizar la varianza de los cuadrados de las correlaciones de cada factor con las variables; lo que conduciría consecuentemente a polarizar los cuadrados de las cargas factoriales hacia 0 ó 1 (dispersarlos lo más posible dentro de su campo de variación), y por tanto hacer que las correlaciones entre variables observadas y factores (cargas factoriales) sean cercana a cero, a +1 ó a –1.

Para un factor F_j, esta varianza será:

$$Var(l_{*j}^2) = \frac{\sum_{i=1}^{p}\left(l_{ij}^2\right)^2}{p} - \left(\frac{\sum_{i=1}^{p} l_{ij}^2}{p}\right)^2$$

Las cantidades l_{ij} representaban, la correlación de la variable X_i con el factor F_j , correlación que puede ser positiva o negativa; por ello su cuadrado, l_{ij}^2 , es una medición cuantitativa del grado de dependencia entre la variable X_i y el factor F_j . Realmente los cuadrados, l_{ij}^2, son los coeficientes de determinación entre las variables observadas y el factor F_j y están siempre entre 0 y 1. Serán iguales a 0 cuando la variables esté incorrelacionada con el factor y valdrán 1 cuando tenga correlación máxima con el factor. En otro caso, tomarán valores intermedios dependiendo del grado de dependencia lineal.

La suma de las varianzas así obtenidas para lo m factores será una función de evaluación global de la relación existente entre variables y factores, y el método de *Rotación Varimax* calcula la transformación ortogonal (rotación) que produce que la expresión de esta evaluación global sea máxima, separando, por tanto, lo más posible las correlaciones entre variables observadas y factores.

$$Max\sum_{j=1}^{m}\left\{\frac{\sum_{h=1}^{p}\left(l_{hj}^{2}\right)^{2}}{p}-\left(\frac{\sum_{h=1}^{p}l_{hj}^{2}}{p}\right)^{2}\right\}$$

Como los factores rotados ortogonalmente permanecen incorrelacionados, la solución obtenida presentará factores altamente correlacionados con ciertas variables observadas, pero de forma separada, de forma que los sucesivos factores presentarán menor correlación con las variables más altamente correlacionados con los factores anteriores y viceversa. De hecho, éste ~~método minimizan el número de variables explicativas de los factores~~.

La diferencia con el método de *Rotación Varimax Normalizado* es que la función objetivo maximizada en este caso es la varianza de los cuadrados de las cargas factoriales normalizados por las respectivas comunalidades de las variables.

$$Max\sum_{j=1}^{m}\left\{\frac{\sum_{i=1}^{p}\left(\frac{l_{ij}^{2}}{h_{i}^{2}}\right)^{2}}{p}-\left(\frac{\sum_{i=1}^{p}\frac{l_{ij}^{2}}{h_{i}^{2}}}{p}\right)^{2}\right\}$$

Existen otros muchos métodos de Rotaciones Ortogonales que, en general, intentan de alguna manera polarizar las cargas factoriales hacia cero o hacia más o menos uno en los distintos ejes factoriales, análogamente a como lo procuran los métodos Varimax.

Así por ejemplo, el método de *Rotación Quartimax,* en lugar de maximizar la suma de las varianzas de los cuadrados de las cargas factoriales, trata de maximizar la suma de sus momentos de segundo orden respecto del origen; o lo que es lo mismo, la suma de los cuadrados de todas las cargas factoriales:

$$Max\sum_{j=1}^{m}\sum_{i=1}^{p}\left(l_{ij}^{2}\right)^{2}$$

Por su parte, este método hace mínimo el número de factores explicativos de las variables.

Otro método de rotación ortogonal es el método de *Rotación Equamax,* que combina los propiedades de los dos anteriores.

Rotaciones Oblicuas

Existen ocasiones en las que el estado del conocimiento de las cosas (teorías) nos induce a tratar de explicar la diversidad de comportamientos observados en función de factores no necesariamente incorrelacionados. Así, algunas teorías del mercado podrían tratar de explicar ciertos hechos observados en base a ciertas características sicológicas o sociológicas de los consumidores y agentes involucrados, observándose ciertas dependencias entre éstas.

En estas situaciones, no sería correcto imponer la condición de incorrelación entre los factores comunes, cuando sabemos que aquellos a los que queremos restringir nuestra investigación estarían correlacionados entre sí de acuerdo con ciertas teorías científicamente asentadas. Por ello es preciso, en estas ocasiones, transformar las soluciones iniciales de nuestro modelo (incorrelacionadas) para permitir que las soluciones finales puedan admitir cierto grado de correlación.

Este tipo de transformaciones que producen unos factores que pueden estar correlacionados entre sí reciben el nombre de Rotaciones Oblícuas, por ser no necesariamente ortogonales.

Estas rotaciones oblicuas obtienen factores finales (rotados) más correlacionados con las variables originales (observadas) que las rotaciones ortogonales antes vistas; lo cual permite hacer una interpretación a priori mucho más clara de los factores; pero no debemos olvidar que si la interpretación de estos factores correlacionadas puede ser más clara desde un punto de vista intuitivo, la interpretación del comportamiento de cada variable en función de éstos se oscurece debido a la dependencia existente entre estos, por lo que una dependencia aparentemente exclusiva de una variable con respecto de un factor(carga factorial no nula para él y nula para los demás factores), no puede excluir la dependencia real también de los otros factores a través de la correlación existente entre ellos.

Así pues, las *Rotaciones Oblicuas* se obtienen empleando transformaciones no necesariamente ortogonales, mediante una matriz no singular H que puede interpretarse como un giro oblicuo, y consiguen obtener factores altamente correlacionados con las variables originales, aunque también correlacionados entre sí; no respetando por tanto estrictamente las hipótesis del modelo problema general clásico del Análisis Factorial tal y como se ha planteado. Si los factores ortogonales f se transforman para dar unos factores correlacionados f^*, mediante la expresión $f^*=Hf$, entonces, la matriz de varianzas y covarianzas de los nuevos factores será HH'.

Pretendiendo una mayor interpretación intuitiva a los factores, entre las propiedades de los factores obtenidos mediante rotaciones oblicuas, cabe destacar que conservan las comunalidades de las variables observadas (siguen explicando conjuntamente la misma proporción de sus diversidades) pero NO la incorrelacción (ortogonalidad) de los factores (los factores resultantes pueden estar correlacionados).

En consecuencia, tras una rotación oblicua, no se conserva necesariamente la ortogonalidad de los factores , aunque sí las comunalidades de las variables observadas, no teniendo que coincidir la *matriz de componentes* (matriz de las cargas factoriales) con la *matriz de estructura* (matriz de correlaciones entre variables observadas y factores).

En cuanto a los métodos que proporcionan rotaciones oblícuas, a continuación se relacionan lo más conocidos, siendo el método Oblimín el más comúnmente empleado para rotar oblicuamente la solución inicial del Análisis Factorial: Rotación Oblimin y Oblimin Directa, Rotación Quartimin, Rotación Covarimin, Rotación Orthoblique, Rotación Oblimax, Rotación Promax, Rotación Maxplan

Adecuación del Análisis Factorial y Validación de Hipótesis

En este apartado se presentan algunos estadísticos y contrastes que nos permiten ver la adecuación o no de abordar el problema de Análisis Factorial ante una situación concreta. Recuérdese que la reducción efectiva de la dimensionalidad del problema bajo el enfoque del Análisis Factorial era conveniente que existiese información redundante entre las variables observadas, o lo que es lo mismo, que hubiese correlación entre las variables.

Lo que comprueban justamente las pruebas presentadas aquí, de una u otra forma, es si la matriz de varianzas y covarianzas es o no aproximadamente diagonal o, equivalentemente, si la correspondiente matriz de correlaciones es parecida a la identidad. Si esto fuera así, no habría correlaciones importantes entre las variables observadas y, por tanto, no tendría sentido el Análisis Factorial; Sí tendría sentido en la situación opuesta; es decir, si existen variables que están altamente correlacionadas.

Contraste de Esfericidad de Barlett

Fundamentalmente, el contraste de esfericidad de Bartlett comprueba la existencia de correlación entre todas las variables observadas de una forma global mediante el descarte de la incorrelación de las variables, viendo si el determinante de la matriz de correlaciones poblacional vale o no vale 1.

$$\begin{cases} H_0 : \left|R_{pobl}\right| = 1 \\ H_1 : \left|R_{pobl}\right| \neq 1 \end{cases}$$

Obsérvese que:

$$R_{pobl} = I \Rightarrow \left|R_{pobl}\right| = 1$$

por lo que:

$$\left|R_{pobl}\right| \neq 1 \Rightarrow R_{pobl} \neq I$$

lo que conduciría a la necesidad de contemplar la existencia de correlaciones no nulas entre algunas de las variables observadas.

Por tanto, en principio y a falta de ver si las correlaciones son relativamente importantes, tendría sentido considerar el Análisis Factorial con el objetivo de reducir la dimensionalidad del problema, cuando se se rechace H_0, para lo que se utiliza el estadístico experimental:

$$B_{\exp} = -\left[n - 1 - \frac{2p+5}{6}\right]\ln|R| \xrightarrow{BajoH_0} \chi^2_{0,5(p^2-p)}$$

Contraste de Adecuación KMO (Kaiser-Mayer-Olkin)

El contraste de Kaiser-Mayer-Olkin comprueba la existencia global de correlaciones entre las variables observadas a partir la comparación de las diferentes correlaciones simples y parciales entre cada dos variables X_i y X_j.

Si llamamos r_{ij} al coeficiente de correlación simple entre las variables X_i y X_j, entonces su cuadrado r^2_{ij} tiene el significado de la proporción de varianza de la variable X_i que explica la X_j. Sin embargo, parte de ella también podría ser explicada por alguna o varias de las demás variables, dada la posible existencia de correlación entre ellas.

Sin embargo, si llamamos a_{ij} a la correlación parcial entre las variables X_i y X_j, fijadas el resto de variables observadas, su cuadrado a^2_{ij} representa la proporción de varianza de la variable X_i que explica la X_j y que no puede ser explicada por el resto de variables fijadas.

Por tanto, si existiese un entramado amplio de intercorrelaciones entre las variables observadas (lo que proporcionaría informaciones redundantes múltiples y la posibilidad de reducir la dimensionalidad del problema eliminándola), al comparar dos variables observadas X_i y X_j ocurrirá que r^2_{ij} será relativamente bastante mayor que a^2_{ij}, siendo $a^2_{ij}=0$ cuando exista una dependencia perfecta de X_i o X_j con el resto de variables, mientras que r^2_{ij} y a^2_{ij} tenderán a ser iguales cuando estas variables estén incorrelacionadas con el resto.

Kaiser, Mayer y Olkin introducen el siguiente estadístico que nos informa globalmente de la existencia de intercorrelaciones y, en consecuencia, de si es o no adecuado realizar un Análisis Factorial en base a la redundancia de información.

$$KMO = \frac{\sum_i \sum_{j \neq i} r^2_{ij}}{\sum_i \sum_{j \neq i} r^2_{ij} + \sum_i \sum_{j \neq i} a^2_{ij}}$$

Desde esta óptica, la aplicación del Análisis Factorial será adecuada tiene este estadístico KMO es relativamente grande. En concreto, sus autores proponen la siguiente quía de aplicación, que se conoce como Escala de Kaiser (1974):

0.90 - Maravilloso	0.80 – Meritorio	0.70 – Medio
0.60 – Mediocre	0.50 - Miserable	<0.50 -Inaceptable

Contraste de Adecuación MSA (Measure of Sampling Adequacy)

De forma análoga al estadístico KMO, pero de forma individualizada para cada variable, evalúa la existencia de entramados de correlaciones de cada variable con el resto. Así, para una variable observada cualquiera X_i se propone el estadístico de prueba

$$MSA_i = \frac{\sum_{j \neq i} r_{ij}^2}{\sum_{j \neq i} r_{ij}^2 + \sum_{j \neq i} a_{ij}^2}$$

de forma que será adecuado considerar la variable X_i en el Análisis Factorial si MSA_i es relativamente grande.