**CUARTO EJERCICIO OBLIGATORIO PARA SUPERAR EL SEMINARIO. Parte II – MÁXIMA VEROSIMILITUD con lavaan**

**Responde este ejercicio en este Word y entrega también anexado el script de R.**

El archivo de datos que vamos a abrir es *AFC\_2\_factores.dat*, cada uno de ellos con 8 ítems (las variables son x1, x2, …, x16, siendo {x1 a x8} los indicadores que definen el primer factor y {x9 a x16} los indicadores que definen el segundo factor.

Este archivo contiene datos simulando una estructura factorial de dos factores con pesos estandarizados = 0.60 y una correlación entre los dos factores = 0.50. Se ha simulado una pérdida de datos MNAR (ya sabes que los métodos modernos no están pensados para pérdida MNAR, pero queremos ver la robustez de las estimaciones).

Realiza dos análisis factoriales confirmatorios en lavaan. Uno utilizando Full Maximum Likelihood (recuerda que en lavaan esto esto se hace con el argumento missing = "ML") y otro sin utilizar FML. Verás que no te permite estimar el modelo con listwise (eliminación por lista) dado que se pierden la mayor parte de los sujetos. Haz una imputación por la media de los valores perdidos (sabemos que esta estrategia no es buena) y vuelve a correr el análisis factorial confirmatorio con lavaan tras esa imputación.

1. *Ajuste de los dos modelos*.Muestra los índices de ajuste de FIML (lavaan) frente a los que arroja la solución con imputación por la media. Compáralos. Informa de los índices de ajuste

| **MODEL** | **NPAR** | **Chi2(103)** | **p (Chi2)** | **CFI** | **TLI** | **RMSEA** | **RMSEA CI** | **SRMR** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| cfa\_fiml | 49 | 114.430 | 0.208 | 0.981 | 0.978 | 0.020 | [0.000, 0.038] | 0.059 |
| cfa\_media | 49 | 118.133 | 0.146 | 0.973 | 0.968 | 0.023 | [0.000, 0.040] | 0.048 |

Ambos modelos presentan un ajuste aceptable según los criterios clásicos, aunque el modelo estimado con FIML muestra índices ligeramente superiores en CFI, TLI y RMSEA. El modelo con imputación por la media obtiene un SRMR algo menor. Dado que la imputación por la media no respeta la estructura de covarianzas, FIML se considera el método más robusto y adecuado en este contexto.

2. *Soluciones factoriales de los dos modelos*. Muestra (basta poner pantallazos) la solución factorial para FML frente a imputación por la media. La solución factorial se refiere a los pesos factoriales, su error típico, la significación estadística, etc. Compara las soluciones factoriales y justifica por qué pueden darse esas diferencias.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **lhs** | **op** | **rhs** | FIML | | | | Media | | | |
| **Est.std** | **se** | **z** | **pvalue** | **Est.std** | **se** | **z** | **pvalue** |
| F1 | =~ | x1 | 0.534 | 0.059 | 9.023 | <0.001 | 0.434 | 0.058 | 7.498 | <0.001 |
| F1 | =~ | x2 | 0.602 | 0.056 | 10.763 | <0.001 | 0.458 | 0.057 | 8.071 | <0.001 |
| F1 | =~ | x3 | 0.544 | 0.058 | 9.324 | <0.001 | 0.465 | 0.056 | 8.238 | <0.001 |
| F1 | =~ | x4 | 0.507 | 0.061 | 8.351 | <0.001 | 0.429 | 0.058 | 7.379 | <0.001 |
| F1 | =~ | x5 | 0.600 | 0.055 | 10.959 | <0.001 | 0.525 | 0.053 | 9.863 | <0.001 |
| F1 | =~ | x6 | 0.566 | 0.058 | 9.838 | <0.001 | 0.462 | 0.057 | 8.175 | <0.001 |
| F1 | =~ | x7 | 0.713 | 0.045 | 15.947 | <0.001 | 0.632 | 0.047 | 13.353 | <0.001 |
| F1 | =~ | x8 | 0.743 | 0.044 | 16.995 | <0.001 | 0.616 | 0.048 | 12.769 | <0.001 |
| F2 | =~ | x9 | 0.660 | 0.055 | 11.952 | <0.001 | 0.560 | 0.055 | 10.179 | <0.001 |
| F2 | =~ | x10 | 0.491 | 0.064 | 7.619 | <0.001 | 0.441 | 0.060 | 7.283 | <0.001 |
| F2 | =~ | x11 | 0.430 | 0.071 | 6.089 | <0.001 | 0.356 | 0.064 | 5.576 | <0.001 |
| F2 | =~ | x12 | 0.570 | 0.060 | 9.425 | <0.001 | 0.452 | 0.060 | 7.531 | <0.001 |
| F2 | =~ | x13 | 0.494 | 0.066 | 7.432 | <0.001 | 0.403 | 0.062 | 6.495 | <0.001 |
| F2 | =~ | x14 | 0.430 | 0.071 | 6.088 | <0.001 | 0.356 | 0.064 | 5.573 | <0.001 |
| F2 | =~ | x15 | 0.626 | 0.058 | 10.875 | <0.001 | 0.543 | 0.056 | 9.745 | <0.001 |
| F2 | =~ | x16 | 0.569 | 0.062 | 9.128 | <0.001 | 0.466 | 0.059 | 7.858 | <0.001 |
| F1 | ~~ | F2 | 0.615 | 0.061 | 10.020 | <0.001 | 0.614 | 0.063 | 9.667 | <0.001 |

El modelo FIML mostró cargas ligeramente más altas y consistentes con los parámetros poblacionales simulados (λ ≈ 0.60), en comparación con la solución obtenida tras imputar los valores perdidos por la media. Mientras que en el modelo con imputación por la media, las cargas factoriales tienden a ser más bajas, especialmente en los ítems con mayor proporción de datos faltantes, como x1, x2, x4 o x11.

En ambos modelos, los valores de significación (z y p) indican que todas las cargas son significativas, pero los estadísticos z tienden a ser más altos en el modelo FIML, lo que sugiere estimaciones más estables. Con respecto a la correlación entre factores (F1 ~~ F2) se mantiene prácticamente igual en ambos modelos (≈ 0.615), lo cual puede indicar que la estructura general del modelo se preserva, aunque el modelo con imputación por la media reduce la precisión en las estimaciones individuales.

3. *Consistencia de los modelos.*  Como conclusión del ejercicio, ¿te parece consistente el método FIML pese a que hayamos generado pérdida de datos MNAR? Para ello puedes ver si **aproximadamente, en promedio**, los pesos estandarizados (los que se encabezan bajo la columna “**Std.all**”) tienen aproximadamente un valor de 0,60. ¿Y es consistente cuando imputamos los valores perdidos por las medias de las variables?

Las cargas factoriales estandarizadas (Std.all) obtenidas mediante FIML oscilan entre 0.49 y 0.74, con una media próxima a 0.60 (0.567). En tanto, la correlación estimada entre los factores es de 0.615, lo cual representa una ligera sobreestimación, pero sigue dentro de un rango razonable de aproximación. Esto sugiere que, a pesar de que los datos tienen una pérdida de tipo MNAR (para la cual FIML no está diseñado teóricamente), el método ha mostrado un comportamiento razonablemente bueno recuperando la estructura factorial simulada con relativa precisión.

4. *Restricción en el modelo*. Por último, aprovechando que lavaan es muy flexible, impón una restricción en el modelo tal que la **correlación entre los dos factores sea = 0 (**. ¿Cómo es el ajuste del modelo ahora? Por ejemplo, qué ocurre con el índice de ajuste SRMR. ¿Por qué crees que ha cambiado?

| **MODEL** | **NPAR** | **Chi2** | **Chi2\_df** | **p (Chi2)** | **CFI** | **TLI** | **RMSEA** | **RMSEA CI** | **SRMR** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| cfa\_ortogonal | 48 | 175.942 | 104 | < .001 | 0.883 | 0.865 | 0.050 | [0.037, 0.062] | 0.140 |
| cfa\_fiml | 49 | 114.430 | 103 | 0.208 | 0.981 | 0.978 | 0.020 | [0.000, 0.038] | 0.059 |
| cfa\_media | 49 | 118.133 | 103 | 0.146 | 0.973 | 0.968 | 0.023 | [0.000, 0.040] | 0.048 |

Se reestimó el modelo factorial confirmatorio imponiendo la restricción F1 ~~ 0\*F2, es decir, suponiendo que los dos factores latentes no se correlacionan. Al evaluarse y observa la tabla se tiene que hay una clara pérdida de ajuste global, siendo el indicador con mayor cambio el SRMR que aumentó de 0.059 a 0.140, reflejando un deterioro considerable en la capacidad del modelo para reproducir las correlaciones observadas entre ítems (residuos más grandes). También los índices CFI y TLI cayeron situándose por debajo del umbral aceptable de 0.90.

Esta pérdida de ajuste se explica debido a que los datos fueron generados con una correlación poblacional entre factores de ρ = 0.50, al forzar una correlación nula entre los factores viola la estructura latente real, lo que impide al modelo ajustarse adecuadamente a las covarianzas observadas entre los ítems.

*Anexo Código R*

##%######################################################%##  
# #  
#### Ejercicio 3 ###  
#### Brian Norman Peña Calero ###  
#### Seminario de Valores Perdidos ###  
# #  
##%######################################################%##  
  
# Carga de Paquetes -------------------------------------------------------  
  
library(tidyverse)  
library(lavaan)  
  
# Importar Datos Ejercicio 1 ----------------------------------------------  
  
afc\_2f <- read\_delim("AFC\_2\_factores.dat")  
afc\_2f  
  
  
  
# Especificación del modelo -----------------------------------------------  
  
modelo\_2f <- '  
 F1 =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8  
 F2 =~ x9 + x10 + x11 + x12 + x13 + x14 + x15 + x16  
'  
  
  
# Estimaciones ------------------------------------------------------------  
  
cfa\_fiml <- cfa(modelo\_2f, data = afc\_2f, missing = "ml")  
  
## Imputación por la media   
afc\_media <- afc\_2f %>%  
 mutate(  
 across(  
 everything(),   
 ~ ifelse(is.na(.), mean(., na.rm = TRUE), .)  
 )  
 )  
  
## Estimación con datos imputados   
cfa\_media <- cfa(modelo\_2f, data = afc\_media,  
 meanstructure = TRUE)  
  
## Comparación de índices  
  
psymetrics::compare\_model\_fit(  
 cfa\_fiml,  
 cfa\_media  
) |>   
 select(-c(NOBS:ESTIMATOR))   
  
  
# Comparación de soluciones factoriales -----------------------------------  
  
standardizedSolution(cfa\_fiml) |>   
 rename\_with(  
 .fn = ~ paste0(.x, "\_fiml"),  
 .cols = est.std:ci.upper  
 ) |>   
 bind\_cols(  
 standardizedSolution(cfa\_media) |>   
 rename\_with(  
 .fn = ~ paste0(.x, "\_media")  
 )  
 ) |>   
 filter(op %in% c("=~", "~~"),  
 lhs != rhs) |>   
 select(  
 lhs:pvalue\_fiml, est.std\_media:pvalue\_media  
 ) |>   
 mutate(  
 across(  
 c(pvalue\_fiml, pvalue\_media),  
 scales::pvalue  
 )  
 )   
  
  
# Promedio cargas ---------------------------------------------------------  
  
standardizedSolution(cfa\_fiml) |>   
 filter(op %in% c("=~")) |>   
 summarise(  
 mean\_lambda = mean(est.std)  
 )  
  
  
# Modelo ortogonal --------------------------------------------------------  
  
# Modelo con restricción de no correlación entre factores  
modelo\_2f\_ortogonal <- '  
 F1 =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8  
 F2 =~ x9 + x10 + x11 + x12 + x13 + x14 + x15 + x16  
 F1 ~~ 0\*F2  
'  
  
# Ajuste usando FIML  
cfa\_ortogonal <- cfa(modelo\_2f\_ortogonal, data = afc\_2f, missing = "ml")  
  
# Comparar índices de ajuste  
psymetrics::compare\_model\_fit(  
 cfa\_ortogonal,  
 cfa\_fiml,  
 cfa\_media  
) |>   
 select(-c(NOBS:ESTIMATOR))