Análisis Multivariado

ID 033521 - Clase 4901

Lina Maria Acosta Avena

Ciencia de Datos
Departamento de Matemáticas
Pontificia Universidad Javeriana

Semana 12: 07/10/24 – 12/10/24

Introducción

Recuerde que:

- El modelo factorial ortogonal (**AF**) se **aplica** en situaciones donde las **variables originales** ($X_i \in \mathbf{X}_{p\times 1} \quad i = 1, 2, \ldots, p$) **estén correlacionadas entre sí**, pues en caso contrario, cada factor ($\mathbf{F}_j,\ j = 1, \ldots, m$) estaría relacionado con sólo una variable original y en ese caso $m = p$.
- En algunos casos, **la interpretacion de los factores no es tan simple, clara o directa**, pues los valores de las cargas factoriales ($l_{ij},\ i = 1, 2, \ldots, p,\ j = 1, 2, \ldots, m$) son muy similares y no despreciables en otros factores. Esto se debe a **violaciones del supuesto de ortogonalidad en los factores**. En estos casos se recurre a una **transformación ortogonal de los factores**, la cual básicamente consiste en **rotaciones de los factores**.

Introducción

Example

Lawley y Maxwell presentan la **matriz de correlación** muestral

$$\mathbf{R} = \begin{bmatrix} 1.0 & 0.439 & 0.410 & 0.288 & 0.329 & 0.248 \\ 0.439 & 1.0 & 0.351 & 0.354 & 0.320 & 0.329 \\ 0.410 & 0.351 & 1.0 & 0.164 & 0.190 & 0.181 \\ 0.288 & 0.354 & 0.164 & 1.0 & 0.595 & 0.470 \\ 0.329 & 0.320 & 0.190 & 0.595 & 1.0 & 0.464 \\ 0.248 & 0.329 & 0.181 & 0.470 & 0.464 & 1.0 \end{bmatrix}$$

de las **puntuaciones de exámenes en Gaélico, Inglés, Historia, Aritmética, Álgebra y Geometría**, para estudiantes varones.

Example 9.8 de Johnson and Wichern (2013) - Applied Multivariate Statistical Analysis, pp. 505.

Introducción

Haciendo un análisis con $m = 2$ **factores** utilizando el método de máxima verosimilitud

```
require(psych)
R <- matrix(c(1.0,0.439,0.410,0.288,0.329,0.248,
              0.439,1.0,0.351,0.354,0.320,0.329,
              0.410,0.351,1.0,0.164,0.190,0.181,
              0.288,0.354,0.164,1.0,0.595,0.470,
              0.329,0.320,0.190,0.595,1.0,0.464,
              0.248,0.329,0.181,0.470,0.464,1.0),
            ncol=6)
AF.ml<-fac(R, nfactors = 2, rotate = "none", fm="ml")
L.m<-AF.ml$loadings      # cargas factoriales
```

Introducción

```
L.m
Loadings:
      ML1    ML2
[1,]  0.553  0.429
[2,]  0.568  0.288
[3,]  0.392  0.450
[4,]  0.740 -0.273
[5,]  0.724 -0.211
[6,]  0.595 -0.132

h2 <- AF.ml$communality  # comunalidades
h2
[1] 0.4898329 0.4059266 0.3562688 0.6226474
[5] 0.5686419 0.3717909
```

Introducción

tenemos:

Variables	Cargas Factoriales		Comunalidades
	F_1	F_2	
Gaélico	0.553	0.429	0.490
Inglés	0.568	0.288	0.406
Historia	0.392	0.450	0.356
Aritmética	0.740	-0.273	0.623
Álgebra	0.724	-0.211	0.569
Geometría	0.595	-0.132	0.372

Introducción

- En $\mathbf{F}_1$, **todas las cargas son positivas**. Los estudiantes que obtengan **puntuaciones altas en todas las pruebas** obtendrán **puntuaciones altas en este factor**. Así que éste factor podría se llamado **Inteligencia general**.

Introducción

- En $\mathbf{F}_2$, la **primera mitad de las cargas son positivas,** mientras que **la segunda son negativas (factor bipolar)**[1]. Así que, por un lado, aquellos estudiantes con puntuaciones superiores a la media en las pruebas verbales, obtendrán puntuaciones superiores a la media en este factor. Y, aquellos que obtengan puntuaciones superiores a la media en las pruebas de matemáticas obtienen puntuaciones inferiores a la media en este factor. Así que éste factor podría se llamado **no-matemático/idiomas/verbal**.

[1] *"la asignación de polos negativo y positivo es arbitraria, porque los signos de las cargas en un factor se pueden invertir sin afectar el análisis"*

Introducción

- Las variables **Aritmética, Álgebra y Geometría**, tienen **cargas factoriales altas en $\mathbf{F}_1$** y **bajas** en $\mathbf{F}_2$. Mientras que las variables **Gaélico, Inglés e Historia**, tienen **cargas relativamente similares en ambos factores**. Entonces, la interpretación de los factores no es del todo clara.

Introducción

- Las variables **Aritmética, Álgebra y Geometría**, tienen **cargas factoriales altas en $\mathbf{F}_1$** y **bajas** en $\mathbf{F}_2$. Mientras que las variables **Gaélico, Inglés e Historia**, tienen **cargas relativamente similares en ambos factores**. Entonces, la interpretación de los factores no es del todo clara.

¿Qué hacemos?

Introducción

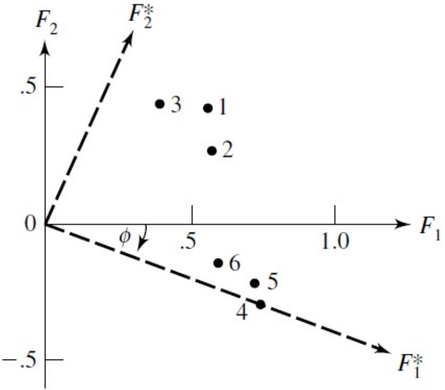

Figura: Figura 9.1 de Johnson and Wichern (2013)

Introducción

Observe que si se **rotan los factores**:

- Las variables Aritmética (4), Álgebra (5) y Geometría (6) se cargan mucho en $\mathbf{F}_1^*$ y tienen cargas insignificantes en $\mathbf{F}_2^*$. Así que, este factor podría llamarse **capacidad matemática**.
- Las variables Gaélico (1), Inglés (2) e Historia (3), tienen cargas altas en $\mathbf{F}_2^*$ y cargas de moderadas a pequeñas en $\mathbf{F}_1^*$. Así que, este factor podría llamarse **capacidad verbal**.
- El factor de **inteligencia general** identificado inicialmente se encuentra sumergido en los factores $\mathbf{F}_1^*$ y $\mathbf{F}_2^*$.

Introducción

Observe que si se **rotan los factores**:

- Las variables Aritmética (4), Álgebra (5) y Geometría (6) se cargan mucho en $\mathbf{F}_1^*$ y tienen cargas insignificantes en $\mathbf{F}_2^*$. Así que, este factor podría llamarse **capacidad matemática**.
- Las variables Gaélico (1), Inglés (2) e Historia (3), tienen cargas altas en $\mathbf{F}_2^*$ y cargas de moderadas a pequeñas en $\mathbf{F}_1^*$. Así que, este factor podría llamarse **capacidad verbal**.
- El factor de **inteligencia general** identificado inicialmente se encuentra sumergido en los factores $\mathbf{F}_1^*$ y $\mathbf{F}_2^*$.

¿ Cómo hacer la rotación?

Rotación de Factores

Recuerde que:

$$\mathbf{P} = \mathbf{L}\,\mathbf{L}^\top + \boldsymbol{\Psi}$$

Como el interés es hacer una rotación en $\mathbf{L}$, considere

$$\widehat{\mathbf{L}}^* = \widehat{\mathbf{L}}\,\mathbf{T}$$

donde

- $\widehat{\mathbf{L}}$ es la matriz obtenida el modelo ortogonal.
- $\mathbf{T}$ es una matriz ortogonal, esto es $\mathbf{T}^\top\,\mathbf{T} = \mathbf{T}\,\mathbf{T}^\top = \mathbf{I}$.

Note que

$$\widehat{\mathbf{L}}^*\widehat{\mathbf{L}}^{*\top} = \left(\widehat{\mathbf{L}}\,\mathbf{T}\right)\left(\widehat{\mathbf{L}}\,\mathbf{T}\right)^\top = \widehat{\mathbf{L}}\,\mathbf{T}\,\mathbf{T}^\top\,\widehat{\mathbf{L}}^\top = \widehat{\mathbf{L}}\,\mathbf{I}\,\widehat{\mathbf{L}}^\top = \widehat{\mathbf{L}}\widehat{\mathbf{L}}^\top$$

Así que la rotación $\widehat{\mathbf{L}}^* = \widehat{\mathbf{L}}\,\mathbf{T}$ no altera $\mathbf{P} = \mathbf{L}\,\mathbf{L}^\top + \boldsymbol{\Psi}$.

Rotación de Factores

Observaciones:

- *La rotación implica que, apartir de una solución* $\boldsymbol{F}$*, siempre es posible obtener una nueva solución* $\boldsymbol{F}^{*} = \boldsymbol{T}^{\top}\boldsymbol{F}$ *que proporciona una interpretación más fácil que* $\boldsymbol{F}$.
- *En términos de bondad de ajuste, **la nueva solución no proporciona nunguna mejora con respecto al ajuste inicial**, pues la matriz residual, las comunalidades y las varianzas específicas no son alteradas.*
- *Cuando **la solución sin rotación es de buena calidad (interpretable), no es necesario hacer rotación ortogonal**. En general, cuando la rotación es innecesaria y se aplica, la interpretación de los factores rotacionados suelen ser de peor calidad.*

Rotación de Factores

Lo ideal sería utilizar/aplicar una transformación **T** que proporcione cargas factoriales con valores altos en un factor y valores pequeño o moderados en el resto de los factores.

Rotación de Factores

Existen varios **métodos/criterios ortogonales** para determinar **T** adecuada:

- **Varimax:** Propuesto por Kaiser (1958), tiene como objetivo principal encontrar factores con máxima variabilidad en las cargas factoriales.
- **Cuartimax:** Propuesto por Jobson (1996), tiene como objetivo encontrar factores que lleven al máximo la variabilidad de los cuadrados de las cargas factoriales sobre todos los factores y todas las variables. Este método tiende a generar factores, donde todas las variables tienen cargas elevadas.
- **Ortomax:** Propuesto por Jobson (1996), básicamente es una medida ponderada de Varimax y Cuartimax.

Rotación de Factores

También existen **criterios no ortogonales u oblicuos**, los cuales alteran los supuestos del modelo factorial original, específicamente:

- las rotaciones no preservan: la estructura de ajuste original del modelo factorial, la matriz residual, las comunalidades y las matrices específicas.
- los factores son correlacionados entre sí y la grandeza de las correlaciones depende del tipo de rotación que sea utilizada, es decir, de los ángulos de rotación utilizados. Esto implica una mayor dificultad en la interpretación de los factores.

Rotación de Factores

Haremos la rotación en nuestro ejemplo usando el criterio varimax:

```
AF.ml<-fac(R,
           nfactors = 2,
           rotate = "varimax",
           fm="ml")
L.m<-AF.ml$loadings      # cargas factoriales
h2 <- AF.ml$communality  # comunalidades
psi<- AF.ml$uniquenesses # especificidades
```

Rotación de Factores

Variables	Cargas Factoriales		Comunalidades
	$\mathbf{F}_1$	$\mathbf{F}_2$	
Gaélico	0.235	0.659	0.490
Inglés	0.323	0.549	0.406
Historia		0.590	0.356
Aritmética	0.771	0.170	0.623
Álgebra	0.724	0.213	0.569
Geometría	0.572	0.210	0.372

Note que ahora, las cargas factoriales de Aritmética, Álgebra y Geometría son altas en $\mathbf{F}_1$ y relativamente bajas en $\mathbf{F}_2$. Así mismo, las cargas factoriales de Gaélico, Inglés e Historia son relativamente bajas en $\mathbf{F}_1$ y moderadas en $\mathbf{F}_2$. Además, observe que las comunalidades se mantinen.

Example

En un estudio de preferencias del consumidor, se pidió a una muestra aleatoria de clientes que calificaran los siguientes atributos de un nuevo producto: gusto, costo, sabor, tamaño por porcion, y caloría suministradas. Se tabularon las respuestas, en una escala diferencial semántica de 7 puntos, y se construyó la matriz de correlación de atributos:

$$\mathbf{R} = \begin{bmatrix} 1.00 & 0.02 & 0.96 & 0.42 & 0.01 \\ 0.02 & 1.00 & 0.13 & 0.71 & 0.85 \\ 0.96 & 0.13 & 1.00 & 0.50 & 0.11 \\ 0.42 & 0.71 & 0.50 & 1.00 & 0.79 \\ 0.01 & 0.85 & 0.11 & 0.79 & 1.00 \end{bmatrix}$$

Example 9.9 de Johnson and Wichern (2013) - Applied Multivariate Statistical Analysis, pp. 508.

Rotación de Factores

Hacemos el AF con $m = 2$ **factores** utilizando el método de **factores principales:**

```
R<-matrix(c(1.00 , 0.02 , 0.96 , 0.42 , 0.01,
            0.02 , 1.00 , 0.13 , 0.71 , 0.85,
            0.96 , 0.13 , 1.00 , 0.50 , 0.11,
            0.42 , 0.71 , 0.50 , 1.00 , 0.79,
            0.01 , 0.85 , 0.11 , 0.79 , 1.00),ncol=5)

# ----- AF sin rotación ----- #
AF.fp<-principal(R, nfactors = 2,rotate = "none")

L.m<-AF.fp$loadings # cargas factoriales
h2 <- AF.fp$communality  # comunalidades
```

Rotación de Factores

```
# ----- AF con rotacion ----- #
AF.fp<-principal(R, nfactors = 2, rotate = "varimax")

L.m<-AF.fp$loadings      # cargas factoriales
h2 <- AF.fp$communality  # comunalidades
```

Variables	Factores		Factores Rotados		Cumunalidades
	$\mathbf{F}_1$	$\mathbf{F}_2$	$\mathbf{F}_1^*$	$\mathbf{F}_2^*$	
Gusto	0.56	0.82		0.99	0.98
Costo	0.78	-0.52	0.94		0.88
Sabor	0.65	0.75	0.13	0.98	0.98
Tamaño	0.94	-0.10	0.84	0.43	0.89
Caloría	0.80	-0.54	0.97		0.93

Rotación de Factores

Apesar de que en ambos casos, las variables costo, tamaño y caloría definen al factor 1, las variables costo y caloría tienen cargas moderadas en $\mathbf{F}_2$, lo cual mejora cuando se hace la rotación ($\mathbf{F}_2^*$). Análogamente sucede con las variables gusto y sabor, tienen cargas altas en $\mathbf{F}_2$ y moderadas en el factor $\mathbf{F}_1$. Las cargas de éstas variables aumentaron en $\mathbf{F}_2^*$ y disminuyeron en $\mathbf{F}_1^*$. Vale la pena resaltar que la variable tamaño, tiene carga factorial alta (0.84) en $\mathbf{F}_1^*$ y moderada (0.43) en $\mathbf{F}_2^*$. Esto implica que esta variable está más estrechamente alineada con el factor 1, aunque tiene aspectos del rasgo representado por el factor 2. Luego, podríamos llamar al **factor 1** un factor **nutricional** y al **factor 2** un factor de **sabor**.

Comparación entre AF y ACP

Apesar de que AF y ACP son técnicas de reducción de la dimensionalidad de los datos, los criterios de optimización usados en cada caso son diferentes:

- ✓ El objetivo principal del ACP es explicar la variabilidad de los datos, mientas que el AF busca explicar las covacianzas (correlaciones) entre las variables por medio de factores comunes.
- ✓ El ACP es una transformación de los datos y no se hacen supuestos sobre las matrices analizar, mientras que en el AF se asume un modelo para los datos, sobre el cual se consideran unos supuestos.

Pontificia Universidad JAVERIANA Bogotá

Comparación entre AF y ACP

- ✓ En el ACP las "nuevas variables" forman un índice, mientras que el AF estos índices (factores) reflejan un atributo no observable (latente).
- ✓ En el ACP la tención recae sobre la variación total contenida en las variables, mientras que en el AF se dirige a la parte del total de varianza que es compartida por las variables.
- ✓ En el ACP las variables se "agregan" adecuadamente para definir nuevas variables, mientras que en el AF las variables se "desagregan" convenientemente en una serie de factores comunes desconocidos y en una parte propia de cada variable.