Modelos de Mercadotecnia

Modelos de Ecuaciones Estructurales: Análisis Factorial

Jorge de la Vega Góngora

Maestría de Mercadotecnia,
Instituto Tecnológico Autónomo de México

Sesión 7

Introducción

Introducción I

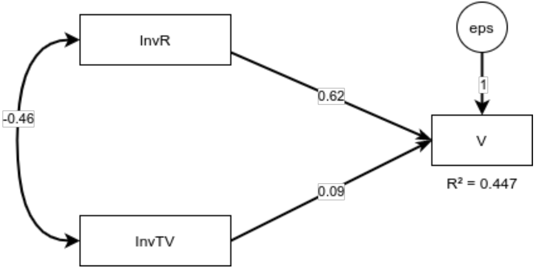
- El tema de modelos de ecuaciones estructurales (SEM por sus siglas en inglés) no es un método estadístico, sino un marco para un número de métodos/técnicas de análisis estadístico multivariado.
- Estos modelos involucran una serie de relaciones entre variables numéricas (observadas o latentes) con la finalidad de descubrir procesos subyacentes que generan las variables.
- Un caso particular que conviene entender primero es el de análisis factorial (FA), que se revisará a continuación como el contexto inicial para SEM. De hecho a veces SEM se conoce como *Análisis Factorial Confirmatorio*.
- SEM es tan general, que incluye FA, Análisis de Correlación canónica (CCA), regresión, análisis de senderos, modelado de crecimiento latente, etc.
- **Desafortunadamente**, SEM es una generalización de muchos modelos lineales que requiere una fuerte notación matricial. Ayuda a entender los fundamentos técnicos, aunque no es necesario para entender en términos generales, las ideas de la modelación vía SEM.

Ejemplo motivacional I

- Para analizar si las ventas (V) de un producto dependen linealmente de la inversión en publicidad por radio ($InvR$) y de la inversión en publicidad por TV ($InvTV$), se puede usar el modelo:

$$V = \beta_0 + \beta_1 InvR + \beta_2 InvTV + \epsilon$$

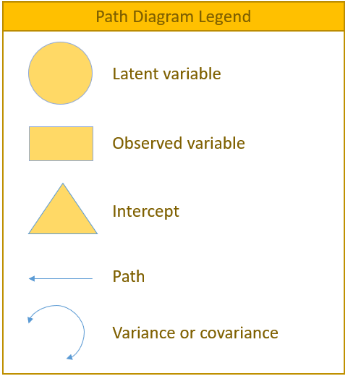
- Una representación gráfica del modelo se puede ver a continuación:



Ejemplo motivacional II

- A las variables que se pueden medir directamente, usualmente se les llama **variables manifiestas** y en las gráficas se representan con rectángulos.
- Las varaibles que no se pueden medir directamente, son las **variables latentes** o **factores**, que se representan con círculos o elipses. Usualmente los residuales son latentes.
- Las flechas dirigidas ($\rightarrow$, $\leftarrow$) indican al inicio las variables independientes, exógenas o predictores y al final son las variables dependientes, endógenas o de respuesta.
- Las flechas de doble sentido ($\leftrightarrow$) indican que hay correlación o covarianza entre las variables, pero no causalidad.
- El coeficiente del residual se fija en el valor 1 de antemano. Esto a veces se hace con la varianza de algunas variables latentes.

Ejemplo motivacional III



Análisis Factorial (exploratorio)

Origen y propósito del análisis factorial (FA) I

- Desarrollado originalmente por Charles Spearman y Karl Pearson (1904) como primeros intentos para medir conceptos abstractos o *constructos*[1] psicométricos, como la inteligencia, la personalidad, satisfacción, la creatividad, actitudes, pasión, confianza, aversión al riesgo, 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 explica 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*).

Origen y propósito del análisis factorial (FA) II

- El principal propósito de FA 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*.
- *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.
- FA se aplica en dos contextos de análisis de datos:
  - **Análisis exploratorio**: cuando no hay hipótesis o teoría y se explora la existencia de un patrón o relaciones de los datos que no son atribuíbles al azar, se puede pensar como *aprendizaje no supervisado*.
  - **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. Es similar al concepto de *aprendizaje supervisado*.

Origen y propósito del análisis factorial (FA) III

- Adicionalmente, AF así como PCA se pueden considerar como técnicas de reducción de dimensión. En la investigación de Marketing una tentación es recolectar tanta información como sea posible de los clientes potenciales u objetivo. La idea de estos métodos es evitar el “doble conteo” de la misma información.

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

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

Ejemplos en Mercadotecnia

- **Reducción de datos**: Por ejemplo, resumir el valor percibido de un servicio o producto, comenzando de un cuestionario multi-item. El valor de un servicio bancario se resume en tres dimensiones: servicio profesional, eficiencia de marketing y comunicación efectiva.
- **Mapas perceptuales**: nuevas oportunidades de mercado y espacio de producto. Por ejemplo, medición de las preferencias de usuario final respecto a productos lácteos light e identificar oportunidades para desarrollar un nuevo producto en yogurth de fresa más light en lugar de quesos tipo Cheddar más light.
- **Preparación preeliminar de datos**: Para segmentación. Por ejemplo segmentación de juegos de computadora, tres principales *drivers* de las conductas de juego:
  1. Actitudes y conocimiento
  2. hábitos de juego
  3. hábitos de compra

  Con éstos se identifican dos tipos de jugadores: los *hardcore* y los casuales

Modelo

Modelo básico I

- Si $\mathbf{x}$ es un vector en $\mathbb{R}^p$ con media $\boldsymbol{\mu}$ y covarianza $\boldsymbol{\Sigma}$, el modelo supone que los datos se generan de acuerdo a un modelo donde la varianza de cada una de ellas se puede descomponer en una varianza común a todas las variables más una varianza debida a errores de medición y propia de cada variable.
- En términos formales, se supone un modelo de la forma:

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

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**
  - $\boldsymbol{\epsilon}_{p\times 1}$ es un vector de errores o **factores específicos**

Modelo básico II

- Sin notación matricial, las ecuaciones anteriores son de la forma:

$$
\begin{aligned}
x_1 &= \mu_1 + l_{11} f_1 + l_{12} f_2 + \cdots + l_{1m} f_m + \epsilon_1 \\
x_2 &= \mu_2 + l_{21} f_1 + l_{22} f_2 + \cdots + l_{2m} f_m + \epsilon_2 \\
\vdots &= \vdots \\
x_p &= \mu_p + l_{p1} f_1 + l_{p2} f_2 + \cdots + l_{pm} f_m + \epsilon_p
\end{aligned}
$$

- Usualmente, **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 (**L** y ϵ suman $pm + p$ parámetros, y además 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:

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

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

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

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

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

Estructura de covarianza subyacente de **x** I

- El modelo de factores ortogonales implica una estructura específica para la covarianza de **x**:

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

  Así, la varianza se puede descomponer en la variabilidad de los factores comunes ($\mathbf{L}\mathbf{L}'$) y de los factores específicos ($\boldsymbol{\Psi}$).
- La matriz **L** se llama matriz de cargas factoriales y es la que sirve para expresar la información contenida en las variables originales.
- El propósito de AF y de SEM en general, es reproducir la matriz de covarianza usando parámetros de los modelos propuestos, para las relaciones de medición y/o estructurales de las variables.

Las comunalidades I

- A partir de la relación $\mathbf{\Sigma} = \mathbf{LL}' + \mathbf{\Psi}$, se ve que la varianza de cada variable es de la forma:

  $$\text{Var}(X_i) = l_{i1}^2 + \cdots + l_{im}^2 + \psi_i$$

  A la suma de los cuadrados $h_i^2 = \sum_{k=1}^{m} l_{ik}^2$ se le llama *comunalidad*
- La varianza de cada variable se descompone en dos partes: una que es explicada por los factores comunes y otra que es propia de la variable y expresada por el factor no común.
- Las comunalidades miden la cantidad de información que los factores comunes expresan de cada variable e indican una manera de evaluar la pertinencia de la aplicación del análisis a las variables iniciales.

Las comunalidades II

- Así que la varianza de cada variable se puede ver descompuesta como sigue:

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

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

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

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 de manera única.

Ejemplo. [Solución no propia]

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

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

$$\begin{array}{rrr} 1 = l_{11}^2 + \psi_1 & 0.90 = l_{11}l_{12} & 0.70 = l_{11}l_{13} \\ & 1 = l_{12}^2 + \psi_2 & 0.40 = l_{12}l_{13} \\ & & 1 = l_{13}^3 + \psi_3 \end{array}$$

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.

$\square$

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 $\mathrm{E}[\mathbf{f}^*] = \mathbf{T}'\mathrm{E}[\mathbf{f}] = \mathbf{0}$ y $\mathrm{Var}(\mathbf{f}^*) = \mathbf{T}'\mathrm{Var}(\mathbf{f})\mathbf{T} = \mathbf{T}'\mathbf{T} = \mathbf{I}$, 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 $\boldsymbol{\Psi}$.
- 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. 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, pero puede ayudar a la interpretación. Usualmente hay un *trade-off*.
- Posteriormente se identifican los factores latentes y se calculan los *scores de los factores*, como en componentes principales.

¿Es apropiado aplicar FA? I

- Hay una medida estadística para medir si la muestra de datos es adecuada para aplicar factores, se conoce como la medida de Kaiser-Meyer-Olkin (KMO). Esta prueba mide la adecuación de cada variable y de todas.
- La estadística es una medida de la proporción de varianza entre las variables que puede ser varianza común. Mientras más baja es la proporción, más adecuados son los datos para FA.
- De acuerdo a los criterios de Kaiser (1974), se sugieren los siguientes valores:
  - KMO > .9 we are marvelous,
  - in the .80s, mertitourious,
  - in the .70s, middling,
  - in the .60s, mediocre,
  - in the 50s, miserable, and
  - less than .5, unacceptable.
- Podemos decir que es razonable aplicar la técnica si $KMO \geq 0.6$
- Más adelante veremos la aplicación.

¿Es apropiado aplicar FA? II

- Otra prueba es la prueba de esfericidad de Bartlett, que compara la matriz de correlación a una matriz identidad. La hipótesis nula de la prueba de Bartlett es

$$H_0 : matriz = identidad$$

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

```r
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)
```

Ejemplo de preferencias de consumo II

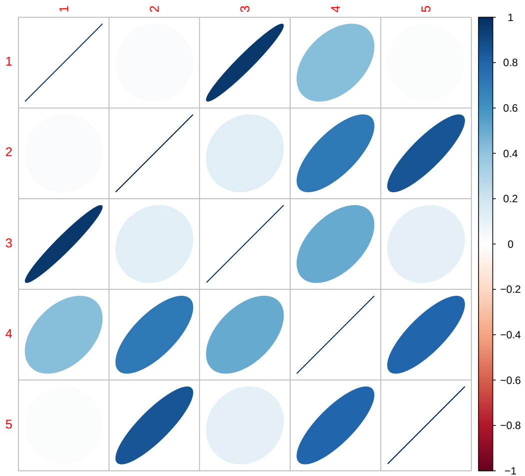
- Veamos las correlaciones de los datos. ¿Es posible pensar que hay factores representativos de los datos?

```
library(corrplot)
```

```
corrplot 0.90 loaded
```

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

Ejemplo de preferencias de consumo III



Ejemplo de preferencias de consumo IV

Las variables que muestran alta correlación son la pareja (Gusto, Sabor), (Energía, buena compra) y (Energía, tentempié). Pareceria que puede haber dos o tres factores asociados al modelo. Usaremos la descomposición en componentes principales para obtener los factores.

- Para medir la factibilidad del análisis de factores, podemos aplicar la prueba KMO:

```
library(psych)
KMO(r = R)  #psych

Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = R)
Overall MSA =  0.65
MSA for each item =
[1] 0.54 0.76 0.56 0.78 0.65
```

El valor de la estadística es 0.65, lo que indica que la factorabilidad de la muestra es mediocre, pero aun así se puede considerar que el análisis factorial puede ser útil.
La prueba de Bartlett nos da:

Ejemplo de preferencias de consumo V

```
cortest.bartlett(R)

Warning in cortest.bartlett(R): n not specified, 100 used

$chisq
[1] 542.0634

$p.value
[1] 4.474383e-110

$df
[1] 10
```

Extracción de los factores

El método más conocido para hallar los coeficientes de carga es el de componentes principales. Hay muchos otros métodos que incluyen:

- Máxima verosimilitud
- Factor principal iterado
- Mínimos cuadrados no ponderados
- Mínimos cuadrados generalizados
- Residuales mínimos.

El pquete `psych` en `R` tiene una amplia colección de metodologías

Método de componentes principales I

- Si **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

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

  donde $\hat{\psi}_i = s_{ii} - \sum_{j=1}^{m} \hat{l}_{ij}^2$
- De este modo, con $m < p$, podemos escribir:

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

Herramientas de estimación de AF en R I

- 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 funcion `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 de preferencias de consumo (continuación) I

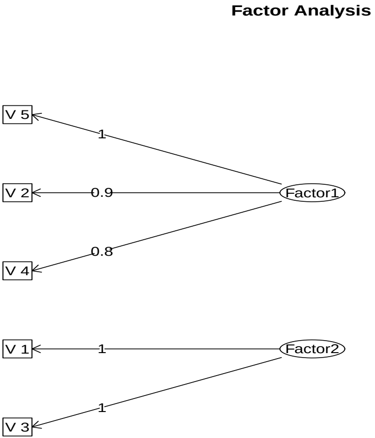
- A continuación se obtienen los factores:

```
m1 <- factanal(covmat = R, factors = 2)
m1
```

```

Call:
factanal(factors = 2, covmat = R)

Uniquenesses:
[1] 0.028 0.237 0.040 0.168 0.052

Loadings:
     Factor1 Factor2
[1,]          0.985
[2,]  0.873
[3,]  0.131   0.971
[4,]  0.817   0.405
[5,]  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
```

```
fa.diagram(m1$loadings)
```

Ejemplo de preferencias de consumo (continuación) II



Ejemplo de preferencias de consumo (continuación) III

Podemos ver que el modelo no permite estimar tres factores, por las restricciones del modelo:

```
factanal(covmat = R, factors =3)
```

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

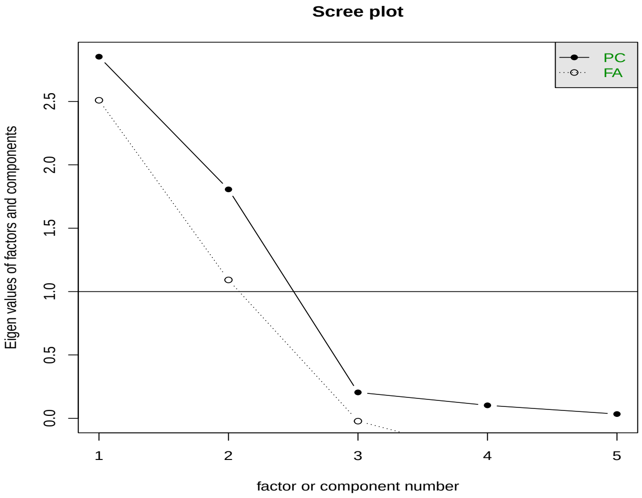
Selección del número de factores m I

- En general, se pueden tener menos o más factores de los necesarios. En ese caso se puede tener un sesgo en los resultados.
  - Usualmente sobre-extracción es menos serio que sub-extracción, entonces es mejor tener más que menos.
- La elección del número de factores se puede hacer con las técnicas usuales (por ejemplo, una gráfica de codo).
- 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:

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

```
scree(R)

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
```

Selección del número de factores m II



Selección del número de factores m III

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

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 $j = 1, \ldots, n$.
- 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**

- Se supone que $\boldsymbol{\mu}$, $\mathbf{L}$ y $\boldsymbol{\Psi}$ conocidas, entonces para el modelo de factores

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

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

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

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

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

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

- 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.

Estimación de scores de los factores: Método de Regresión I

- Sabemos que $\mathbf{x}-\boldsymbol{\mu} \sim \mathcal{N}_p\left(\mathbf{0}, \mathbf{L}\mathbf{L}'+\boldsymbol{\Psi}\right)$ y $\mathbf{f} \sim \mathcal{N}_m\left(\mathbf{0}, \mathbf{I}_m\right)$, y además, $\mathrm{cov}\left(\mathbf{x}, \mathbf{f}\right) = \mathbf{L}$.
- 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

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

- $\mathbf{L}'(\mathbf{L}\mathbf{L}'+\boldsymbol{\Psi})^{-1}$ 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:

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

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.

Interpretación de los factores

Interpretación de los factores I

- La interpretación está basada en el valor de las cargas para cada factor.
- Las cargas de los factores son las correlaciones entre cada una de las variables originales y el factor y están acotadas entre -1 y 1, donde valores altos (positivos o negativos) indican que la liga entre el factor y la variable considerada es muy fuerte.
- La relevancia estadística de una carga depende de su valor absoluto y del tamaño de muestra, así que el mismo valor para una carga supone diferentes relevancias dependiendo del tamaño de muestra. Una regla de dedo:
  - Con una muestra de tamaño 50, tomar 0.75
  - Con una muestra de tamaño 100 usar límite de 0.55
  - Con una muestra de tamaño 200 el valor cae a 0.40

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:

- 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 $\mathbf{X}$. Si hay pocas o ninguna correlación alta, entonces no tiene sentido intentar siquiera FA.