Econometría para la gestión

2023

Análisis Factorial

Rodrigo Ortiz

Universidad Alberto Hurtado

Santiago, 2023

Planificación

Contenidos clase de hoy

Introducción

Tenemos una matriz de observaciones por variables. Queremos resumir esta información en unas pocas variables.

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}$$

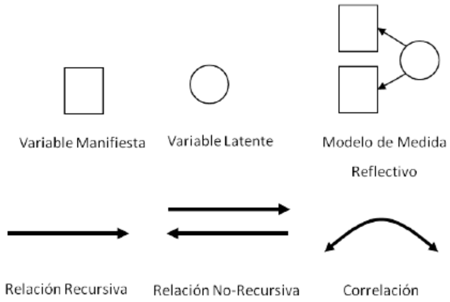
- **Habilidad en matemáticas**
  - Rendimiento en aritmética
  - Rendimiento en álgebra
  - Rendimiento en geometría
- **Importancia de la recompensa para un inventor**
  - Compensación monetaria
  - Oportunidades de un mejor trabajo
  - Satisfacción personal
- **Competitividad industrial**
  - Valor añadido a la manufactura per capita
  - Manufactura exportada per capita
- **Rendimiento organizacional**
  - Operación de mercado
  - Enfoque de aprendizaje organizacional
  - Estilo de emprendimiento
  - Flexibilidad organizacional

Modelo de un factor (**one-factor model**): Charles Spearman en 1904. Propuso un factor general "g" de inteligencia.
Modelo de varios factores (**multi-factor analysis**): Truman Kelly and Louis Leon Thursone (1931).
Dos métodos de estimación:

- Componentes principales
- Máxima verosimilitud (Karl Joreskog, 1969, 1970)

Introducción

Convenciones



Tenemos una variable aleatoria vectorial **x** con p características, con media μ y matriz de varianzas y covarianzas Σ.
El **modelo factorial** postula que **x** es linealmente dependiente de unas pocas variables latentes (no observadas) llamadas **factores**, y p fuentes adicionales de variación $\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots, \varepsilon_p$ llamados **errores** o **factores específicos**.

$$
\begin{aligned}
x_1 - \mu_1 &= l_{11}F_1 + l_{12}F_2 + \cdots + l_{1m}F_m + \varepsilon_1 \\
x_2 - \mu_2 &= l_{21}F_1 + l_{22}F_2 + \cdots + l_{2m}F_m + \varepsilon_2 \\
&\vdots \\
x_p - \mu_p &= l_{p1}F_1 + l_{p2}F_2 + \cdots + l_{pm}F_m + \varepsilon_p
\end{aligned}
$$

$$\mathbf{X}_{px1} - \mu_{px1} = \mathbf{L}_{pxm}F_{mx1} + \varepsilon_{px1}$$

El coeficiente l_{ij}, que a veces se escribe λ_{ij}, se llama **loading** o **carga**, y **L** es la **matriz de factor loadings** o **matriz de cargas**. Hay $m + p$ variables no observadas: $F_1, \ldots, F_m, \varepsilon_1, \ldots, \varepsilon_p$

Modelo con m-factores ortogonales

Modelo con m-factores ortogonales

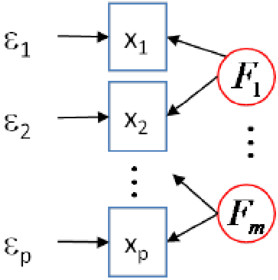


Diagrama de trayectoria
(path diagram)

$$\mathbf{X}_{px1} - \mu_{px1} = \mathbf{L}_{pxm}\mathbf{F}_{mx1} + \varepsilon_{px1}$$

- $\mathbf{F}$ y ε son independientes
- $E(\mathbf{F}) = \mathbf{0}_{mx1}, cov(\mathbf{F}) = E(\mathbf{F}\mathbf{F}') = \mathbf{I}_{mxm}$ (factores ortogonales)
- $E(\varepsilon) = \mathbf{0}_{px1}, cov(\varepsilon) = E(\varepsilon\varepsilon') = \psi_{pxp}$
- $\Psi = \begin{bmatrix} \psi_{11} & 0 & \dots & 0 \\ 0 & \psi_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \psi_{pp} \end{bmatrix}$
- $cov(\varepsilon, \mathbf{F}) = E(\varepsilon\mathbf{F}') = \mathbf{0}_{pxm}$

El modelo factorial ortogonal implica una **estructura de covarianzas para x**.

$$(x-\mu)(x-\mu)' = (\mathbf{LF}+\varepsilon)(\mathbf{LF}+\varepsilon)' = (\mathbf{LF}+\varepsilon)((\mathbf{LF})'+\varepsilon')$$

$$(x-\mu)(x-\mu)' = \mathbf{LF}(\mathbf{LF})' + \varepsilon(\mathbf{LF})' + \mathbf{LF}\varepsilon' + \varepsilon\varepsilon'$$

$$\Sigma = cov(x) = E(x-\mu)(x-\mu)' = \mathbf{L}E(\mathbf{FF}')\mathbf{L}' + E(\varepsilon\mathbf{F}')\mathbf{L}' + \mathbf{L}E(\mathbf{F}\varepsilon') + E(\varepsilon\varepsilon')$$
$$\Sigma = \mathbf{LL}' + \Psi$$

$$cov(\mathbf{x},\mathbf{F}) = E(x-\mu)\mathbf{F}' = \mathbf{L}E(\mathbf{FF}') + E(\varepsilon\mathbf{F}') = \mathbf{L}$$

El modelo factorial ortogonal implica una **estructura de covarianzas para x**.

$$\Sigma = \mathbf{LL}' + \Psi$$

$$Var(x_i) = l_{i1}^2 + \ldots + l_{im}^2 + \psi_i$$

$$cov(x_i, x_k) = l_{i1} l_{k1} + \ldots + l_{im} l_{km}$$

$$cov(x, F) = L$$

$$cov(x_i, F_j) = l_{ij}$$

Supuesto de linealidad.
El modelo $x - \mu = \mathbf{LF} + \varepsilon$ es **lineal** en los factores. Esto significa que no hay relaciones del tipo (interacciones).

$$x_i - \mu_i = l_{11} F_1 F_3 + \varepsilon_1$$

En este caso, la estructura supuesta para $\Sigma, \mathbf{LL}', \Psi$, puede no ser adecuada (cov($\mathbf{F}$)=$\mathbf{I}$)

La proporción de varianza explicada.

- **i-esima comunalidad**, h_i^2: la proporción de la varianza de la i-esima variable explicada por los m factores.
- **Unicidad o varianza específica**: la porción de $Var(x_i) = \sigma_{ij}$ debido al factor específico o error.

$$Var(x_i) = \text{communalidad y varianza específica}$$

$$\sigma_{ii} = l_{i1}^2 + l_{i2}^2 + \ldots + l_{im}^2 + \psi_i$$

$$h_i^2 = l_{i1}^2 + l_{i2}^2 + \ldots + l_{im}^2$$

$$\sigma_{ii} = h_i^2 + \psi_i, \; i = 1, 2, \ldots, p$$

Indeterminancia - Hay una ambiguedad asociada al modelo factorial.

$$x - \mu = \mathbf{LF} + \varepsilon$$

Si $\mathbf{T}_{mxm}$ es una matriz ortogonal tal que $\mathbf{TT}' = \mathbf{T}'\mathbf{T} = \mathbf{I}$ entonces:

$$x - \mu = \mathbf{LF} + \varepsilon = \mathbf{LTT}'\mathbf{F} + \varepsilon = \mathbf{L}^*\mathbf{F}^*$$

No podemos distinguir entre $\mathbf{L}$ y $\mathbf{L}^*$, tampoco entre $\mathbf{F}$ y $\mathbf{F}^*$, aunque diferentes, generan la misma matriz de varianzas y covarianzas Σ.

Rotación de los loadings

Debemos rotar la matriz $\mathbf{L}$ utilizando una matriz ortogonal $\mathbf{T}$ hasta que los valores de las matrices $\mathbf{L}$ y $\mathbf{L}^*$ entreguen la misma representación. Las comunalidades dadas por las diagonales de $\mathbf{LL}'=(\mathbf{F})(\mathbf{F}^*)'$ no se ven afectadas por $\mathbf{T}$.

En el análisis factorial se deben imponer las restricciones necesarias de manera que $\mathbf{L}$ y Ψ sean determinados de forma única. Una vez que la matriz de loadings y varianzas específicas son determinadas, el valor de los factores $\mathbf{F}$ (factor scores) puede ser determinado.

Métodos de estimación

Tenemos dos métodos de estimación:

- Componentes principales
- Máxima verosimilitud

Calculando los **valores y vectores propios** de la matriz de varianzas y covarianzas **S** podemos estimar las matrices **L** y Ψ.
Tenemos $(\hat{\lambda}_1, \hat{u_1}), \ldots, (\hat{\lambda}_p, \hat{u_p})$, donde $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_p$, y $m < p$, entonces:

$$\tilde{\mathbf{L}} = [\sqrt{\hat{\lambda}_1}\hat{u_1} \vdots \sqrt{\hat{\lambda}_2}\hat{u_1} \vdots \ldots \vdots \sqrt{\hat{\lambda}_m}\hat{u_m}]$$

$$\tilde{\Psi} = S - \tilde{L}\tilde{L}', \text{ con } \tilde{\psi}_i = s_{ii} - \textstyle\sum_{j=1}^{m} \tilde{l}_{ij}^2$$

$$\tilde{h}_i^2 = \tilde{l}_{i1}^2 + \tilde{l}_{i2}^2 + \ldots + \tilde{l}_{im}^2$$

Proporción de la varianza muestral total explicada por el factor j:

$$\frac{\tilde{\lambda}_J}{S_{11}+s_{22}+...+s_{pp}}, \text{ análisis factorial de la matriz } \mathbf{S}$$

$$\frac{\tilde{\lambda}_J}{p}, \text{ análisis factorial de la matriz } \mathbf{R}$$

$$S_{11} + s_{22} + \ldots + s_{pp} = traza(S)$$

Criterio para seleccionar el número de factores. Otro criterio:

- $m =$ número de valores propios de R¿1
- $m =$ número de valores propios de S¿0

$$x - \mu = \mathbf{LF} + \varepsilon$$

Si $\mathbf{F}$ y ε pueden ser asumidos normalmente distribuidos, entonces las observaciones $\mathbf{x}$ son también normales, y $\mathbf{L}$ y Ψ pueden ser estimados por máxima verosimilitud.

Seleccionamos el valor de los parámetros de manera de maximizar la densidad conjunta evaluada en estas observaciones. Esta es la estimación por máxima verosimilitud y los parámetros se denominan **estimadores de máxima verosimilitud**.

Los estimadores máximo verosímiles de las comunalidades:

$$\tilde{h}_i^2 = \tilde{l}_{i1}^2 + \tilde{l}_{i2}^2 + \ldots + \tilde{l}_{im}^2 \text{ para } i = 1, 2, \ldots, p$$

Entonces:

$$\text{Proporción varianza por factor } j = \frac{\tilde{l}_{1j}^2 + \tilde{l}_{2j}^2 + \ldots + \tilde{l}_{pj}^2}{s_{11} + s_{22} + \ldots + s_{pp}}$$

Rotación de los loadings

Rotamos la matriz de loadings, $\hat{L}$, utilizando una matriz ortogonal T hasta que los valores de las matrices $\hat{L}$ y $\hat{L}^*$ entreguen la misma representación.

- La matriz residual, $S_n - \hat{L}\hat{L}' - \hat{\Psi} = S_n - \hat{L}^*\hat{L}^{*\prime} - \hat{\Psi}$
- Las comunalidades dadas por las diagonales de $\hat{L}\hat{L}' = (F^*)(F^*)'$ no se ven afectadas por **T**.

Rotamos la matriz de loadings hasta que podamos obtener una interpretación simple. Recuerda que los factores representan variables no observadas y queremos que cada variable se relacione **más** con un factor (que lo refleje) de manera de interpretar el modelo. Por lo que es deseable que esta estructura simple logre este objetivo.

Rotación de los loadings

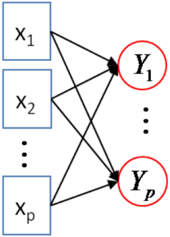
Se llama rotación varimax cuando rotamos los loadings y maximizamos la suma de las varianzas de los cuadrados de los loadings. Definimos entonces una medida proporcional a este valor. Si $\tilde{l}_{ij}^{*} = \hat{l}_{ij}^{*}/\hat{h}_i$ es el coeficiente rotado por la raiz cuadrada de la comunalidad, entonces el procedimiento varimax selecciona la matriz ortogonal $\mathbf{T}$ tal que la medida:

$$V = \frac{1}{p}\sum_{j=1}^{m}\left[\sum_{i=1}^{p}\tilde{l}_{ij}^{*4} - \left(\frac{(\sum_{i=1}^{p}\tilde{l}_{ij}^{*2})^2}{p}\right)\right]$$

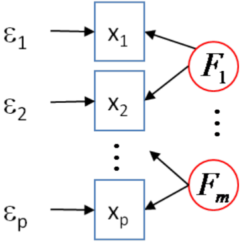
sea tan grande como sea posible.

Comparando componentes principales y análisis factorial

Componentes Principales



Análisis Factorial



Comparando componentes principales y análisis factorial

Análisis de Componentes Principales	**Análisis Factorial**
Karl Pearson (1901)	Charles Spearman (1904)
Ciencias Naturales	Ciencias Sociales
Sin errores de medida	Errores de medida
Sin variables no observadas	Variables no observadas
Explicar la varianza común y unica de las variables	Explicar la varianza común
Maximizar la varianza total de las variables manifiestas	Explicar las correlaciones (fuera de la diagonal) entre las variables manifiestas
Loadings tienden a ser más grandes	Loadings tienden a ser cercanos a los parámetros de la población (Cud07)
Loadings con menores errores estandar	Loadings con mayores errores estandar (Oga03)
Las componentes son únicas	Las componentes no son únicas
No hay problemas de indeterminancia	Hay problemas de indeterminancia (Ste79,Ste96)
Menor uso intensivo del computador	Mayor uso intensivo del computador (VelJac90)

Bécue-Bertaut, M. (2010) Minería de Textos. Aplicación a Preguntas Abiertas en Encuestas, La Muralla, Madrid.

Escofier, B., Pages, J. (1992) Análisis Factoriales Simples y Múltiples, Universidad del País Vasco, Bilbao.

Johnson, R. Wichern, D. (2007) Applied Multivariate Statistical Analysis, Prentice Hall, New Jersey.

Peña, D. (2002) Análisis de Datos Multivariantes, McGraw–Hill, Madrid.