6. ANÁLISIS DE FACTORES (AF)

EL MODELO DEL ANÁLISIS DE FACTORES (AF)

Principio básico:

Un conjunto de p variables X_1, X_2,…..,X_p puede describirse en términos de un menor número de índices o factores, con el propósito de dilucidar la relación existente entre ellas.

Diferencia ACP vs AF

El A.C.P. busca explicar la varianza contenida en una matriz de datos, en tanto que el A.F. está orientado a la identificación de correlaciones entre variables.

EL MODELO DEL ANÁLISIS DE FACTORES (AF)

ORÍGENES (Charles Spearman, 1904)

Spearman observó que, en muchos casos, la matriz de correlación tiene la siguiente interesante propiedad: ***dos filas o columnas cualesquiera son aproximadamente proporcionales si se ignoran los elementos de su diagonal.***

EL MODELO DEL ANÁLISIS DE FACTORES (AF)

EJEMPLO:

Matriz de correlación obtenida a partir de las calificaciones en Clásicos (C), Francés (F), Inglés (E), Matemáticas (M), Discernimiento (D) y Música (Mu), de los chicos de una escuela preparatoria.

	C	F	E	M	D	Mu
C	1.00	0.83	0.78	0.70	0.66	0.63
F	0.83	1.00	0.67	0.67	0.65	0.57
E	0.78	0.67	1.00	0.64	0.54	0.51
M	0.70	0.67	0.64	1.00	0.45	0.51
D	0.66	0.65	0.54	0.45	1.00	0.40
Mu	0.63	0.57	0.51	0.51	0.40	1.00

Columnas → $\dfrac{C}{E}$ $\dfrac{0,83}{0,67} \approx \dfrac{0,70}{0,64} \approx \dfrac{0,66}{0,54} \approx \dfrac{0,63}{0,51} \approx 1,2$

EL MODELO DEL ANÁLISIS DE FACTORES (AF)

En base a ello, Spearman propuso que las calificaciones **(scores)** de las seis pruebas pueden expresarse según el siguiente modelo:

$$X_i = a_i F + e_i \qquad (i = 1,2,...6)$$

X_i = i-ésimo score estandarizado ($m=0$, $s^2=1$)

a_i = constante

F = valor "factorial" (con $m=0$, $s^2=1$ para las observaciones /individuos en conjunto)

e_i = componente de X_i específica para la i-ésima prueba

EL MODELO DEL ANÁLISIS DE FACTORES (AF)

Además se cumple que:

$$\mathrm{var}(X_i) = \mathrm{var}(a_iF+e_i) = \mathrm{var}(a_iF) + \mathrm{var}(e_i) = (a_i)^2\,\mathrm{var}(F) + \mathrm{var}(e_i) = (a_i)^2 + \mathrm{var}(e_i)$$

Puesto que:

a_i = cte.

F y e_i son independientes

var (F) = 1

var (X_i) = 1

Por lo tanto: $\boxed{1 = (a_i)^2 + \mathrm{var}(e_i)}$

De aquí, la constante **a_i**, que se denomina **carga factorial**, es tal que su cuadrado representa la proporción de la varianza de X_i explicada por el factor.

EL MODELO DEL ANÁLISIS DE FACTORES (AF)

Posteriormente, esta teoría fue modificada para permitir que el resultado de cada prueba consista en una parte debida a varios factores comunes más una parte específica de la prueba. El modelo general del AF se expresa, por tanto, como:

$$X_i = a_{i1}F_1 + a_{i2}F_2 + \cdots + a_{im}F_m + e_i$$

donde:

EL MODELO DEL ANÁLISIS DE FACTORES (AF)

X_i = **valor/score** de la i-ésima prueba ($m=0$, $s^2=1$)

$a_{i1}, a_{i2},\ldots, a_{im}$ = **cargas factoriales** de la i-ésima prueba

$F_1, F_2,\ldots,F_m$ = **m factores comunes** no correlacionados, cada uno con $m=0$, $s^2=1$)

e_i = **factor específico** para la i-ésima prueba, no correlacionado con ninguno de los factores comunes y con $m=0$.

EL MODELO DEL ANÁLISIS DE FACTORES (AF)

Con este modelo:

$$\mathrm{Var}(X_i) = 1 = a_{i1}^2\,\mathrm{Var}(F_1) + a_{i2}^2\,\mathrm{Var}(F_2) + \ldots + a_{im}^2\,\mathrm{Var}(F_m) + \mathrm{Var}(e_i)$$
$$= a_{i1}^2 + a_{i2}^2 + \ldots + a_{im}^2 + \mathrm{Var}(e_i)$$

donde: $a_{i1}^2 + a_{i2}^2 + \cdots + a_{im}^2$ se denomina ***comunalidad*** de X_i (la parte de su varianza que está relacionada con los factores comunes), mientras que var(e_i) se denomina la ***especificidad*** de X_i (la parte de su varianza que no está correlacionada con los factores comunes)

EL MODELO DEL ANÁLISIS DE FACTORES (AF)

La correlación entre X_i y X_j está dada por:

$$r_{ij} = a_{i1}a_{j1} + a_{i2}a_{j2} + \cdots + a_{im}a_{jm}$$

De aquí, los ***scores*** de dos pruebas/tests sólo pueden estar fuertemente correlacionados si tienen ***cargas*** elevadas en los mismos factores.

Además, $-1 \leq a_{ij} \leq +1$ puesto que la comunalidad no puede ser superior a 1.

PROCEDIMIENTO DE CÁLCULO PARA EL AF

Datos: Matriz constituida por p variables y n observaciones.

El cálculo comprende 3 etapas:

1. **Determinación de cargas factoriales provisionales, a_{ij}**
   Un modo de hacerlo consiste en efectuar un ACP, seleccionar m CP (por ej.: m = nro. de valores propios > 1, obtenidos a partir de la matriz de correlación), que en este caso serán considerados como factores.

PROCEDIMIENTO DE CÁLCULO PARA EL AF

Los ***factores*** obtenidos de este modo están ***no-correlacionados entre sí y tampoco lo están con los factores específicos (e_i).*** Sin embargo, ***los factores específicos no están no-correlacionados entre sí***, lo que significa que ***no cumplen con una de las hipótesis del modelo de AF.*** No obstante, esto probablemente no constituya problema asegurando que las comunalidades sean altas.

PROCEDIMIENTO DE CÁLCULO PARA EL AF

Cualquiera sea el modo en que se obtengan las cargas factoriales provisionales, se tiene que:

Hipótesis: F_1, F_2,……..F_m son factores provisionales

Tesis: No son únicos

Demostración:

Las combinaciones lineales de los factores provisionales:

$$F_1^* = d_{11}F_1 + d_{12}F_2 + \ldots + d_{1m}F_m$$
$$F_2^* = d_{21}F_1 + d_{22}F_2 + \ldots + d_{2m}F_m$$
$$\vdots$$
$$F_m^* = d_{m1}F_1 + d_{m2}F_2 + \ldots + d_{mm}F_m$$

PROCEDIMIENTO DE CÁLCULO PARA EL AF

Pueden construirse de forma que no estén correlacionados y que "expliquen" los datos tan bien como los factores provisionales.

Existe un número infinito de soluciones alternativas para el modelo de AF, lo que conduce a:

2. **Rotación de factores**

Los factores provisionales se transforman con el propósito de encontrar nuevos factores que faciliten la interpretación.

En este contexto, **rotar** significa esencialmente elegir los valores de d_{ij} en las ecuaciones anteriores.

PROCEDIMIENTO DE CÁLCULO PARA EL AF

3. Cálculo de los scores factoriales

Éstos son los valores de los factores rotados $F_1^*, F_2^*, \ldots., F_m^*$ para cada uno de los individuos/observaciones

ROTACIÓN DE FACTORES

- **Ortogonal**
  Los nuevos factores están no correlacionados, del mismo modo que los factores originales.
- **Oblicua**
  En este caso, los nuevos factores están correlacionados.

Cualquiera sea el tipo de rotación utilizado, es deseable que **las cargas factoriales de los nuevos factores sean próximas a 0 o muy diferentes de 0.**

a_{ij} próximo a 0	**X_i tiene poca relación con el factor F_j**
$a_{ij} \neq 0$ (valor alto, positivo o negativo)	**X_i está determinado por F_j y en una magnitud considerable**

ROTACIÓN DE FACTORES

- **Rotación Varimax (Kaiser, 1958)**

  Es un método de rotación ortogonal, que se basa en asumir que **la interpretabilidad del factor j** puede medirse por la **varianza del cuadrado de sus cargas factoriales**, es decir, la varianza de

  $$a_{1j}^2, a_{2j}^2, \dots, a_{mj}^2$$

  Si esta varianza es alta, luego los valores de a_{ij}^2 tienden a estar o próximos a 0 o próximos a la unidad. Por consiguiente, **la rotación varimax maximiza la suma de dichas varianzas para todos los factores.**

ROTACIÓN DE FACTORES

Rotación Varimax (cont.)

Posteriormente, Kaiser(1958) introdujo la normalización de las cargas factoriales antes de maximizar la varianza de sus cuadrados, puesto que produce mejores resultados.

ANÁLISIS DE FACTORES UTILIZANDO C.P.

La metodología para obtener los factores no rotados es la siguiente:

- Dadas p variables se obtiene el mismo número de C.P.

$$
\begin{aligned}
Z_1 &= b_{11}X_1 + b_{12}X_2 + \ldots + b_{1p}X_p \\
Z_2 &= b_{21}X_1 + b_{22}X_2 + \ldots + b_{2p}X_p \\
&\;\;\vdots \\
Z_p &= b_{p1}X_1 + b_{p2}X_2 + \ldots + b_{pp}X_p
\end{aligned}
\qquad \text{[Ec. 6.1]}
$$

donde los valores de b_{ij} están dados por los vectores propios de la matriz de correlación.

ANÁLISIS DE FACTORES UTILIZANDO C.P.

- Puesto que ésta es una transformación ortogonal, la relación inversa es simplemente:

$$
\begin{aligned}
X_1 &= b_{11}Z_1 + b_{21}Z_2 + \ldots + b_{p1}Z_p \\
X_2 &= b_{12}Z_1 + b_{22}Z_2 + \ldots + b_{p2}Z_p \\
&\;\;\vdots \\
X_p &= b_{1p}Z_1 + b_{2p}Z_2 + \ldots + b_{pp}Z_p
\end{aligned}
$$