

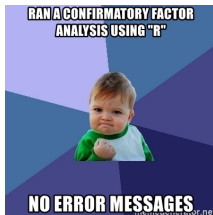
Estadística Aplicada III

Análisis Factorial

Jorge de la Vega Góngora

Departamento de Estadística,
Instituto Tecnológico Autónomo de México

Semana 9

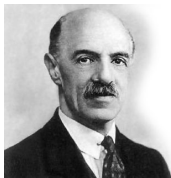


ITAM

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 latentes o *constructos*¹ psicométricos, como la inteligencia, la personalidad, satisfacción, la creatividad, etc.



Spearman



Pearson

- También hay constructos o *variables latentes* en ámbitos como la economía, mercadotecnia, finanzas. Ejemplos son: productividad, calidad de servicio, satisfacción, competencia, eficiencia financiera, etc.

Origen y propósito del Análisis Factorial (AF) II

- 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 pueden estar 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 explica porque en realidad las personas variaban en una característica latente, la inteligencia, y sus scores eran dependientes de su nivel de inteligencia. A este nivel se le llama usualmente el *coeficiente intelectual* (o IQ).
- *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 latente que es el responsable de la alta correlación en ese grupo.*

Origen y propósito del Análisis Factorial (AF) III

El principal propósito del análisis factorial (AF) es *describir* las relaciones de dependencia (a través de las covarianzas) entre variables, en términos de unas cuantas variables explicativas subyacentes *no observables* llamados *factores latentes*.

- El análisis factorial se puede considerar como una extensión del Análisis de componentes principales, en el sentido de que ambos métodos buscan aproximar la matriz de covarianzas Σ .

¹Un constructo es una entidad hipotética que no es fácil definir dentro de una teoría científica. Se sabe que existe, pero cuya definición es difícil o controvertida.

Tipos de Análisis Factorial

- El AF se aplica en dos contextos de análisis de datos:
 - **Análisis exploratorio:**
 - Se utiliza en las fases de desarrollo de un modelo.
 - Cuando no hay hipótesis o teoría y se explora la existencia de un patrón o relaciones de los datos que no son atribuibles al azar, se puede pensar como *aprendizaje no supervisado*.
 - Evalúa variables.
 - **Análisis confirmatorio:**
 - Cuando hay una teoría o hipótesis subyacente y se pretende validar o evaluar la estructura propuesta.
 - En este caso, se pretende hacer una inferencia y se requieren verificar normalidad de los datos.
 - Es similar al concepto de *aprendizaje supervisado*.

Tanto matemática y conceptualmente son diferentes procedimientos. Esto se entenderá mejor en la práctica.

Ejemplos I

- Algunos ejemplos de 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
volatilidad de mercados financieros	incertidumbre, riesgo, utilidad

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

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

Modelo

Modelo básico I

- Si $\mathbf{x} \in \mathbb{R}^p$ de una población con media μ y covarianza Σ , el modelo supone que los datos se generan de acuerdo al siguiente modelo:

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

donde

- $\mathbf{L}_{p \times m}$ es una matriz que contiene **coeficientes de carga**
- $\mathbf{f}_{m \times 1}$ es el vector de **factores comunes**
- $\epsilon_{p \times 1}$ es un vector de errores o **factores específicos** o **idiosincrásicos**
- Usualmente, \mathbf{f} y ϵ 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 (\mathbf{L} y ϵ suman $pm + p$ parámetros, y se tienen p variables), se requiere imponer restricciones complementarias para especificar adecuadamente el problema. Estas restricciones son:

- i. los factores comunes y los específicos se consideran independientes:

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

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

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

- iii. los factores específicos también son independientes entre sí, con

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

Estructura de covarianza subyacente de \mathbf{x} I

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

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

Entonces, tomando esperanza de ambos lados:

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

Así, la varianza se puede descomponer en la variabilidad de los factores comunes ($\mathbf{L}\mathbf{L}'$) y de los factores específicos ($\boldsymbol{\Psi}$).

- También hay que notar la dependencia entre \mathbf{x} y \mathbf{f} . Notando que $\mathbf{x} - \boldsymbol{\mu} = \mathbf{L}\mathbf{f} + \boldsymbol{\epsilon}$:

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

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

Ejemplo 1 I

- Primero notemos que cualquier matriz definida positiva se puede escribir de la forma $\Sigma = \mathbf{L}\mathbf{L}'$ a través de su descomposición espectral, por lo que en ese caso $\Psi = \mathbf{0}$. Por ejemplo:

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

      [,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 1 II

- La misma matriz se puede escribir considerando sólo dos factores y agregando una matriz 'residual':

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

- La matriz $\mathbf{L}_{4 \times 2}$ tiene elementos l_{ij} . Entonces, $\sigma_{1,1}$ se puede escribir, por ejemplo, como:

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

- Entonces podemos ver que podemos aproximarnos a Σ de diferentes modos, no necesariamente se tiene una solución única. Es por esto la necesidad de agregar restricciones adicionales.

Posibles variaciones al modelo

Los supuestos escogidos para especificar el modelo factorial se pueden escoger de otras formas:

- **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.

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

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

Estos casos más complicados no serán cubiertos en este curso, pero pueden considerarse como un buen proyecto final.

- Consideremos la descomposición de la varianza de \mathbf{x} en dos partes:
 - La proporción de la varianza de la i -ésima variable que es contribuida por los factores comunes a todas las variables, se conoce como *la i -ésima comunalidad*:

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

- La proporción de la varianza que se debe a los factores específicos se conoce como *unicidad o varianza específica*
- Así que la varianza de cada variable se puede ver descompuesta como sigue:

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 en la descomposición, porque $\Sigma = \mathbf{L}\mathbf{L}'$ y $\Psi = \mathbf{0}$.

Prueba de Kaiser-Meyer-Olkin (KMO) I

- El análisis factorial no tiene mucho sentido si hay demasiada o muy poca correlación entre las variables. Para evaluar la correlación entre las variables se utiliza la estadística KMO (Kaiser, 1970) que mide la proporción de varianza entre las variables que puede ser considerada varianza común.
- Mientras más alta es la proporción, el valor de la estadística KMO es más alta, y los datos son más adecuados para un análisis factorial.
- La fórmula de la prueba está dada por:

$$KMO = \frac{\sum_j \sum_{j \neq k} r_{jk}^2}{\sum_j \sum_{j \neq k} r_{jk}^2 + \sum_j \sum_{j \neq k} p_{jk}^2}$$

donde

- r_{jk} es la correlación entre variables X_j y X_k
- p_{kj} es la correlación parcial.
- La prueba $KMO \in (0, 1)$ y su interpretación es la siguiente (de acuerdo al propio Kaiser):

Subjective appraisal, based on extensive experience, suggests that these Indices of Factorial Simplicity may be evaluated according to the following table:

in the .90s	marvelous,	
in the .80s	meritorious,	
in the .70s	middling,	(2)
in the .60s	mediocre,	
in the .50s	miserable,	
below .50	unacceptable.	

Prueba de esfericidad de Bartlett

- Otra prueba que puede ser útil para evaluar si un modelo factorial es razonable, es la prueba de esfericidad de Bartlett (1951), que sirve para probar la hipótesis $H_0 : \mathbf{R} = \mathbf{I}$.
- La hipótesis alternativa es la de que las variables no son ortogonales. Si se rechaza la hipótesis nula, entonces podemos decir que las variables tienen cierta correlación, y sería factible un análisis factorial.
- La prueba de esfericidad de Bartlett, bajo el supuesto de normalidad de los datos, está dada por:

$$\chi^2 = -\log(|\mathbf{R}|) \left(n - 1 - \frac{2p + 5}{6} \right) \sim \chi^2_{p(p-1)/2}$$

- Si la condición de normalidad no se cumple, se puede utilizar de manera directa la prueba de KMO.

Posibles problemas con el modelo

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 de manera única.
- El primer problema no tiene una solución general. Entonces no todas las matrices de covarianza pueden factorizarse en factores comunes.
 - El segundo problema es similar al problema de los signos que se da en componentes principales. La solución en este caso será tratar de imponer restricciones adicionales.

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:

$$\begin{aligned} 1 &= l_{11}^2 + \psi_1 & 0.90 &= l_{11}l_{12} & 0.70 &= l_{11}l_{13} \\ & & 1 &= l_{12}^2 + \psi_2 & 0.40 &= l_{12}l_{13} \\ & & & & 1 &= l_{13}^2 + \psi_3 \end{aligned}$$

Resolviendo el sistema de ecuaciones se obtiene $l_{11} = \pm 1.255$ (despejando l_{11} e igualando en las dos últimas ecuaciones del primer renglón, y luego sustituyendo en la última ecuación del segundo renglón). Pero resulta que

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

Problema 1: Soluciones no factibles II

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. Esto se debe a que si \mathbf{T} es una matriz ortogonal, entonces cumple la condición que $\mathbf{T}\mathbf{T}' = \mathbf{T}'\mathbf{T} = \mathbf{I}$.
- Entonces esto se puede reflejar en la siguiente transformación:

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

Entonces, como $E[\mathbf{f}^*] = \mathbf{T}'E[\mathbf{f}] = \mathbf{0}$ 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 Σ :

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

Problema 2: Ambigüedad en la solución II

- Para intentar resolver la ambigüedad, usualmente, se imponen restricciones (puramente matemáticas) para encontrar \mathbf{L} y Ψ .
- Una vez resuelto el problema de estimación, la matriz de cargas \mathbf{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 (*cherry picking*). 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 latentes y se calculan los *scores de los factores*, como en componentes principales.

Estimación

Métodos de estimación I

- 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 *propriamente* 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 $\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\Psi}$ para encontrar los factores. Ejemplos de estos otros métodos son los siguientes:
 - 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*: usando la teoría del contexto disponible o alguna 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 cumple la desigualdad:

$$\text{SumCuadrados}(\mathbf{S} - \hat{\Sigma}_m) = \text{SumCuadrados}(\mathbf{S} - (\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\Psi})) \leq \sum_{i=m+1}^p \hat{\lambda}_i^2$$

Demostración.

La matriz $\mathbf{S} - (\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\Psi})$ tiene ceros en la diagonal, por lo que la suma de cuadrados de sus elementos es menor que la suma de cuadrados de $\mathbf{S} - \hat{\mathbf{L}}\hat{\mathbf{L}}'$, que no tiene ceros en la diagonal.

Además, como vimos $\mathbf{S} = \sum_{i=1}^p \hat{\lambda}_i \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i'$ y $\hat{\mathbf{L}}\hat{\mathbf{L}}' = \sum_{i=1}^m \hat{\lambda}_i \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i'$. Entonces $\mathbf{S} - \hat{\mathbf{L}}\hat{\mathbf{L}}' = \sum_{i=m+1}^p \hat{\lambda}_i \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i'$. Esta última parte se puede escribir de la forma: $\mathbf{P}\mathbf{\Lambda}\mathbf{P}'$. Por último, recordar que $\text{tr}\{\mathbf{A}\mathbf{A}'\} = \sum_{i,j} a_{ij}^2$ y de este modo:

$$\text{tr}\{\mathbf{P}\mathbf{\Lambda}\mathbf{P}'\} = \text{tr}\{\mathbf{\Lambda}\mathbf{\Lambda}\} = \sum_{i=m+1}^p \hat{\lambda}_i^2$$



Selección del número de factores m II

- 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 eigenvalores que cumplen $\hat{\lambda}_i > 1$.

Ejemplo 1 (cont.) 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) )
}
```


Ejemplo 1 (cont.) II

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:

```
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
```

Ejemplo 1 (cont.) III

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:

```
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
```

Ejemplo 1 (cont.) IV

Considerando ahora $m = 3$:

```
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 \mathbf{LL}' reproduce toda la matriz original Σ . Los errores mostrados son errores numéricos

Ejemplo 1 (cont.) 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	2.48690e-14	2.664535e-15	7.105427e-15
[2,]	2.486900e-14	0.00000e+00	-1.243450e-14	0.000000e+00
[3,]	2.664535e-15	-1.24345e-14	0.000000e+00	1.136868e-13
[4,]	7.105427e-15	0.00000e+00	1.136868e-13	0.000000e+00

```
$SumRes2
```

```
[1] 2.751074e-26
```

```
$proporcion
```

```
[1] 1
```

```
$sumalambdas_colita
```

```
[1] NA
```

Ejemplo 2: Preferencias de consumo I

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

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

La matriz de correlación, obtenida de las respuestas en una escala de Likert de 7 puntos es la siguiente:

Ejemplo 2: Preferencias de consumo II

```
# Matriz de correlación dada:
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)

colnames(R) <- row.names(R) <- c("G", "B", "S", "T", "E")
eigen(R) # Descomposición espectral

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

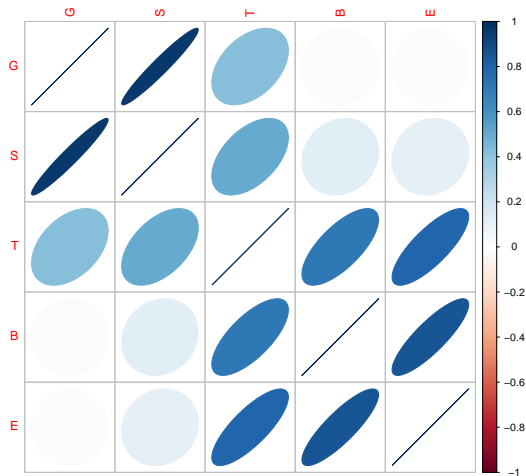
$vectors
      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 0.3314539 -0.60721643 0.09848524 0.1386643 0.701783012
[2,] 0.4601593 0.39003172 0.74256408 -0.2821170 0.071674637
[3,] 0.3820572 -0.55650828 0.16840896 0.1170037 -0.708716714
[4,] 0.5559769 0.07806457 -0.60158211 -0.5682357 0.001656352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

```
library(corrplot)

corrplot 0.94 loaded

corrplot(R, method = "ellipse", order = "hclust")
```

Ejemplo 2: Preferencias de consumo III



Ejemplo 2: Preferencias de consumo IV

Las variables que muestran alta correlación son la pareja (Gusto, Sabor), (Energía, buena compra) y (Energía, tentempié). Parecería que puede haber dos o tres factores asociados al modelo. 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      # 3 factores
```

```
      [,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 para dos factores?

```
L[,1:2]
      [,1] [,2]
[1,] 0.56 -0.82
[2,] 0.78  0.52
[3,] 0.65 -0.75
[4,] 0.94  0.10
[5,] 0.80  0.54
```


Ejemplo 2: Preferencias de consumo V

- ¿Cuál es la proporción de la varianza explicada por el primer, los dos primeros y los tres primeros factores?

```
100*cumsum(eigen(R)$values)/5  
[1] 57.06181 93.18846 97.27826 99.32645 100.00000
```

- ¿Cuáles son las comunidades?

```
diag(L[,1:2] %*% t(L[,1:2]))  
[1] 0.9860 0.8788 0.9850 0.8936 0.9316
```

Valores grandes de las comunidades en el caso de correlación indican que los factores considerados explican una buena proporción de la varianza de cada variable.

- ¿Cuáles son las unicidades o factores específicos?

```
(psihat <- rep(1,5) - diag(L[,1:2] %*% t(L[,1:2])))  
[1] 0.0140 0.1212 0.0150 0.1064 0.0684
```

Ejemplo 2: Preferencias de consumo VI

- Calcular la varianza residual con dos factores.

```
(VarRes <- R-(L[,1:2] %*% t(L[,1:2]) + diag( rep(1,5) - diag(L[,1:2] %*% t(L[,1:2])) ) )
```

	G	B	S	T	E
G	0.0000	0.0096	-0.019	-0.0244	0.0048
B	0.0096	0.0000	0.013	-0.0752	-0.0548
S	-0.0190	0.0130	0.000	-0.0360	-0.0050
T	-0.0244	-0.0752	-0.036	0.0000	-0.0160
E	0.0048	-0.0548	-0.005	-0.0160	0.0000

Método de máxima verosimilitud I

- Máxima verosimilitud requiere distribuciones de probabilidad sobre las variables aleatorias en el modelo.
- En el Análisis Factorial clásico se supone que tanto los factores como las unidades son normales:

$$\mathbf{f} \sim \mathcal{N}_m(\mathbf{0}, \mathbf{I}), \quad \boldsymbol{\epsilon} \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Psi})$$

Entonces se sigue que

$$\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}).$$

- La función de log-verosimilitud que hay que optimizar es:

$$l(\boldsymbol{\mu}, \mathbf{L}, \boldsymbol{\Psi}) = -\frac{np}{2} \log(2\pi) - \frac{n}{2} \det(\mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}) - \left\{ -\frac{1}{2} \text{tr} [(\mathbf{L}\mathbf{L}' + \boldsymbol{\Psi})^{-1} (n\mathbf{S}_n + n(\bar{\mathbf{x}} - \boldsymbol{\mu})(\bar{\mathbf{x}} - \boldsymbol{\mu})')] \right\}$$

- Para evitar ambigüedad en la solución (que está determinada salvo rotaciones ortogonales), se impone alguna condición adicional de unicidad, que es una restricción puramente matemática. La condición varía según la implementación computacional, usualmente alguno de los dos criterios siguientes:

- 1 $\mathbf{L}'\mathbf{L} = \mathbf{\Delta}$, donde $\mathbf{\Delta}$ es diagonal. Esta restricción establece que los factores, además de ser independientes, producirán los efectos lo más distinto posible en las variables.
- 2 *Restricción de unicidad*: $\mathbf{L}'\mathbf{\Psi}^{-1}\mathbf{L} = \mathbf{\Delta}$ diagonal. En esta normalización, se supone a los efectos de los factores sobre las variables, ponderados por las varianzas de los errores, independientes.
- Incorporando alguna de las dos restricciones anteriores, la solución de máxima verosimilitud se obtiene por optimización numérica.

Ejemplo: Preferencias de consumo vía máxima verosimilitud

- Para obtener los estimadores máximo verosímiles se considerará la función básica factorial

Ejemplo: Preferencias de consumo vía máxima verosimilitud II

```
(m1 <- factanal(covmat = R, factors = 2)) # Cálculo con dos factores
```

Call:

```
factanal(factors = 2, covmat = R)
```

Uniquenesses:

	G	B	S	T	E
	0.028	0.237	0.040	0.168	0.052

Loadings:

	Factor1	Factor2
G		0.985
B	0.873	
S	0.131	0.971
T	0.817	0.405
E	0.973	

	Factor1	Factor2
SS loadings	2.396	2.078
Proportion Var	0.479	0.416
Cumulative Var	0.479	0.895

The degrees of freedom for the model is 1 and the fit was 0.0233

```
factanal(covmat = R, factors = 3) # con tres factores ya no me permite hacer la estimación (no puede calcular grados de libertad).
```

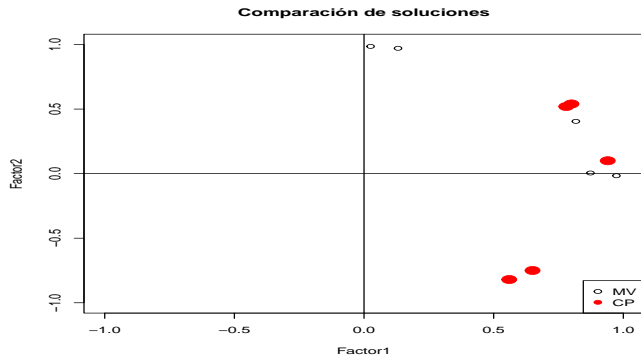
```
Error in factanal(covmat = R, factors = 3): 3 factors are too many for 5 variables
```

Ejemplo: Preferencias de consumo vía máxima verosimilitud III

Podemos comparar los resultados contra los resultados obtenidos con el método de componentes principales (notar la rotación de los puntos):

```
plot(mi$loadings,ylim=c(-1,1),xlim=c(-1,1), main = "Comparación de soluciones")
abline(h=0,v=0)
points(L[,1],L[,2], col="red",pch=19,cex=2)
legend("bottomright",legend=c("MV","CP"),col=c("black","red"),pch=c(1,19))
```

Ejemplo: Preferencias de consumo vía máxima verosimilitud IV



Inferencia sobre m en el caso normal I

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

$H_0 : \Sigma$ es de la forma $\mathbf{LL}' + \Psi$ vs. $H_1 : \Sigma > 0$ es cualquier otra matriz

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

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

- La aproximación a la χ^2 se puede mejorar considerando la corrección de Bartlett, siempre que n y $n - p$ sean grandes:

$$(n - 1 - (2p + 4m + 5)/6) \log \left(\frac{|\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\Psi}|}{|\mathbf{S}_n|} \right) \sim \chi^2_{\frac{1}{2}((p-m)^2 - p - m)}$$

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

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

Ejemplo: preferencias de consumo I

- En el ejemplo 2 sobre preferencias de consumo, se tiene que $|R| = 0.003635$, y $|\hat{\mathbf{L}}_2 \hat{\mathbf{L}}_2' + \hat{\Psi}_m|$ está dado por

```
det(m1$loadings %*% t(m1$loadings) + diag(m1$uniquenesses))  
[1] 0.003720373
```

por lo que $\frac{|\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\Psi}|}{|\mathbf{S}_n|} = 1.023579$ Suponiendo un valor de n (no se tiene en este ejemplo), $n = 100$, se tiene que $(100 - 1 - (2 * 5 + 4 * 2 + 5)/6) * \log(1.023579) = 2.217889$ y

```
pchisq(2.217889, 0.5 * ((5-2)^2 - 5-2), lower.tail = F)  
[1] 0.1364195
```

- Por lo tanto, no se tiene evidencia estadística para rechazar la hipótesis nula, con el supuesto valor $n = 100$.
- El lado derecho de la desigualdad para los grados de libertad es $5 + \frac{1 - \sqrt{41}}{2} = 2.29$, por lo que $m \leq 2$.

Herramientas computacionales

Herramientas de estimación de AF en R I

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

- `factanal` es la función por default. Realiza la estimación por máxima verosimilitud sobre los datos o sobre la matriz de covarianza. Puede calcular los scores de los factores (si se dan los datos originales) 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.
- `GPArotation`: Paquete con montón de funciones para rotar factores.
- `factoextra` permite varios modelos de análisis factorial: *multiple factor analysis*, cuando los datos están organizados por grupos; *hierarchical MFA* cuando además tienen estructura jerárquica; *Factor analysis of mixed data* para datos cualitativos y cuantitativos.

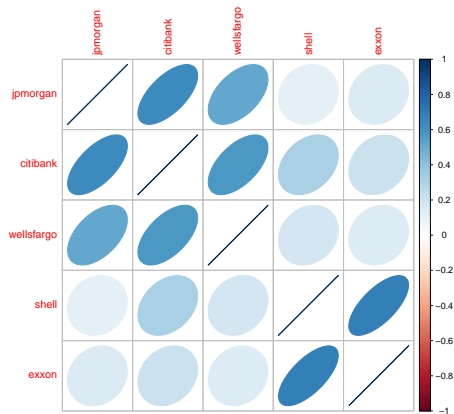
- FactoMineR, desarrollado por franceses, permite hacer análisis exploratorio multivariado.
- EFAtools tiene por objetivo proveer implementaciones flexibles de diferentes algoritmos para el análisis factorial exploratorio.

Ejemplo 3: Datos de precios de acciones I

- Los datos consisten en 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.
- Se utilizará `factanal`, sin verificar los supuestos de normalidad y sin usar todas sus características, para obtener las estimaciones básicas.
- Notar que se requiere tener 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.
- La estructura de correlación no es tan clara.

```
stock <- read.csv("https://raw.githubusercontent.com/fcbarbi/AMSA/master/csv/T0804.csv", header = T)
corrplot(cor(stock), method = "ellipse")
```

Ejemplo 3: Datos de precios de acciones II



Ejemplo 3: Datos de precios de acciones III

- ¿Qué interpretación darían a los coeficientes de los factores?

Ejemplo 3: Datos de precios de acciones IV

```
head(stock, 2) # Cómo se ven los datos

  jpmorgan  citibank wellsfargo      shell      exxon
1 0.0130338 -0.0078431 -0.0031889 -0.0447693 0.0052151
2 0.0084862 0.0166886 -0.0062100 0.0119560 0.0134890

# factanal(x = stock, factors = 3, rotation = "none") #arroja error por la restricción sobre grados de libertad
(maf1 <- factanal(x = stock, factors = 2, rotation = "none"))
```

Call:

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

Uniquenesses:

jpmorgan	citibank	wellsfargo	shell	exxon
0.417	0.275	0.542	0.005	0.530

Loadings:

	Factor1	Factor2
jpmorgan	0.121	0.754
citibank	0.328	0.786
wellsfargo	0.188	0.650
shell	0.997	
exxon	0.685	

	Factor1	Factor2
SS loadings	1.622	1.610
Proportion Var	0.324	0.322
Cumulative Var	0.324	0.646

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 1.97 on 1 degree of freedom.
The p-value is 0.16

Interpretación a través de rotaciones

Rotación de factores, $m = 2$ 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, particularmente en la estimación de máxima verosimilitud.
- 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 X_i
 - 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* (en el sentido de las manecillas del reloj) o *levógiro* respectivamente.

Rotación de factores, $m > 2$ |

- 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.
- 2 La rotación **T** se puede elegir de varios modos:

- **varimax**: Sea $\tilde{l}_{ij}^* = \frac{\hat{l}_{ij}}{\hat{h}_i}$ el coeficiente rotado escalado por la raíz cuadrada de las comunales. 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 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.

Rotación de factores, $m > 2$ II

- **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 implementadas en el paquete `GPArotation`, incluyendo rotaciones no ortogonales, oblicuas).

Ejemplo 4: matriz de Evaluaciones de exámenes I

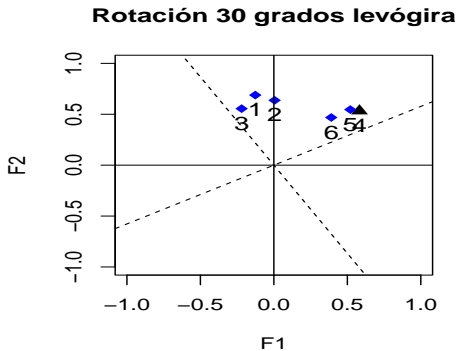
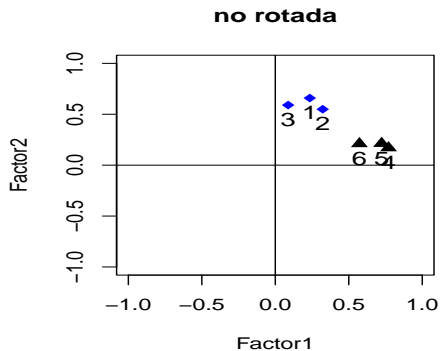
Se cuenta con una matriz de correlaciones de calificaciones en 6 materias para 220 estudiantes. Las materias son: Galés, Inglés, Historia, Aritmética, Álgebra y Geometría.

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

	Galés	Inglés	Historia	Aritmética	Álgebra	Geometría
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", mfrow=c(1,2))
fa.plot(m2, title = "no rotada", pch=16, cex=1.2, ylim = c(-1,1), xlim = c(-1,1)) #función del paquete psych
factor.rotate(m2, angle = -30, plot = T, ylim = c(-1,1), xlim = c(-1,1), cex = 1.2, pch = 16,
              title = "Rotación 30 grados levógira") #Rotación "manual", función del paquete psych
```

Ejemplo 4: matriz de Evaluaciones de exámenes II



Ejemplo 3 : Datos de precios de acciones (cont.) I

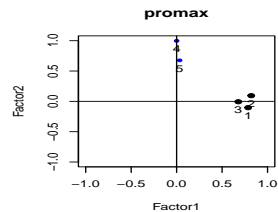
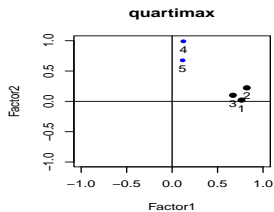
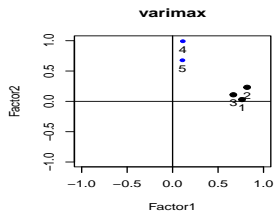
Consideremos de nuevo el ejemplo de los rendimientos de las acciones. Consideremos la rotación varimax, la rotación quartimax.

```
par(mfrow = c(1,3), pty = "s")  
library(GPArotation)
```

```
Adjuntando el paquete: 'GPArotation'  
The following objects are masked from 'package:psych':  
equamax, varimin
```

```
fa.plot(factanal(x = stock, factors = 2, rotation = "varimax"), ylim = c(-1,1), xlim = c(-1,1),  
        title = "varimax")  
fa.plot(factanal(x = stock, factors = 2, rotation = "quartimax"), ylim = c(-1,1), xlim = c(-1,1),  
        title = "quartimax")  
fa.plot(factanal(x = stock, factors = 2, rotation = "promax"), ylim = c(-1,1), xlim = c(-1,1),  
        title = "promax")
```

Ejemplo 3 : Datos de precios de acciones (cont.) II



Cálculo de scores de los factores

- Los valores estimados de los factores comunes, se llaman *puntajes de los factores* (o scores). Usualmente se utilizan para realizar diagnósticos o como variables de entrada para diferentes tipos de análisis.
- Los puntajes 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 , exceden en número los valores estimados de \mathbf{x}_j .
- Se pueden considerar 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, ya que las fórmulas en cada caso son las mismas.

Estimación de puntajes de los factores: Método de Bartlett I

Mínimos cuadrados ponderados

- Se supone que μ , \mathbf{L} y Ψ conocidas, entonces para el 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 el método de *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 puntajes de los factores: Método de Bartlett II

Mínimos cuadrados ponderados

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

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

- Si la estimación es por componentes principales:

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

Este último estimador coincide con los scores de componentes principales.

Ejemplo 4 (cont.) 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")
```

```
Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, : The estimated weights for the factor scores are probably incorrect. Try a different factor score estimation method.
```

```
Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate = rotate, : An ultra-Heywood case was detected. Examine the results carefully
```

```
plot(m1$scores, pch = 16, main = "gráfica de los scores de los factores")  
abline(h = 0, v = 0)
```

Ejemplo 4 (cont.) II

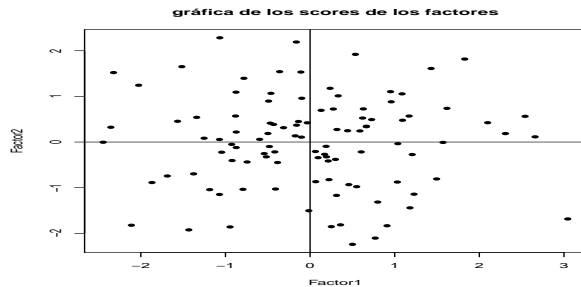
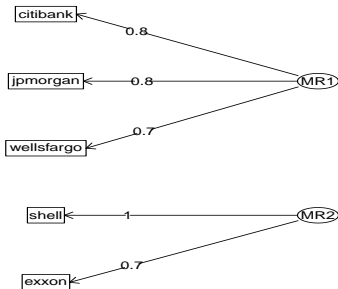


Diagrama de relaciones I

```
diagram(m2)
```

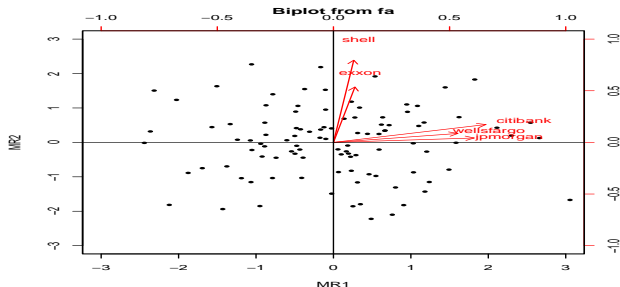
Factor Analysis



Biplot I

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

```
biplot(m2)  
abline(h = 0, v = 0)
```



Estimación de scores de los factores: Método de Regresión I

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

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

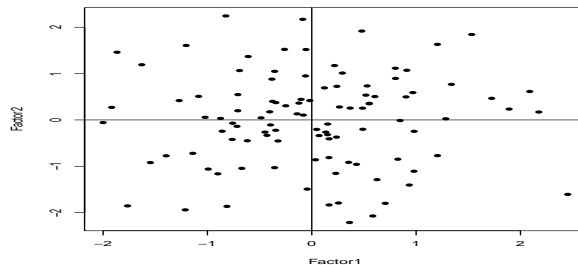
- $\mathbf{L}'(\mathbf{LL}' + \boldsymbol{\Psi})^{-1}$ son los coeficientes de la regresión entre \mathbf{f} y \mathbf{x} .
- Sustituyendo por los valores estimados obtenemos los scores de los factores estimados por regresión:

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

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)  
abline(v=0, h=0)
```



- 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.

Aplicación en la práctica

Estrategia de aplicación de AF

- 1 Realizar un análisis de factores basado en componentes principales.
 - Buscar observaciones sospechosas graficando los scores de los factores.
 - Intentar la rotación varimax.
- 2 Desarrollar un análisis de factores basado en máxima verosimilitud, incluyendo la rotación varimax
- 3 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.
- 4 Repetir los pasos anteriores aumentando el número de factores comunes m ¿Estos factores extras contribuyen al entendimiento e interpretación de los resultados?
- 5 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



- Es importante notar que el Análisis Factorial no garantiza un descubrimiento satisfactorio de factores significativos latentes, o que su interpretación sea sencilla cuando no hay una teoría subyacente.
- 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 el Análisis de Factores.
- En el proceso de estimación puede haber complicaciones para su obtención. Depende mucho de la implementación computacional y de la calidad de los datos.