

Estadística Aplicada III

Análisis Factorial

Jorge de la Vega Góngora

Departamento de Estadística,
Instituto Tecnológico Autónomo de México

Semana 7: 26/28 de septiembre de 2018

Introducción

Origen y propósito del análisis factorial (AF) I

- Desarrollado originalmente por Charles Spearman y Karl Pearson (1904) como primeros intentos para medir conceptos abstractos o *constructos* en psicometría, como *la inteligencia, la personalidad, satisfacción*, etc.
- Cuando una muestra de individuos toma una batería de pruebas diseñadas para medir habilidades cognitivas (eg: matemáticas, razonamiento, comprensión de lectura), es un hecho de que sus scores están altamente correlacionados:
 - ▶ Una persona que obtuvo altos scores en una prueba tiende a tener altos scores en todas.
 - ▶ Las personas que tuvieron altos scores en todas las pruebas son aquellas que usualmente se consideran como 'más inteligentes'.
- Se postuló que la correlación se daba porque en realidad las personas variaban en una característica no observable directamente, la inteligencia, y sus scores eran dependientes de su nivel de inteligencia. A este nivel se le llama usualmente el *coeficiente intelectual* (o *IQ*).
- El principal propósito de AF es *describir* las relaciones de dependencia (a través de las covarianzas) entre muchas variables, en términos de unas cuantas variables explicativas subyacentes *no observables* o *latentes* llamados *factores*.
- El AF se aplica en dos contextos de análisis de datos:

Origen y propósito del análisis factorial (AF) II

- ▶ **Análisis exploratorio:** cuando no hay hipótesis o teoría y se explora una estructura subyacente
 - ▶ **Análisis confirmatorio:** cuando hay una teoría o hipótesis subyacente y se pretende confirmar o negar la estructura propuesta. En este caso, se pretende hacer una inferencia y se requieren verificar normalidad de los datos.
- Si las variables se pueden agrupar por sus correlaciones, entonces se puede concebir que cada grupo de variables son realizaciones de un *constructo* o *factor* que es el responsable de la alta correlación en ese grupo.

Ejemplos I

- Algunos constructos generales:

Factor	Variables correlacionadas
stress laboral	desempeño, horas trabajadas, tipo de trabajo
clase social	ingreso, educación, ocupación
Inteligencia	matemáticas, lengua, música
Capacidad atlética	Fuerza, destreza, gimnasia
Personalidad	Introversión, juicio, temperamental
raza	medidas de cuerpo humano

- Ejemplos de constructos en marketing de autos (Kachigan, 1982):

Factor	Variables correlacionadas
comodidad	espacio interior; espacio cajuela; cómodo
eficiencia de costo	costo reparaciones; espacio cajuela; consumo gasolina; cómodo
estilo	colores; diseño; apariencia; atractivo a la vista
facilidad de manejo	facilidad estacionar; accesorios; suavidad manejo; velocidades

Modelo

Modelo básico

- Si \mathbf{x} es un vector con media $\boldsymbol{\mu}$ y covarianza $\boldsymbol{\Sigma}$, el modelo supone que los datos se generan de acuerdo al siguiente modelo:

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{L}\mathbf{f} + \boldsymbol{\epsilon}$$

donde

- $\mathbf{L}_{p \times m}$ es una matriz de *coeficientes de carga*
 - $\mathbf{f}_{m \times 1}$ es el vector de *factores comunes*
 - $\boldsymbol{\epsilon}_{p \times 1}$ es un vector de errores o *factores específicos*
- Usualmente, \mathbf{f} y $\boldsymbol{\epsilon}$ son vectores *no observables* o *latentes*, a diferencia de un modelo de regresión, en donde se supone que los predictores son observables.
- Como el modelo cuenta con más parámetros que variables, se requiere imponer restricciones complementarias para especificar el problema:
 - los factores comunes y los específicos son independientes:

$$\mathbf{f} \perp \boldsymbol{\epsilon}$$

- los factores comunes son independientes (ortogonales) entre sí:

$$E(\mathbf{f}) = \mathbf{0} \text{ y } \text{Var}(\mathbf{f}) = \mathbf{I}_m$$

- los factores específicos también son independientes entre sí, con

$$E(\boldsymbol{\epsilon}) = \mathbf{0} \text{ y } \text{Var}(\boldsymbol{\epsilon}) = \boldsymbol{\Psi} = \text{diag}(\psi_i)$$

Estructura de covarianza subyacente de \mathbf{x}

- El modelo de factores ortogonales implica una estructura específica para la covarianza de \mathbf{x} :

$$\begin{aligned}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})' &= (\mathbf{L}\mathbf{f} + \boldsymbol{\epsilon})(\mathbf{L}\mathbf{f} + \boldsymbol{\epsilon})' \\ &= (\mathbf{L}\mathbf{f})(\mathbf{L}\mathbf{f})' + \boldsymbol{\epsilon}\boldsymbol{\epsilon}' + (\mathbf{L}\mathbf{f})\boldsymbol{\epsilon}' + \boldsymbol{\epsilon}(\mathbf{L}\mathbf{f})'\end{aligned}$$

Entonces, tomando esperanza de ambos lados:

$$\begin{aligned}\boldsymbol{\Sigma} &= E[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})'] \\ &= \mathbf{L}E[\mathbf{f}\mathbf{f}']\mathbf{L} + E[\boldsymbol{\epsilon}\boldsymbol{\epsilon}'] + \mathbf{L}E[\mathbf{f}\boldsymbol{\epsilon}'] + E[\boldsymbol{\epsilon}\mathbf{f}']\mathbf{L}' \\ &= \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}\end{aligned}$$

Entonces la varianza se puede descomponer en la variabilidad de los factores comunes y de los factores específicos.

- También hay que notar la dependencia entre \mathbf{x} y \mathbf{f} :

$$\text{cov}(\mathbf{x}, \mathbf{f}) = E[(\mathbf{x} - \boldsymbol{\mu})\mathbf{f}'] = E[(\mathbf{L} + \boldsymbol{\epsilon})\mathbf{f}'] = \mathbf{L}E[\mathbf{f}\mathbf{f}'] + E[\boldsymbol{\epsilon}\mathbf{f}'] = \mathbf{L}$$

Entonces la dependencia de las variables originales con los factores no observados está dada por la matriz de cargas.

Ejemplo I

Cualquier matriz definida positiva se puede escribir de la forma $\Sigma = LL'$ a través de la descomposición espectral, por lo que en ese caso $\Psi = 0$. Por ejemplo:

```
Sigma <- matrix(c(19,30,2,12,30,57,5,23,2,5,38,47,12,23,47,68),ncol=4)
Sigma

      [,1] [,2] [,3] [,4]
[1,]   19   30   2   12
[2,]   30   57   5   23
[3,]    2    5   38   47
[4,]   12   23   47   68

L <- NULL
for(i in 1:4) L <- cbind(L,eigen(Sigma)$vectors[,i]*sqrt(eigen(Sigma)$values[i]))
L

      [,1]      [,2]      [,3]      [,4]
[1,] -2.538012  3.266721  1.3246848 -0.36364966
[2,] -4.754940  5.809211 -0.8010499 -0.04392532
[3,] -5.075972 -3.337857 -0.2620398 -1.01220303
[4,] -7.839177 -2.419968  0.2266801  0.79979361

L %*% t(L)

      [,1] [,2] [,3] [,4]
[1,]   19   30   2   12
[2,]   30   57   5   23
[3,]    2    5   38   47
[4,]   12   23   47   68
```

Ejemplo II

La misma matriz se puede escribir considerando sólo dos factores:

$$\begin{pmatrix} 19 & 30 & 2 & 12 \\ 30 & 57 & 5 & 23 \\ 2 & 5 & 38 & 47 \\ 12 & 23 & 47 & 68 \end{pmatrix} = \begin{pmatrix} 4 & 1 \\ 7 & 2 \\ -1 & 6 \\ 1 & 8 \end{pmatrix} \begin{pmatrix} 4 & 7 & -1 & 1 \\ 1 & 2 & 6 & 8 \end{pmatrix} + \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$

La matriz $\mathbf{L}_{4 \times 2}$ tiene elementos l_{ij} . Por ejemplo:

$$19 = \sigma_{11} = l_{11}^2 + l_{12}^2 + \psi_1 = 4^2 + 1^2 + 2 = 19$$

Posibles variaciones al modelo

Los supuestos escogidos para especificar el modelo factorial pueden ser diferentes:

- **Modelo con factores oblicuos:** Cuando se supone que los factores no son ortogonales, y se permite una estructura de covarianza entre ellos:

$$\text{Var}(\mathbf{f}) = \Phi$$

- ▶ Son mucho más complicados de modelar matemáticamente
- ▶ Se pierden algunas propiedades del modelo
- ▶ En algunos contextos, facilitan la interpretación de los factores
- **Modelos no lineales de factores:** Cuando la relación entre los factores y las variables originales no es lineal.

$$\mathbf{x} - \boldsymbol{\mu} = \mathbf{G}(\mathbf{f}) + \boldsymbol{\epsilon}$$

- ▶ El modelo es mucho más difícil de estimar
- ▶ Incluye algunas versiones bayesianas

Estos casos no serán cubiertos en este curso.

Comunalidades

- Consideremos la descomposición de la varianza de x en dos partes:
 - ▶ La proporción de la varianza de la i -ésima variable contribuida por los factores comunes se conoce como *la i -ésima comunalidad*:

$$h_i^2 = \sum_{j=1}^m l_{ij}^2$$

- ▶ La porción de la varianza que se debe a los factores específicos se conoce como *unicidad o varianza específica*
- Entonces la varianza de cada variable se puede ver como:

$$\text{Variabilidad de } X_i = \text{Comunalidad} + \text{Unicidad}$$

ó

$$\text{Variabilidad de } X_i = \left(\begin{array}{c} \text{Factores} \\ \text{comunes} \end{array} \right) + \left(\begin{array}{c} \text{Factores} \\ \text{específicos} \end{array} \right)$$

ó

$$\text{Var}(X_i) = \sum_{j=1}^m l_{ij}^2 + \psi_i = h_i^2 + \psi_i$$

- El modelo de factores supone que los $p(p+1)/2$ elementos de Σ se pueden reproducir de los $pm + p = p(m+1)$ elementos de \mathbf{L} y ϵ
- Cuando m es pequeño comparado con p es cuando el modelo de factores es más útil.
- Cuando $m = p$ no hay ningún beneficio, porque $\Sigma = \mathbf{L}\mathbf{L}'$ y $\Psi = \mathbf{0}$.

Problemas con el modelo

Problemas generales con el modelo básico de FA

Hay dos problemas básicos generales:

1. El problema puede no tener una solución estadística válida.
2. La descomposición propuesta para la varianza no está completamente determinada.

Problema 1: Soluciones no factibles I

Ejemplo. [Solución no propia]

Con $p = 3$ y $m = 1$ y suponiendo que la matriz de covarianzas de $\Sigma = \begin{pmatrix} 1 & .9 & .7 \\ .9 & 1 & .4 \\ .7 & .4 & 1 \end{pmatrix}$.

En este ejemplo, $\mathbf{L}_{3 \times 1} = \begin{pmatrix} l_{11} \\ l_{12} \\ l_{13} \end{pmatrix}$ y la estructura del modelo $\Sigma = \mathbf{L}\mathbf{L}' + \Psi$ define las ecuaciones:

$$\begin{aligned} 1 &= l_{11}^2 & 0.90 &= l_{11}l_{21} & 0.70 &= l_{11}l_{31} \\ & & 1 &= l_{21}^2 + \psi_2 & 0.40 &= l_{21}l_{31} \\ & & & & 1 &= l_{31}^2 + \psi_3 \end{aligned}$$

Resolviendo el sistema de ecuaciones se obtiene $l_{11} = \pm 1.255$. Pero resulta que

$$\text{cov}(X_1, f_1) = \text{cor}(X_1, f_1) = l_{11},$$

por lo que l_{11} no debería ser mayor que 1. Además también se obtiene que $\psi_1 = 1 - l_{11}^2 = -0.575$, pero es una varianza, por lo que tampoco es válido.

□

Problema 2: Ambigüedad en la solución I

- Para $m > 1$ la solución es única, salvo por rotaciones ortogonales. Si \mathbf{T} es una matriz ortogonal, entonces cumple que $\mathbf{T}\mathbf{T}' = \mathbf{T}'\mathbf{T} = \mathbf{I}$.
- Entonces

$$\begin{aligned}\mathbf{x} &= \boldsymbol{\mu} + \mathbf{L}\mathbf{f} + \boldsymbol{\epsilon} \\ &= \boldsymbol{\mu} + \mathbf{L}\mathbf{T}\mathbf{T}'\mathbf{f} + \boldsymbol{\epsilon} \\ &= \boldsymbol{\mu} + (\mathbf{L}\mathbf{T})(\mathbf{T}'\mathbf{f}) + \boldsymbol{\epsilon} \\ &= \boldsymbol{\mu} + \mathbf{L}^*\mathbf{f}^* + \boldsymbol{\epsilon}\end{aligned}$$

Entonces, como $\mathbf{E}[\mathbf{f}^*] = \mathbf{T}'\mathbf{E}[\mathbf{f}] = \mathbf{0}$ y $\text{Var}(\mathbf{f}^*) = \mathbf{T}'\text{Var}(\mathbf{f})\mathbf{T} = \mathbf{T}'\mathbf{T} = \mathbf{I}$, no es posible distinguir entre \mathbf{f}^* y \mathbf{f} , aun cuando las cargas de la matriz \mathbf{L}^* son diferentes de las de \mathbf{L} y ambas generan a $\boldsymbol{\Sigma}$:

$$\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi} = \mathbf{L}^*\mathbf{L}^{*'} + \boldsymbol{\Psi}$$

- Para intentar resolver la ambigüedad, usualmente, se imponen restricciones (puramente matemáticas) para encontrar \mathbf{L} y $\boldsymbol{\Psi}$.

Problema 2: Ambigüedad en la solución II

- Una vez resuelto el problema, la matriz de cargas L se rota multiplicando por una matriz ortogonal con la rotación que facilite la interpretación de las cargas. **Este procedimiento fue muy criticado porque parece que permite a los investigadores imponer sobre los datos cualquier tipo de solución que estuvieran buscando.** Sin embargo, la rotación no altera la estructura de la solución sino sólo cómo se describe la solución.
- La rotación no cambia las propiedades matemáticas subyacentes.
- Posteriormente se identifican los factores y se calculan los *scores de los factores*.

Métodos de estimación

Métodos de estimación

Se considerarán tres métodos (aunque los dos primeros en realidad son variaciones del mismo):

- 1 Método de componentes principales
- 2 Método de factor principal
- 3 Método de máxima verosimilitud. Este último supuesto requiere el supuesto de normalidad para \mathbf{x} , \mathbf{f} y ϵ .

En realidad los dos primeros métodos son básicamente el mismo, y no corresponden *propia*mente a un análisis de factores verdadero. El método de máxima verosimilitud es el modelo formal.

Hay otros métodos de estimación que no consideraremos aquí. Cada uno de estos métodos busca minimizar algún criterio que mide la discrepancia entre \mathbf{S} (o \mathbf{R}) y el estimado $\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\mathbf{\Psi}}$ para encontrar los factores.

- Factor principal iterado
- Mínimos cuadrados no ponderados
- Mínimos cuadrados generalizados

Método de componentes principales I

- Si \mathbf{S} es la matriz de covarianzas muestral, sabemos que la podemos escribir de manera espectral como:

$$\mathbf{S} = \sum_{i=1}^p \hat{\lambda}_i \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i'$$

- Entonces, podemos construir la matriz de cargas considerando los primeros $m < p$ vectores propios:

$$\hat{\mathbf{L}}_m = \left[\sqrt{\hat{\lambda}_1} \hat{\mathbf{e}}_1 | \sqrt{\hat{\lambda}_2} \hat{\mathbf{e}}_2 | \cdots | \sqrt{\hat{\lambda}_m} \hat{\mathbf{e}}_m \right]$$

y podemos definir la matriz residual como

$$\hat{\Psi} = \text{diag}(\hat{\psi}_i)$$

donde $\hat{\psi}_i = s_{ii} - \sum_{j=1}^m \hat{l}_{ij}^2$

- De este modo, con $m < p$, podemos escribir:

$$\hat{\Sigma}_m = \hat{\mathbf{L}}_m \hat{\mathbf{L}}_m' + \hat{\Psi}_m$$

Selección del número de factores m

Para seleccionar m :

- Por consideraciones *a priori*: teoría disponible o investigación.
- Por el tamaño de los valores propios, como en CP. Se puede probar analíticamente que la suma de los cuadrados de las entradas de la matriz residual $\mathbf{S} - \hat{\Sigma}_m = \mathbf{S} - (\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\Psi}) \leq \sum_{i=m+1}^p \hat{\lambda}_i^2$. (tarea).
- Al igual que en componentes principales, la contribución de los primeros factores a la explicación de la varianza se mide como la proporción de la varianza total debida al factor j :

$$\text{La proporción de la varianza total del factor } j = \begin{cases} \frac{\hat{\lambda}_j}{\text{tr}(\mathbf{S})} & \text{para } \mathbf{S} \\ \frac{\hat{\lambda}_j}{p} & \text{para } \mathbf{R} \end{cases}$$

- En algunas implementaciones (SAS, SPSS), se usa el criterio de que m sea el número de $\hat{\lambda}_i > 1$.

Ejemplo I

Retomando el primer ejemplo dado, podemos tomar un enfoque incremental, para ver el comportamiento de la matriz residual. El siguiente código calcula las aproximaciones tomando $i = 1, 2, 3, 4$ eigenvectores de Σ :

```
Sigma

      [,1] [,2] [,3] [,4]
[1,]   19   30    2   12
[2,]   30   57    5   23
[3,]    2    5   38   47
[4,]   12   23   47   68

eigenS <- eigen(Sigma)
eigenS$values #varianzas de las CP

[1] 116.269155  61.415930   2.516520   1.798395

M <- list(NULL) #guarda las aproximaciones a Sigma usando i vectores propios
L <- NULL
for(i in 1:4){
  L <- cbind(L, eigen(Sigma)$vectors[,i]*sqrt(eigen(Sigma)$values[i])) #matriz de cargas con i componentes
  Psihat <- diag(diag(Sigma)-diag((L %*% t(L)))) # matriz residual estimada
  Sigmahat = L %*% t(L) + Psihat # matriz estimada
  M[[i]] <- list(Sigmahat = Sigmahat,
                 Res = Sigma - Sigmahat,
                 SumRes2 = sum((Sigma - Sigmahat)^2), #suma los cuadrados de las entradas de la matriz residual
                 proporcion = sum(eigenS$values[1:i])/sum(diag(Sigma)),
                 sumalambdas_colita = sum(eigenS$values[(i+1):4]^2) )
}
```

Con $m = 1$ se obtiene las siguientes matrices y sumas de cuadrados de la matriz residual (Res2). La matriz residual aún es muy grande con un factor:

Ejemplo II

```
M[[1]]

$Sigmahat
      [,1]      [,2]      [,3]      [,4]
[1,] 19.00000 12.06810 12.88288 19.89593
[2,] 12.06810 57.00000 24.13594 37.27482
[3,] 12.88288 24.13594 38.00000 39.79145
[4,] 19.89593 37.27482 39.79145 68.00000

$Res
      [,1]      [,2]      [,3]      [,4]
[1,]  0.000000 17.93190 -10.882878 -7.895926
[2,] 17.931905  0.00000 -19.135945 -14.274820
[3,] -10.882878 -19.13594  0.000000  7.208554
[4,] -7.895926 -14.27482  7.208554  0.000000

$SumRes2
[1] 2248.508

$proporcion
[1] 0.6388415

$sumalambdas_colita
[1] 3781.484
```

Considerando ahora $m = 2$, vemos que la componente de los factores comunes reproduce mejor a la matriz original y se reduce de manera considerable la suma de cuadrados de la matriz residual:

Ejemplo III

```
M[[2]]

$Sigmahat
      [,1]      [,2]      [,3]      [,4]
[1,] 19.000000 31.045165  1.979033 11.99056
[2,] 31.045165 57.000000  4.745632 23.21671
[3,]  1.979033  4.745632 38.000000 47.86895
[4,] 11.990565 23.216713 47.868953 68.00000

$Res
      [,1]      [,2]      [,3]      [,4]
[1,]  0.000000000 -1.0451653  0.0209671  0.009435029
[2,] -1.045165280  0.0000000  0.2543683 -0.216713279
[3,]  0.020967097  0.2543683  0.0000000 -0.868952738
[4,]  0.009435029 -0.2167133 -0.8689527  0.000000000

$SumRes2
[1] 3.919292

$proporcion
[1] 0.9762917

$sumalambdas_colita
[1] 9.567097
```

Considerando ahora $m = 3$:

Ejemplo IV

```
M[[3]]

$Sigmahat
      [,1]      [,2]      [,3]      [,4]
[1,] 19.000000 29.984027  1.631913 12.29084
[2,] 29.984027 57.000000  4.955539 23.03513
[3,]  1.631913  4.955539 38.000000 47.80955
[4,] 12.290845 23.035131 47.809554 68.00000

$Res
      [,1]      [,2]      [,3]      [,4]
[1,]  0.00000000  0.01597343  0.36808729 -0.29084467
[2,]  0.01597343  0.00000000  0.04446134 -0.03513119
[3,]  0.36808729  0.04446134  0.00000000 -0.80955352
[4,] -0.29084467 -0.03513119 -0.80955352  0.00000000

$SumRes2
[1] 1.757844

$proporcion
[1] 0.9901187

$sumalambdas_colita
[1] 3.234226
```

Para $m = 4$ se vió que LL' reproduce toda la matriz original Σ . Los errores mostrados son errores numéricos

Ejemplo V

```
M[[4]]

$Sigmahat
      [,1] [,2] [,3] [,4]
[1,]   19   30   2   12
[2,]   30   57   5   23
[3,]    2    5  38   47
[4,]   12   23   47   68

$Res
      [,1]      [,2]      [,3]      [,4]
[1,]  0.000000e+00  3.197442e-14 -3.552714e-15  1.065814e-14
[2,]  3.197442e-14  0.000000e+00 -1.065814e-14  2.486900e-14
[3,] -3.552714e-15 -1.065814e-14  0.000000e+00  8.526513e-14
[4,]  1.065814e-14  2.486900e-14  8.526513e-14  0.000000e+00

$SumRes2
[1] 1.830157e-26

$proporcion
[1] 1

$sumalambdas_colita
[1] NA
```

Método de factor principal I

- El análisis de factor de componente principal sustituye a \mathbf{S} por la matriz de correlaciones \mathbf{R} . Esto es equivalente a la estandarización de las variables originales.
- La estandarización es útil en el caso en que las variables tengan diferentes unidades y escalas, y evita el problema de que las variables con alta variabilidad dominen las cargas de los factores.

Ejemplo de preferencias de consumo I

Considérese la matriz para 5 variables obtenidas en un estudio de preferencias de consumo. Las variables consideradas son:

- Sabor
- Buena compra por el dinero pagado
- Sabor
- Adecuado como tentempie
- Provee mucha energía.

La matriz de correlación es la siguiente:

```
R <- matrix(c(1, 0.02, 0.96, 0.42, 0.01,
              0.02, 1, 0.13, 0.71, 0.85,
              0.96, 0.13, 1, 0.5, 0.11,
              0.42, 0.71, 0.5, 1, 0.79,
              0.01, 0.85, 0.11, 0.79, 1), ncol=5)

eigen(R)

eigen() decomposition
$values
[1] 2.85309042 1.80633245 0.20449022 0.10240947 0.03367744

$vectors
      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] -0.3314539  0.60721643  0.09848524  0.1386643  0.701783012
[2,] -0.4601593 -0.39003172  0.74256408 -0.2821170  0.071674637
[3,] -0.3820572  0.55650828  0.16840896  0.1170037 -0.708716714
[4,] -0.5559769 -0.07806457 -0.60158211 -0.5682357  0.001656352
[5,] -0.4725608 -0.40418799 -0.22053713  0.7513990  0.009012569
```

Ejemplo de preferencias de consumo II

Usaremos la descomposición en componentes principales para obtener los factores

```
L <- NULL
for(i in 1:3) L <- cbind(L, round(eigen(R)$vectors[,i]*sqrt(eigen(R)$values[i]), 2))
L

      [,1] [,2] [,3]
[1,] -0.56  0.82  0.04
[2,] -0.78 -0.52  0.34
[3,] -0.65  0.75  0.08
[4,] -0.94 -0.10 -0.27
[5,] -0.80 -0.54 -0.10
```

- ¿Cuál es la matriz de cargas?
- ¿Cuál es la proporción de la varianza explicada por el primer, los dos primeros y los tres primeros factores?
- ¿Cuáles son las comunales?
- ¿Cuáles son las unicidades o factores específicos?
- Calcular la varianza residual con dos factores.

Método de máxima verosimilitud I

- Para poder estimar parámetros de máxima verosimilitud se requiere imponer distribuciones de probabilidad sobre las variables consideradas aleatorias en el modelo.
- En el AF se supone que $\mathbf{f} \sim \mathcal{N}_m(\mathbf{0}, \mathbf{I})$ y $\epsilon \sim \mathcal{N}_p(\mathbf{0}, \Psi)$. Entonces $\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \mathbf{L}\mathbf{L}' + \Psi)$.
- Para evitar la ambigüedad del problema (que está determinado salvo una rotación ortogonal), se impone una condición de unicidad, que es una restricción puramente matemática. Esta condición varía según la implementación computacional, y se han considerado principalmente dos criterios:
 - ① Considerar la restricción: $\mathbf{L}'\mathbf{L} = \mathbf{\Delta}$, donde $\mathbf{\Delta}$ es diagonal. Esta condición establece que los factores, además de ser independientes, producirán los efectos lo más distinto posible en las variables.
 - ② Considerar la restricción: $\mathbf{L}'\Psi^{-1}\mathbf{L} = \mathbf{\Delta}$ diagonal. En esta normalización, los efectos de los factores sobre las variables, ponderados por las varianzas de los errores, se suponen independientes.
- Incorporando la restricción correspondiente, la solución de máxima verosimilitud se obtiene por maximización numérica.

Inferencia sobre m en el caso normal I

- Una prueba asintótica sobre el tamaño de la muestra para m considera como hipótesis a contrastar:

$$H_0 : \Sigma = \mathbf{L}\mathbf{L}' + \Psi \text{ vs. } H_1 : \Sigma > 0 \text{ es cualquier otra matriz}$$

- La prueba, basada en la razón de verosimilitud (LRT), establece que

$$-2 \log \lambda = n \log \left(\frac{|\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\Psi}|}{|\mathbf{S}_n|} \right) \sim \chi^2_{\frac{1}{2}((p-m)^2 - p - m)}$$

- Sin embargo, para poder aplicar la prueba, los grados de libertad deben ser positivos, así que se impone la restricción adicional sobre m :

$$m < p + \frac{1 - \sqrt{8p + 1}}{2}$$

para poder aplicar la prueba.

Además de las funciones que ya se vieron para componentes principales, se pueden ver estas funciones adicionales:

- La función por default `factanal` realiza la estimación por máxima verosimilitud sobre los datos o sobre la matriz de covarianza. Puede calcular los scores de los factores por diferentes métodos y también puede aplicar varios tipos de rotaciones.
- **psych**: paquete para funciones de psicometría e investigación psicológica. Cuenta con la función `fa` que pide matriz de correlación y el número de factores a considerar. Las funciones de este paquete están explicadas con mucho detalle en el sitio del autor William Revelle, donde además pone a disposición mucha más información y los capítulos de su libro: **Personality Project**. Vale mucho la pena.
- **GPA Factor Rotation**: Montón de funciones para rotar factores

Ejemplo 1: Datos de precios de acciones I

- Los datos consisten en las tasas de rendimiento semanal de 5 acciones: (JP Morgan, Citibank, Wells Fargo, Royal Dutch Shell y ExxonMobil) listados en NYSE para el periodo de enero 2004 a diciembre 2005.
- Aplicaremos la función `factanal`, sin verificar los supuestos de normalidad y sin usar todas sus características, para obtener las estimaciones básicas.
- Noten que para aplicar esta función, se requiere pasar de antemano el número de factores a considerar. Como esta función se basa en máxima verosimilitud sólo aceptará el número de factores que permitan hacer la prueba de hipótesis sobre la cola de los eigenvalores.
- En este ejemplo, ¿qué interpretación darían a los coeficientes de los factores? Adicionalmente, noten que la varianza explicada no alcanza el 60 %.

Ejemplo 1: Datos de precios de acciones II

```
stock <- read.table("../data/T8-4.DAT",header = F)
names(stock) <- c("jpmorgan","citibank","wellsfargo","royalDutch","Exxon")
head(stock,4)

  jpmorgan  citibank wellsfargo royalDutch  Exxon
1 0.000000  0.000000   0.000000   0.039473 0.000000
2 0.027027 -0.044855  -0.003030  -0.014466 0.043478
3 0.122807  0.060773   0.088146   0.086238 0.078124
4 0.057031  0.029948   0.066808   0.013513 0.019512
```

factanal(x = stock,factors=3,rotation = "none") *#arroja error por la restricción sobre grados de libertad*

Error in factanal(x = stock, factors = 3, rotation = "none"): 3 factors are too many for 5 variables

factanal(x = stock,factors=2,rotation = "none")

Call:
factanal(x = stock, factors = 2, rotation = "none")

Uniquenesses:

jpmorgan	citibank	wellsfargo	royalDutch	Exxon
0.497	0.252	0.474	0.610	0.176

Loadings:

	Factor1	Factor2
jpmorgan	0.683	0.192
citibank	0.692	0.519
wellsfargo	0.680	0.251
royalDutch	0.621	
Exxon	0.794	-0.439

	Factor1	Factor2
SS loadings	2.424	0.567
Proportion Var	0.485	0.113
Cumulative Var	0.485	0.598

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 0.58 on 1 degree of freedom.
The p-value is 0.448

Interpretación

Rotación de factores I

- Como hemos visto, las cargas de los factores originales pueden no ser fáciles de interpretar y usualmente se aplican rotaciones ortogonales para lograr una estructura más simple.
- Cuando $m = 2$,
 - 1 los factores comunes no correlacionados se consideran como vectores unitarios sobre ejes perpendiculares.
 - 2 Se grafican como puntos las p pares de cargas $(\hat{l}_{i1}, \hat{l}_{i2})$, donde cada par corresponde a una variable
 - 3 Los ejes se pueden rotar considerando un ángulo ϕ generando nuevas cargas \hat{l}_{ij}^* determinados por la relación:

$$\hat{\mathbf{L}}^* = \hat{\mathbf{L}}\mathbf{T}$$

donde $\mathbf{T} = \begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix}$, o $\mathbf{T} = \begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix}$, según se considere una rotación en sentido destrógiro o levógiro respectivamente.

- Cuando $m > 2$,
 - 1 En este caso no es posible visualizar fácilmente, entonces la magnitud de las cargas rotadas debe revisarse para tratar de interpretar los datos originales. En este caso, la rotación \mathbf{T} se puede elegir de varios modos:

Rotación de factores II

- ★ **varimax**: Sea $\tilde{l}_{ij}^* = \frac{\hat{l}_{ij}}{\hat{h}_i}$ el coeficiente rotado escalado por la raíz cuadrada de las communalidades. La rotación varimax es la que maximiza la suma de la varianzas de los cuadrados de las cargas escaladas para el j -ésimo factor:

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

Maximizar V corresponde a dispersar los cuadrados de las cargas en cada factor tanto como sea posible. Lo que se espera encontrar grupos de coeficientes grandes y otros muy pequeños en cualquiera de las columnas de la matriz rotada $\hat{\mathbf{L}}^*$.

- ★ **promax**: (Hendrickson y White (1964)). Rotación oblicua que se aplica posterior a aplicar varimax y eleva esa solución a alguna potencia, al estilo Box-Cox.
- ★ **quartimax**: (Carroll 1953), busca es forzar a una variable dada a tener una alta correlación con un factor y muy poco con los otros factores. No es tan popular como varimax.
- ★ **oblimin**: (Jennrich y Sampson 1966) otra rotación oblicua que intenta encontrar estructuras simples usando como parámetro de control la correlación entre los factores.
- ★ **quartimin**
- ★ **simplimax**
- ★ ... (muchas implememtadas en el paquete `GPArotation`, incluyendo rotaciones no ortogonales, oblicuas).

Ejemplo 1: matriz de Evaluaciones de exámenes I

Se cuenta con una matriz de correlaciones de calificaciones en 6 materias para 220 estudiantes. Las materias son: Galés, Inglés, Historia, Aritmética, Álgebra y Geometría.

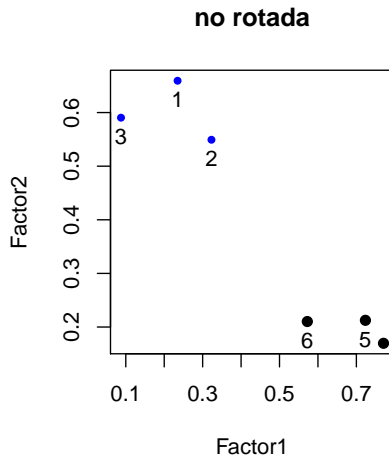
```
library(psych)
R1 <- matrix(c(1, 0.439, 0.410, 0.288, 0.329, 0.248,
               0.439, 1, 0.351, 0.354, 0.320, 0.329,
               0.410, 0.351, 1, 0.164, 0.190, 0.181,
               0.288, 0.354, 0.164, 1, 0.595, 0.470,
               0.329, 0.320, 0.190, 0.595, 1, 0.464,
               0.248, 0.329, 0.181, 0.470, 0.464, 1), ncol = 6)

rownames(R1) <- c("Galés", "Inglés", "Historia", "Aritmética", "Álgebra", "Geometría")
R1

      [,1] [,2] [,3] [,4] [,5] [,6]
Galés  1.000 0.439 0.410 0.288 0.329 0.248
Inglés  0.439 1.000 0.351 0.354 0.320 0.329
Historia 0.410 0.351 1.000 0.164 0.190 0.181
Aritmética 0.288 0.354 0.164 1.000 0.595 0.470
Álgebra  0.329 0.320 0.190 0.595 1.000 0.464
Geometría 0.248 0.329 0.181 0.470 0.464 1.000

m2 <- factanal(covmat=R1,factors=2)
par(pty="s")
fa.plot(m2,title="no rotada") #función del paquete psych
```

Ejemplo 1: matriz de Evaluaciones de exámenes II



Ejemplo 1: matriz de Evaluaciones de exámenes III

```
factor.rotate(m2,angle=-20,plot=F) #Rotación "manual", función del paquete psych
```

	[,1]	[,2]
Galés	-0.004921359	0.6998565
Inglés	0.115559306	0.6265582
Historia	-0.119711203	0.5847603
Aritmética	0.666113204	0.4230223
Álgebra	0.607178649	0.4471752
Geometría	0.465919362	0.3933351

Ejemplo 2: Datos de precios de acciones I

Consideremos de nuevo el ejemplo de los rendimientos de las acciones. Consideremos la rotación `varimax`, la rotación `quartimax`.

```
library(GPArotation)
factanal(x = stock, factors=2, rotation = "varimax")$loadings
```

Loadings:

	Factor1	Factor2
jpmorgan	0.601	0.378
citibank	0.849	0.165
wellsfargo	0.643	0.336
royalDutch	0.365	0.507
Exxon	0.207	0.884

	Factor1	Factor2
SS loadings	1.671	1.321
Proportion Var	0.334	0.264
Cumulative Var	0.334	0.598

```
factanal(x = stock, factors=2, rotation = "quartimax")$loadings
```

Loadings:

	Factor1	Factor2
jpmorgan	0.640	0.307
citibank	0.862	
wellsfargo	0.676	0.261
royalDutch	0.420	0.463
Exxon	0.306	0.855

	Factor1	Factor2
SS loadings	1.880	1.112
Proportion Var	0.376	0.222
Cumulative Var	0.376	0.598

Ejemplo 2: Datos de precios de acciones II

Un ejemplo de una rotación oblicua (ejes no ortogonales) es la rotación promax. Sin embargo, la salida no nos da las correlaciones entre los factores.

```
factanal(x = stock,factors=2, rotation = "promax")$loadings
```

Loadings:

	Factor1	Factor2
jpmorgan	0.576	0.175
citibank	1.011	-0.232
wellsfargo	0.653	
royalDutch	0.202	0.466
Exxon	-0.202	1.037

	Factor1	Factor2
SS loadings	1.863	1.387
Proportion Var	0.373	0.277
Cumulative Var	0.373	0.650

Una alternativa es usar la función `Promax` del paquete `psych` para poder obtener las correlaciones entre los factores oblicuos:

Ejemplo 2: Datos de precios de acciones III

```
library(psych)
Promax(factanal(x = stock, factors=2)$loadings, m=2)
```

Call: NULL

Standardized loadings (pattern matrix) based upon correlation matrix

	Factor1	Factor2	h2	u2
jpmorgan	0.56	0.24	0.50	0.50
citibank	0.89	-0.07	0.75	0.25
wellsfargo	0.62	0.18	0.53	0.47
royalDutch	0.26	0.45	0.39	0.61
Exxon	-0.01	0.91	0.82	0.18

	Factor1	Factor2
SS loadings	1.71	1.28
Proportion Var	0.34	0.26
Cumulative Var	0.34	0.60
Proportion Explained	0.57	0.43
Cumulative Proportion	0.57	1.00

	Factor1	Factor2
Factor1	1.00	0.48
Factor2	0.48	1.00

Scores de los factores

- Los valores estimados de los factores comunes, se llaman *scores de los factores*. Usualmente se utilizan para realizar diagnósticos o como variables de entrada para diferentes tipos de análisis.
- Los scores de los factores son estimados de los valores que pueden tomar los vectores no observables \mathbf{f}_j para $j = 1, \dots, n$.
- La estimación de los scores es problemática porque junto con los estimados de ϵ_j sobrepasan los valores estimados de \mathbf{x}_j .
- Hay dos enfoques para resolver este problema de estimación:
 - ① Método de mínimos cuadrados ponderados
 - ② Método de regresión
- Ambos métodos tienen las siguientes características:
 - ▶ Tratan a las matrices de cargas $\hat{\mathbf{L}}$ y de varianzas específicas $\hat{\Psi}$ estimadas como si fueran los verdaderos valores.
 - ▶ Involucran transformaciones lineales de los datos originales, posiblemente centrados y estandarizados. En ambos casos se pueden usar las cargas rotadas o las no rotadas, no hay diferencia entre los métodos.

Estimación de scores de los factores: Método de Bartlett I

Mínimos cuadrados ponderados

- Si se supone que μ , \mathbf{L} y Ψ conocidas, entonces para modelo de factores

$$\mathbf{x} - \mu = \mathbf{L}\mathbf{f} + \epsilon$$

se puede interpretar como si fuera un modelo de regresión. Como los residuales tienen distribución $\epsilon \sim \mathcal{N}_p(\mathbf{0}, \Psi)$ donde la matriz de covarianzas es diagonal pero no tiene valores constantes, entonces se puede usar mínimos cuadrados ponderados.

- Sea $RSS = \sum_{i=1}^p \epsilon_i^2 / \psi_i = \epsilon' \Psi^{-1} \epsilon = (\mathbf{x} - \mu - \mathbf{L}\mathbf{f})' \Psi^{-1} (\mathbf{x} - \mu - \mathbf{L}\mathbf{f})$
- Podemos minimizar RSS como función de \mathbf{f} y obtenemos por mínimos cuadrados ponderados:

$$\hat{\mathbf{f}} = (\mathbf{L}' \Psi^{-1} \mathbf{L})^{-1} \mathbf{L}' \Psi^{-1} (\mathbf{x} - \mu)$$

Tomando los valores de $\hat{\mathbf{L}}, \hat{\Psi}, \hat{\mu} = \bar{\mathbf{x}}$ como si fueran los verdaderos valores, entonces obtenemos los estimadores de los scores de los factores:

$$\hat{\mathbf{f}}_j = (\hat{\mathbf{L}}' \hat{\Psi}^{-1} \hat{\mathbf{L}})^{-1} \hat{\mathbf{L}}' \hat{\Psi}^{-1} (\mathbf{x}_j - \bar{\mathbf{x}})$$

Estimación de scores de los factores: Método de Bartlett II

Mínimos cuadrados ponderados

- Si se estiman \mathbf{L} y $\mathbf{\Psi}$ por máxima verosimilitud, entonces además se cumple la restricción $\hat{\mathbf{L}}' \hat{\mathbf{\Psi}}^{-1} \hat{\mathbf{L}} = \hat{\mathbf{\Delta}}$ y por lo tanto:

$$\hat{\mathbf{f}}_j = \hat{\mathbf{\Delta}}^{-1} \hat{\mathbf{L}}' \hat{\mathbf{\Psi}}^{-1} (\mathbf{x}_j - \bar{\mathbf{x}})$$

- Si la estimación es por componentes principales:

$$\hat{\mathbf{f}}_j = (\hat{\mathbf{L}}' \hat{\mathbf{L}})^{-1} \hat{\mathbf{L}}' \hat{\mathbf{\Psi}}^{-1} (\mathbf{x}_j - \bar{\mathbf{x}})$$

Este último estimador coincide con los scores de componentes principales.

Ejemplo I

Aplicando de nuevo a la matriz de rendimientos. Los scores estimados corresponden a pares de puntos de los dos factores para cada uno de los datos

```
par(pty="s")  
m1 <- factanal(x=stock, factors=2, scores="Bartlett")  
m2 <- fa(r=stock, nfactors=2, scores = "Bartlett", rotate="varimax")  
plot(m1$scores, pch=16, cex=0.5)
```

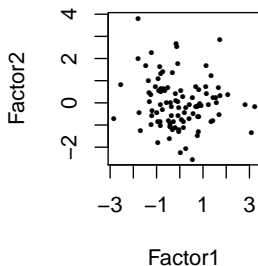
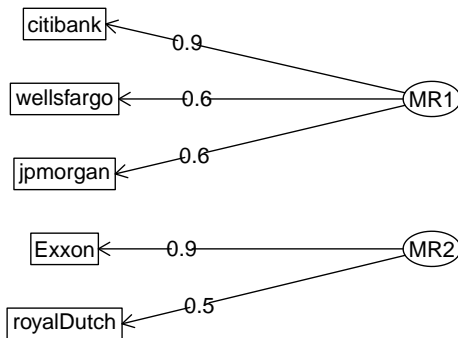


Diagrama de relaciones I

diagram (m2)

Factor Analysis

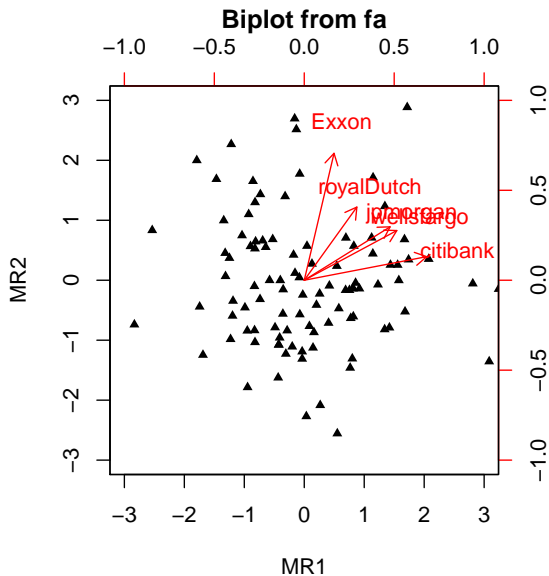


Biplot I

Una gráfica biplot muestra scores y variables en las direcciones de los factores definidos:

`biplot` (m2)

Biplot II



Estimación de scores de los factores: Método de Regresión I

- Sabemos que $\mathbf{x} - \boldsymbol{\mu} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi})$ y $\mathbf{f} \sim \mathcal{N}_m(\mathbf{0}, \mathbf{I}_m)$, y además, $\text{cov}((\mathbf{x}), \mathbf{f}) = \mathbf{L}$.
- Entonces la distribución de $\begin{pmatrix} \mathbf{x} - \boldsymbol{\mu} \\ \mathbf{f} \end{pmatrix} \sim \mathcal{N}_{p+m}\left(\mathbf{0}, \begin{pmatrix} \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi} & \mathbf{L} \\ \mathbf{L}' & \mathbf{I}_m \end{pmatrix}\right)$
- De acuerdo a lo anterior, podemos determinar la regresión

$$\begin{aligned} E[\mathbf{f}|\mathbf{x}] &= \mathbf{L}'(\mathbf{L}\mathbf{L}' + \boldsymbol{\Psi})^{-1}(\mathbf{x} - \boldsymbol{\mu}) \\ \text{Var}(\mathbf{f}|\mathbf{x}) &= \mathbf{I} - \mathbf{L}'(\mathbf{L}\mathbf{L}' + \boldsymbol{\Psi})^{-1}\mathbf{L} \end{aligned}$$

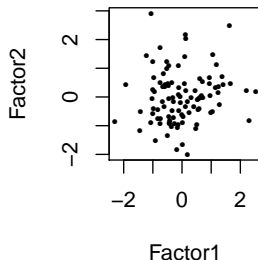
- $\mathbf{L}'(\mathbf{L}\mathbf{L}' + \boldsymbol{\Psi})^{-1}$ son los coeficientes de la regresión entre \mathbf{f} y \mathbf{x} .
- Sustituyendo por los valores estimados obtenemos los scores de los factores estimados por regresión:

$$\hat{\mathbf{f}}_j = \hat{\mathbf{L}}'(\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}})^{-1}(\mathbf{x}_j - \bar{\mathbf{x}})$$

Ejemplo I

Aplicando de nuevo a la matriz de rendimientos. Los scores estimados corresponden a pares de puntos de los dos factores para cada uno de los datos

```
par(pty="s")  
plot(factanal(x=stock, factors=2, scores="regression")$scores, pch=16, cex=0.5)
```



Estrategia de aplicación de AF

- En el análisis de factores, se requiere tomar varias decisiones, en orden de importancia:
 - ▶ Cuántos factores m se van a considerar.
 - ▶ Tipo de método de solución
 - ▶ Tipo de rotación a usar
- La decisión se puede basar en una mezcla de los siguientes criterios:
 - ▶ La proporción de la varianza muestral explicada.
 - ▶ conocimiento del área de aplicación.
 - ▶ la razonabilidad de los resultados.
- Para tratar de seguir una metodología estricta, se sugiere aplicar las siguientes estrategias.

Estrategia de aplicación de AF

- ➊ Realizar un análisis de factores basado en componentes principales.
 - ▶ Buscar observaciones sospechosas graficando los scores de los factores.
 - ▶ Intentar la rotación varimax.
- ➋ Desarrollar un análisis de factores basado en máxima verosimilitud, incluyendo la rotación varimax
- ➌ Comparar las soluciones obtenidas por los dos métodos de los pasos anteriores.
 - ▶ ¿Las cargas se agrupan de la misma manera?
 - ▶ Graficar los scores obtenidos por los dos métodos uno contra otro.
- ➍ Repetir los pasos anteriores aumentando el número de factores comunes m ¿Estos factores extras contribuyen al entendimiento e interpretación de los resultados?
- ➎ Para conjuntos de datos grandes, partir en segmentos y aplicar el análisis en cada segmento para verificar estabilidad y consistencia.

Diferencias entre AF y CP

- el AF trata de explicar las covarianzas o correlaciones de las variables observadas por medio de unos cuantos factores comunes. **CP trata de explicar la varianza de las variables observadas.**
- Si se incrementa el número de componentes de m a $m + 1$, **las CP no cambian.** AF tiene cambios sustanciales en todos los factores cuando se cambia el número de factores.
- **El cálculo de los scores de CP es directo.** El cálculo de los scores de los factores es mucho más complejo, y hay varios métodos para hacerlo.
- **Usualmente no hay relación entre las CP de la matriz de correlaciones o de covarianzas** En el caso de AF, para máxima verosimilitud el análisis es esencialmente equivalente (excepto en el caso de análisis de factor principal).

Aun con estas diferencias, los resultados de ambos tipos de análisis es frecuentemente muy similar.

Conclusión



Conclusión:

- El análisis factorial no garantiza un descubrimiento satisfactorio de factores significativos latentes.
- Usualmente siempre es bueno iniciar observando y analizando la matriz de correlaciones de un conjunto de datos X . Si hay pocas o ninguna correlación alta, entonces no tiene sentido intentar siquiera FA.