

Análisis Factorial

Ignacio Aguilera Martos

7 de enero de 2019

Estadística Multivariante

Link: [Documentos LaTeX en GitHub](#)

1. Explicación teórica del modelo
2. Explicación del ejemplo en R

Explicación teórica del modelo

Idea de AF

El objetivo de este modelo es, dada una matriz de covarianzas o de correlación, ser capaz de explicar esta matriz a partir de factores no observados llamados factores comunes, de forma que se pueda explicar la matriz con un número menor de variables que en un punto inicial.

De esta forma matricialmente si tenemos n observaciones de dimensión p

tendríamos una matriz $F = \begin{pmatrix} F_{11} & \dots & F_{1n} \\ \dots & \dots & \dots \\ F_{k1} & \dots & F_{kn} \end{pmatrix}$ de factores y

$L = \begin{pmatrix} l_{11} & \dots & l_{1k} \\ \dots & \dots & \dots \\ l_{p1} & \dots & l_{pk} \end{pmatrix}$ una matriz de coeficientes de forma que

$x - \mu = LF + \epsilon$ donde ϵ es un vector de errores.

Tipos

- EFA (Exploratory Factor Analysis): se usa para identificar relaciones complejas entre conceptos o grupos de conceptos.

Tipos

- EFA (Exploratory Factor Analysis): se usa para identificar relaciones complejas entre conceptos o grupos de conceptos.
- CFA (Confirmatory Factor Analysis): está dirigida a la confirmación de factores que ya se presuponen importantes para explicar la matriz de correlación o covarianza.

Procesos de ajuste

- Máxima verosimilitud: es una buena opción cuando los datos se distribuyen según una normal. Se intenta que el modelo de los factores obtenidos tenga máxima verosimilitud.

Procesos de ajuste

- Máxima verosimilitud: es una buena opción cuando los datos se distribuyen según una normal. Se intenta que el modelo de los factores obtenidos tenga máxima verosimilitud.
- Factorización en el eje principal: la intención es ir obteniendo factores de forma que el primero tenga la varianza lo más próxima al objetivo, el segundo factor la segunda varianza más próxima a la varianza objetivo, etc. Maximiza la fórmula del modelo.

La rotación de factores se emplea para obtener la estructura de factores más simple escogiendo una orientación de los mismos.

La rotación de factores se emplea para obtener la estructura de factores más simple escogiendo una orientación de los mismos.

Rotación de factores

- Ortogonal: implica que los factores estén incorrelados y busca la estructura más simple.

La rotación de factores se emplea para obtener la estructura de factores más simple escogiendo una orientación de los mismos.

Rotación de factores

- Ortogonal: implica que los factores estén incorrelados y busca la estructura más simple.
- Oblicua: permite que los factores estén correlados y busca la estructura más simple.

No sólo tenemos que estudiar la generación de los factores, si no también cuántos factores debemos escoger.

No sólo tenemos que estudiar la generación de los factores, si no también cuántos factores debemos escoger.

Métodos para escoger el número de factores

- Regla de Kaiser: tomamos los valores propios de la matriz de entrada y comprobamos cuántos de ellos son mayores que 1. Este es el número de factores a tomar. En caso de no haberlo se toma un factor.
- Criterio de la gráfica de Cattell's: obtenemos los valores propios de la matriz de entrada y los pintamos de mayor a menor. Analizamos el cambio entre los valores propios y donde se produzca el último cambio brusco contamos el número de valores propios hasta él. Este es el número de factores. Es un método subjetivo y ampliamente criticado.

Métodos para escoger el número de factores

- VSS (Very Simple Structure): este procedimiento toma un modelo simplificado del problema e intenta ver para qué número de factores los valores obtenidos se acercan más a los que deberían ser.
- Comparación de modelos: intentan obtener una medida de cómo de bueno y complejo es el modelo creado con un número de factores dado, de forma que se intenta maximizar el resultado y minimizar la complejidad.
- OC (Optimal Coordinate): intenta eliminar la subjetividad del método de Cattell. Se calculan los gradientes en la misma gráfica de dicho método y se comprueba dónde hay un cambio más abrupto. Esto delimita el número de factores.
- AF (Acceleration Factor): persigue el mismo objetivo que el método anterior, salvo que en este caso se realiza el cálculo con la pendiente de la curva asociada.

Métodos para escoger el número de factores

- MAP (Minimum Average Partial) test: se realiza, desde $k=1$ hasta el número de variables menos 1 un análisis PCA del modelo con k número de factores. Se estudia cómo se comporta dicho modelo con k factores y se toma el valor de k para el cual se ha obtenido el mejor resultado.
- PA (Parallel Analysis): tomamos la misma gráfica que en el método de Cattell y generamos aleatoriamente un conjunto de valores. Estos valores representan los valores medios de los valores propios de matrices aleatorias con el mismo número de variables y datos que la original. Hallamos la media para estos valores y tomamos como número de factores el de los que superen este valor medio. Se puede ver como un refinamiento de la regla de Kaiser ajustando la cota.

Métodos para escoger el número de factores

- Comparación de datos: se realiza una comparación entre modelos con una separación en factores ya conocida como correcta comparando los valores propios de la matriz y los factores escogidos en cada caso. Se toma el número de factores del modelo cuyos valores propios se parezcan más a los del caso que estamos analizando.
- Convergencia de múltiples tests: esta estrategia busca de forma empírica el mejor número de factores analizando la convergencia del modelo.

El objetivo de CFA es, dada una hipótesis sobre el número de factores, si los factores están o no correlacionados y qué variables se relacionan con cada factor. Es por esto que EFA y CFA se usan en contextos muy diferentes aún siendo los dos tipos del mismo método.

Métodos para evaluar la hipótesis

- Ajuste absoluto: consiste en obtener una medida absoluta de cómo de ajusta el modelo hipotético a los valores proporcionados.
- Ajuste relativo: obtiene una medida comparativa de cómo funciona con respecto a un modelo llamado base. Este modelo es fijo y nos sirve para obtener una valoración absoluta del mismo muy mala, de forma que con el ajuste relativo obtenemos una medida de cuánto se acerca la medida de nuestra hipótesis a la de este modelo con mal ajuste.

Explicación del ejemplo en R

Paquetes a instalar

- `install.packages("psych")`: implementa Análisis Factorial

Paquetes a instalar

- `install.packages("psych")`: implementa Análisis Factorial
- `install.packages("GPArotation")`: implementa la Rotación de Factores

Paquetes a instalar

- `install.packages("psych")`: implementa Análisis Factorial
- `install.packages("GPArotation")`: implementa la Rotación de Factores

Paquetes a instalar

- `install.packages("psych")`: implementa Análisis Factorial
- `install.packages("GPArotation")`: implementa la Rotación de Factores

Importa los paquetes

- `library(psych)`

Paquetes a instalar

- `install.packages("psych")`: implementa Análisis Factorial
- `install.packages("GPArotation")`: implementa la Rotación de Factores

Importa los paquetes

- `library(psych)`
- `library(GPArotation)`

Paquetes a instalar

- `install.packages("psych")`: implementa Análisis Factorial
- `install.packages("GPArotation")`: implementa la Rotación de Factores

Importa los paquetes

- `library(psych)`
- `library(GPArotation)`

Funciones a usar

- fa: función del paquete psych que implementa Análisis Factorial.

Funciones a usar

- fa: función del paquete psych que implementa Análisis Factorial.
- factanal: función del paquete stats que implementa Análisis Factorial Exploratorio con máxima verosimilitud.

Funciones a usar

- fa: función del paquete psych que implementa Análisis Factorial.
- factanal: función del paquete stats que implementa Análisis Factorial Exploratorio con máxima verosimilitud.
- prcomp: función del paquete stats que implementa PCA.

Funciones a usar

- `fa`: función del paquete `psych` que implementa Análisis Factorial.
- `factanal`: función del paquete `stats` que implementa Análisis Factorial Exploratorio con máxima verosimilitud.
- `prcomp`: función del paquete `stats` que implementa PCA.
- `principal`: función del paquete `psych` que implementa PCA devolviendo siempre los mejores factores.
- `cfa`: función del paquete `cfa` que implementa el análisis confirmatorio de factores.

Funciones a usar

- `fa`: función del paquete `psych` que implementa Análisis Factorial.
- `factanal`: función del paquete `stats` que implementa Análisis Factorial Exploratorio con máxima verosimilitud.
- `prcomp`: función del paquete `stats` que implementa PCA.
- `principal`: función del paquete `psych` que implementa PCA devolviendo siempre los mejores factores.
- `cfa`: función del paquete `cfa` que implementa el análisis confirmatorio de factores.

En este caso he usado la función `fa` como función principal y `factanal`. Para comparar con PCA he usado `principal`. Para comprar con CFA he usado `cfa`.

¿Preguntas?

