

# Clase 14

## Análisis Factorial

## Exploratorio II

Análisis Avanzado de Datos

Gabriel Sotomayor



# Contenidos de la sesión

- Extracción de factores comunes
- Obtención e interpretación de la matriz factorial
- Evaluación del modelo factorial
- Cálculo de las puntuaciones factoriales



# La extracción de factores comunes



# Selección de variables

La selección de variables en un Análisis Factorial Exploratorio es un paso crucial para obtener una solución factorial interpretable y teóricamente coherente. Algunas consideraciones para seleccionar variables incluyen:

- **Relevancia teórica:** Las variables deben estar relacionadas con el concepto subyacente que se desea medir y tener una base teórica sólida.
- **Distribución de las variables:** Es conveniente verificar si las variables tienen distribuciones normales o al menos simétricas, ya que las distribuciones sesgadas o asimétricas pueden afectar la solución factorial.
- **Correlaciones entre variables:** Examinar la matriz de correlaciones para identificar variables que estén altamente correlacionadas entre sí, lo que puede indicar redundancia en la información proporcionada por esas variables.
- **Ajuste con el modelo factorial:** Una vez extraídos los factores comunes debe evaluarse el ajuste de cada variable y la posibilidad de modificar las variables.



# Criterios de selección del número de factores

Existen diferentes criterios para determinar el número óptimo de factores a extraer. Algunos de los más comunes son:

- **Regla del codo (scree plot):** Se grafican los autovalores de los factores en función del número de factores. Se busca el punto en el que la curva cambia de pendiente, el “codo”.
- **Criterio de Kaiser:** Se extraen solo los factores con autovalores mayores a 1. En el caso de AFE estamos trabajando solo con la varianza común, por lo que se consideran los que tienen un autovalor mayor a la varianza común promedio.
- **Criterios de ajuste (solo aplicable a ML):** Se comparan diferentes modelos con distintos números de factores utilizando índices de ajuste como el AIC, BIC, RMSEA.



# Criterios de selección del número de factores

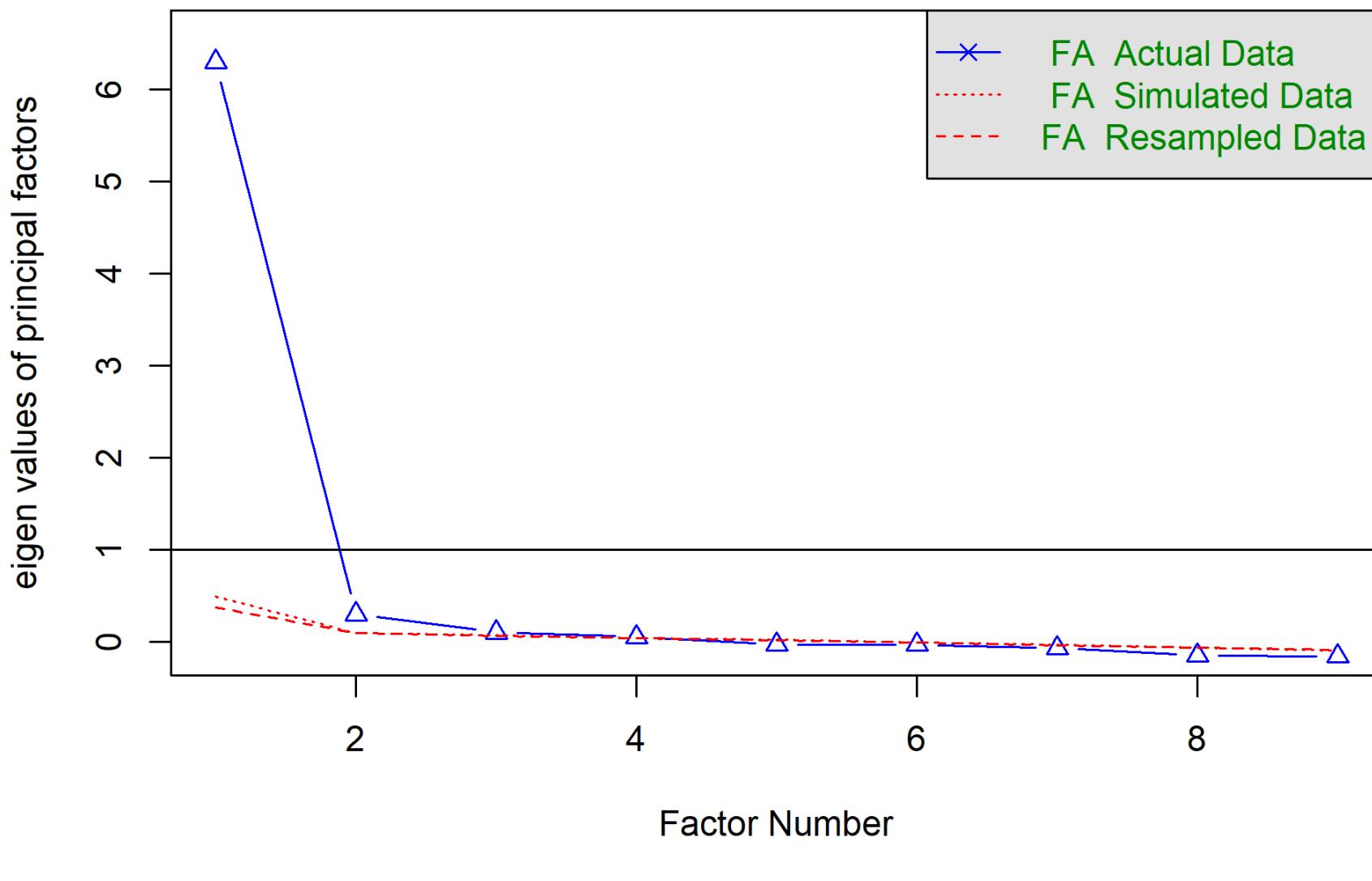
Existen diferentes criterios para determinar el número óptimo de factores a extraer. Algunos de los más comunes son:

- **Criterio de análisis paralelo:** Este enfoque compara los autovalores observados en los datos con los autovalores generados a partir de datos aleatorios con las mismas dimensiones. Los factores se retienen si los autovalores observados son mayores que los autovalores esperados por azar.
- **Interpretación teórica:** Retener los factores que se ajustan a las expectativas teóricas y proporcionan una solución interpretable.

A menudo, es útil considerar varios criterios al tomar esta decisión.



## Parallel Analysis Scree Plots



# Métodos de extracción de factores

Existen varios métodos de extracción de factores. Algunos de los más comunes son:

- **Mínimos cuadrados residuales (LSR):** Este método busca minimizar la suma de las diferencias al cuadrado entre las correlaciones observadas y las correlaciones estimadas por el modelo factorial, teniendo en cuenta las correlaciones residuales entre las variables. Es útil para trabajar con datos que presentan correlaciones no lineales o no normales entre las variables. **Es el que R usa por defecto en la función fa()**
- **Máxima verosimilitud (ML):** Este método busca extraer factores que maximicen la probabilidad de obtener la matriz de correlaciones observada. Asume normalidad multivariante. Permite la inferencia estadística.



# Rotación de factores

La rotación de factores es un paso importante en la interpretación de los resultados del AFE. Los factores extraídos inicialmente a menudo no son fácilmente interpretables, ya que las cargas factoriales pueden estar distribuidas de manera uniforme entre los factores. La rotación busca simplificar la estructura factorial, haciendo que cada variable tenga cargas altas en un factor y cargas bajas en otros. Redistribuye la varianza explicada entre las estructuras latentes.

Existen dos tipos principales de rotación:

- **Rotación ortogonal:** Los factores se mantienen no correlacionados entre sí (por ejemplo, varimax, quartimax).
- **Rotación oblicua:** Se permite que los factores estén correlacionados entre sí (por ejemplo, promax, oblimin). La elección del tipo de rotación dependerá de las expectativas teóricas y de si se espera que los factores estén relacionados entre sí.



# Interpretación de la matriz factorial



# Matriz factorial

La matriz factorial es la tabla que muestra las cargas factoriales de cada variable en cada factor. Las cargas factoriales representan la relación entre las variables observadas y los factores latentes. Cuanto mayor sea la carga factorial de una variable en un factor, mayor será la contribución de esa variable al factor.

Para interpretar la matriz factorial, se busca un patrón claro de cargas altas y bajas en los factores. Las variables con cargas altas en un factor se consideran relacionadas entre sí y pueden representar una dimensión subyacente del concepto que se mide.



```
## Factor Analysis using method = minres
## Call: fa(r = datosLW, nfactors = 4, rotate = "none")
## Standardized loadings (pattern matrix) based upon correlation matrix
##          MR1    MR2    MR3    MR4    h2    u2 com
## SALUD  0.80  0.11 -0.14  0.00  0.68  0.325 1.1
## INGR   0.84  0.38 -0.18  0.26  0.95  0.045 1.7
## TRAB   0.86 -0.02 -0.15 -0.11  0.77  0.228 1.1
## EDUC   0.89  0.05 -0.07 -0.20  0.84  0.156 1.1
## VIVI   0.89  0.01 -0.01 -0.18  0.82  0.182 1.1
## SEGUR  0.76  0.31  0.40 -0.01  0.84  0.164 1.9
## MEDIO  0.83 -0.11  0.14  0.05  0.72  0.276 1.1
## LIBER  0.88 -0.42  0.07  0.18  0.98  0.024 1.5
## PROYE  0.85 -0.27 -0.01  0.03  0.79  0.210 1.2
##
##          MR1    MR2    MR3    MR4
## SS loadings       6.42  0.52  0.27  0.19
## Proportion Var    0.71  0.06  0.03  0.02
## Cumulative Var   0.71  0.77  0.80  0.82
## Proportion Explained 0.87  0.07  0.04  0.03
## Cumulative Proportion 0.87  0.94  0.97  1.00
```



# Interpretación de la matriz factorial

Una vez que se haya rotado la matriz factorial, los siguientes pasos pueden ayudar a interpretar los resultados:

- **Identificar las variables con cargas altas en cada factor:** Establecer un umbral de carga factorial (por ejemplo,  $> 0.30$ ) y considerar las variables que superan ese umbral como importantes para cada factor. Idealmente se esperan que cada variable tenga una carga sobre 0,7 con un solo factor y ningún otro sobre 0,3.
- **Interpretar el significado de cada factor:** Analizar el conjunto de variables con cargas altas en cada factor e intentar identificar un tema o dimensión común que las relate.



# Interpretación de la matriz factorial

Una vez que se haya rotado la matriz factorial, los siguientes pasos pueden ayudar a interpretar los resultados:

- **Asignar nombres a los factores:** Dar a cada factor un nombre descriptivo que refleje el tema o dimensión común identificado.
- **Evaluar la solidez y coherencia de los factores:** Comprobar si los factores son teóricamente coherentes y si se ajustan a las expectativas previas. Además, se pueden realizar análisis de fiabilidad y validez convergente y discriminante para evaluar la calidad de la solución factorial.



```
## Factor Analysis using method = minres
## Call: fa(r = datosLW, nfactors = 4, rotate = "promax")
## Standardized loadings (pattern matrix) based upon correlation matrix
##          MR1    MR2    MR4    MR3    h2    u2 com
## SALUD    0.46   0.10   0.34 -0.02  0.68  0.325 2.0
## INGR     0.00   0.00   0.92   0.08  0.95  0.045 1.0
## TRAB     0.72   0.16   0.11 -0.09  0.77  0.228 1.2
## EDUC     0.85   0.01   0.00   0.08  0.84  0.156 1.0
## VIVI     0.75   0.10  -0.06   0.14  0.82  0.182 1.1
## SEGUR    0.03  -0.02   0.06   0.86  0.84  0.164 1.0
## MEDIO    0.13   0.54   0.02   0.25  0.72  0.276 1.5
## LIBER   -0.09   1.08   0.01 -0.04  0.98  0.024 1.0
## PROYE    0.28   0.69   0.00 -0.06  0.79  0.210 1.3
##
##          MR1    MR2    MR4    MR3
## SS loadings    2.72  2.37  1.25  1.05
## Proportion Var 0.30  0.26  0.14  0.12
## Cumulative Var 0.30  0.57  0.70  0.82
## Proportion Explained 0.37  0.32  0.17  0.14
## Cumulative Proportion 0.37  0.69  0.86  1.00
...
```



```
## Factor Analysis using method = minres
## Call: fa(r = datosLW, nfactors = 4, rotate = "promax")
## Standardized loadings (pattern matrix) based upon correlation matrix
##          MR1    MR2    MR4    MR3     h2     u2 com
## SALUD   0.46  0.10  0.34 -0.02  0.68  0.325  2.0
## INGR    0.00  0.00  0.92  0.08  0.95  0.045  1.0
## TRAB    0.72  0.16  0.11 -0.09  0.77  0.228  1.2
## EDUC    0.85  0.01  0.00  0.08  0.84  0.156  1.0
## VIVI    0.75  0.10 -0.06  0.14  0.82  0.182  1.1
## SEGUR   0.03 -0.02  0.06  0.86  0.84  0.164  1.0
## MEDIO   0.13  0.54  0.02  0.25  0.72  0.276  1.5
## LIBER   -0.09 1.08  0.01 -0.04  0.98  0.024  1.0
## PROYE   0.28  0.69  0.00 -0.06  0.79  0.210  1.3
##
##          MR1    MR2    MR4    MR3
## SS loadings   2.72  2.37  1.25  1.05
## Proportion Var 0.30  0.26  0.14  0.12
## Cumulative Var 0.30  0.57  0.70  0.82
## Proportion Explained 0.37  0.32  0.17  0.14
## Cumulative Proportion 0.37  0.69  0.86  1.00
##
```



# La evaluación del modelo factorial



# Ajuste y cambios en el modelo

La interpretación de los resultados y la mejora del modelo pueden llevarse a cabo considerando los siguientes aspectos:

- **Patrones de carga:** Evaluar la existencia de una estructura simple y si la estructura de carga de los factores es interpretable, y/o se ajusta con las expectativas teóricas, en caso de que existan.
- **Complejidad de los ítems:** Ítems con cargas elevadas en múltiples factores pueden indicar ambigüedad. Considere redefinir o eliminar estos ítems.
- **Correlaciones entre factores:** Factores altamente correlacionados pueden sugerir redundancia o solapamiento. Evaluar un número diferente de factores.
- **Ajuste del modelo:** Utilizar los estadísticos de ajuste para evaluar la calidad del modelo y realizar cambios si es necesario.



# Estadísticos de ajuste y rangos esperados

Los estadísticos de ajuste evalúan la bondad del ajuste entre el modelo y los datos. Algunos de los índices de ajuste más comunes son:

- RMSR (Root Mean Square of the Residuals): Raíz cuadrada de la media de los cuadrados de los residuos. Valores cercanos a 0 indican buen ajuste.
- RMSEA (Root Mean Square Error of Approximation): Medida del error de aproximación por grado de libertad. Valores  $< 0.05$  indican buen ajuste, mientras que valores entre 0.05 y 0.08 sugieren ajuste razonable.
- TLI (Tucker-Lewis Index): Índice que compara el modelo propuesto con un modelo nulo. Valores  $> 0.95$  indican buen ajuste.
- BIC (Bayesian Information Criterion): Penaliza la complejidad del modelo. Valores menores indican un mejor ajuste.

Los últimos 3 solo son directamente aplicables al utilizar máxima verosimilitud (ML)



# Cálculo y uso de puntuaciones factoriales



# Cálculo de puntuaciones factoriales

Una vez que se hayan interpretado y validado los factores, se pueden calcular las puntuaciones factoriales para cada individuo en la muestra. Es importante considerar que estos puntajes corresponden a estimaciones, y no a una medición directa de la variable latente.

Por defecto R usa del método “Thurstone” o basado en regresión. Podemos extraer estas estimaciones a partir de la función “factore.scores()” en R o guardando nuestro análisis factorial como un objeto y seleccionando el elemento scores.



# Uso de puntuaciones factoriales

Las puntuaciones factoriales pueden utilizarse en investigaciones futuras de diversas maneras:

- **Comparaciones entre individuos:** Las puntuaciones factoriales permiten comparar a los individuos en función de sus puntuaciones en las dimensiones subyacentes identificadas en el AFE.
- **Clasificación de individuos:** Se pueden utilizar las puntuaciones factoriales para clasificar a los individuos en grupos en función de sus puntuaciones en cada factor y analizar las diferencias entre los grupos en otras variables de interés.
- **Análisis de regresión:** Las puntuaciones factoriales pueden emplearse como variables independientes o dependientes en análisis de regresión para explorar las relaciones entre las dimensiones subyacentes y otras variables.



