Aplicación de los Índices en Modelos Bifactor

Asignatura: Validez

Brian N. Peña-Calero
Universidad Complutense de Madrid

11 marzo, 2025

Evaluaciones de los índices en Bifactor

Rodriguez, Reise, y Haviland (2016) indica 3 aspectos interpretativos relevantes en los índices bifactor:

1. **¿Los puntajes totales reflejan variación en una sola variable latente?** (Indicadores: ω, ω_H); y, de forma relacionada, ¿los puntajes de subescala reflejan varianza confiable independiente del factor general? (Indicadores: ω_S, ω_{HS})
2. **¿Pueden los ítems usarse para especificar variables latentes en un contexto SEM?** Indicadores: FD, H
3. **¿Son las medidas esencialmente unidimensionales y, por lo tanto, deben especificarse como una única variable latente en SEM?** Indicadores: ECV, PUC

Puntajes totales en un factor general o específicos

Fiabilidad en la Teoría Clásica de los Test (TCT)

En la **TCT**, la fiabilidad de una medida se define como la proporción de la varianza observada que se debe a la **varianza verdadera**, excluyendo el error de medición. Se parte de la ecuación fundamental:

$$O = V + e$$

Donde:

- O es la puntuación observada en un test
- V es la puntuación verdadera
- e es el error de medición

Definición de Fiabilidad

Dado que la varianza de una suma de variables es:

$$\mathrm{Var}(O) = \mathrm{Var}(V) + \mathrm{Var}(e) + 2\mathrm{Cov}(V, e)$$

Y asumiendo en la TCT que el error es **aleatorio** y no está correlacionado con la puntuación verdadera ($\mathrm{Cov}(V, e) = 0$), tenemos:

$$\mathrm{Var}(O) = \mathrm{Var}(V) + \mathrm{Var}(e)$$

Definición de Fiabilidad

Por lo que la fiabilidad (ρ) se expresa como:

$$\rho = \frac{\mathrm{Var}(V)}{\mathrm{Var}(O)} = \frac{\mathrm{Var}(V)}{\mathrm{Var}(V) + \mathrm{Var}(E)}$$

Es decir, la proporción de la varianza de la puntuación observada que es atribuible a la varianza verdadera.

En el análisis factorial, se puede hacer la asunción que la carga factorial al cuadrado ($\lambda^2 = h^2$) puede ser entendida como la $\mathrm{Var}(V)$; mientras que la unicidad ($1 - \lambda^2 = 1 - h^2 = u$) puede ser considerado como la $\mathrm{Var}(E)$.

Puntajes Totales

Uno de los más reportados sería el ω, que en un modelo bifactor, representa la varianza común atribuido a la suma del factor general y los factores específicos. *Un valor alto, refleja multidimensionalidad.*

$$\omega = \frac{\left(\sum \lambda_{\text{gen}}\right)^2 + \sum_{k=1}^{K} \left(\sum \lambda_{\text{grp}_k}\right)^2}{\left(\sum \lambda_{\text{gen}}\right)^2 + \sum_{k=1}^{K} \left(\sum \lambda_{\text{grp}_k}\right)^2 + \sum \left(1 - h^2\right)}$$

- λ_{gen}: Cargas factoriales de los ítems sobre el **factor general**.
- λ_{grp_k}: Cargas factoriales de los ítems sobre el **factor específico** (o grupo)
- $1 - h^2$: Representa el error de medición.

Puntajes Totales

En tanto, el ω_H estudia esta varianza común pero solo la parte que se encuentra atribuída al Factor General, eliminando la parte de los factores específicos.

$$\omega_H = \frac{\left(\sum \lambda_{\text{gen}}\right)^2}{\left(\sum \lambda_{\text{gen}}\right)^2 + \sum_{k=1}^{K}\left(\sum \lambda_{\text{grp}_k}\right)^2 + \sum(1 - h^2)}$$

- λ_{gen}: Cargas factoriales de los ítems sobre el **factor general**.
- λ_{grp_k}: Cargas factoriales de los ítems sobre el **factor específico** (o grupo)
- $1 - h^2$: Representa el error de medición.

Puntajes Sub-Escala:

Estima la fiabilidad de la sub-escala contando la varianza común del factor general y del factor en específico. Habitualmente esto, podría ser alto y sería un peligro evaluarlo así.

$$\omega_S = \frac{\left(\sum \lambda_{\text{gen}}\right)^2 + \left(\sum \lambda_{k=1}\right)^2}{\left(\sum \lambda_{\text{gen}}\right)^2 + \left(\sum \lambda_{k=1}\right)^2 + \sum(1 - h^2)}$$

- λ_{gen}: Cargas factoriales de los ítems sobre el **factor general**.
- λ_{grp_k}: Cargas factoriales de los ítems sobre el **factor específico** (o grupo)
- $1 - h^2$: Representa el error de medición.

Puntajes Sub-Escala:

Este omega jerárquico evaluado en la sub-escala (ω_{HS}) , indica que tanto de la varianza común del factor específico es únicamente del factor específico. *En un modelo bifactor que se sostiene, este indicador debería ser bajo*.

$$\omega_{HS} = \frac{\left(\sum \lambda_{\text{grp}_k}\right)^2}{\left(\sum \lambda_{\text{gen}}\right)^2 + \sum_{k=1}^{K}\left(\sum \lambda_{\text{grp}_k}\right)^2 + \sum\left(1 - h^2\right)}$$

- λ_{gen}: Cargas factoriales de los ítems sobre el **factor general**.
- λ_{grp_k}: Cargas factoriales de los ítems sobre el **factor específico** (o grupo)
- $1 - h^2$: Representa el error de medición.

Interpretaciones e indicaciones

- Instrumentos modelados de forma unifactorial pueden presentar una estimación de fiabilidad ω muy alta, a pesar de que ese único factor podría tener un ajuste cuestionable o no estar representando realmente un único factor.
- De forma similar sucederían en los modelos multifactoriales, en los que las estimaciones de sus factores (ω_s) presenten valores altos, a pesar de que un gran % de esa ψ_{S_k} realmente sean exactamente los mismos en los otros factores (un factor general no modelado).
- Los modelos bifactor pueden contemplar FE's que no contengan ψ_{S_k} suficiente para ser interpretados individualmente (ω_{HS} bajo con respecto a ω_s), pero que a la vez contengan suficiente ψ_{S_k} para tener que modelarlos y no prescindir de su especificación factorial.

Especifcaciones de los modelos en contexto SEM

Índices de FD y H

Para evaluar si los ítems son adecuados para definir variables latentes en un modelo SEM, se utilizan:

- **Factor Determinacy (FD):**
  Mide la correlación entre los puntajes estimados y la verdadera variable latente.

$$FD \;=\; \operatorname{diag}\left(\Phi\,\Lambda^T\,\Sigma^{-1}\,\Lambda\,\Phi\right)^{\frac{1}{2}}$$

  - Valores altos (≥ 0.90) indican que el factor está medido con excelente precisión
  - Valores moderados (0.70–0.80) pueden considerarse aceptables para algunas aplicaciones

Índices de FD y H

Para evaluar si los ítems son adecuados para definir variables latentes en un modelo SEM, se utilizan:

- **Índice H (Construct Replicability):**
  Cuan bien está definido el constructo a partir de sus indicadores.

$$H = \frac{1}{1 + \frac{1}{\sum_{i=1}^{k} \frac{\lambda_i^2}{1-\lambda_i^2}}}$$

  - Se suele considerar un valor H $\geq$ 0.80 como indicador de un constructo robusto.
  - Se considera H $\geq$ 0.70 como aceptable

Interpretaciones e indicaciones

- En un modelo SEM, solo se debería modelar factores específicos en los que tanto H como FD sean al menos ≥ 0.70.
- Una alternativa a esto, es modelar la escala como *esencialmente unidimensional*.
- Aunque el índice H pueda ser un buen indicador hay que prestar atención a las cargas factoriales de los ítems. Podría ser que el factor esté siendo modelado únicamente por una parte de ellos.
- Cargas factorialmente inusualmente altas (por ej. 0.80, 0.90, 0.30, 0.35, 0.28), podría estar debiéndose a algo distinto al factor específico al que pertenece, y no debería confiarse en las estimaciones del índice H o FD.

Unidimensionalidad esencial

Índices ECV y PUC

Para determinar si la medida puede tratarse como esencialmente unidimensional, se utilizan:

- **Explained Common Variance (ECV):**

$$\text{ECV} = \frac{\sum \lambda_{\text{gen}}^2}{\sum \lambda_{\text{gen}}^2 + \sum_{k=1}^{K} \sum \lambda_{\text{grp}_k}^2}$$

  - Un valor alto (> .70 o .80) sugiere que la mayor parte de la varianza común se debe al factor general.
  - Valores cercanos a 1 indican unidimensionalidad suficiente.
  - Valores bajos probablemente se asocian a que una estructura multidimensional pueda explicar mejor los datos.

Índices ECV y PUC

Para determinar si la medida puede tratarse como esencialmente unidimensional, se utilizan:

- **Percentage of Uncontaminated Correlations (PUC):**
  % de correlaciones entre ítems debido únicamente al factor general.

$$\text{PUC} = 1 - \frac{\#\text{ de correlaciones entre ítems del mismo factor}}{\#\text{ total de correlaciones}}$$

  - Un PUC elevado (por ejemplo, > 0.80) refuerza la interpretación unidimensional.
  - Valores cercanos a 1 indican unidimensionalidad suficiente.
  - Valores bajos probablemente se asocian a que una estructura multidimensional pueda explicar mejor los datos.

Interpretaciones e indicaciones

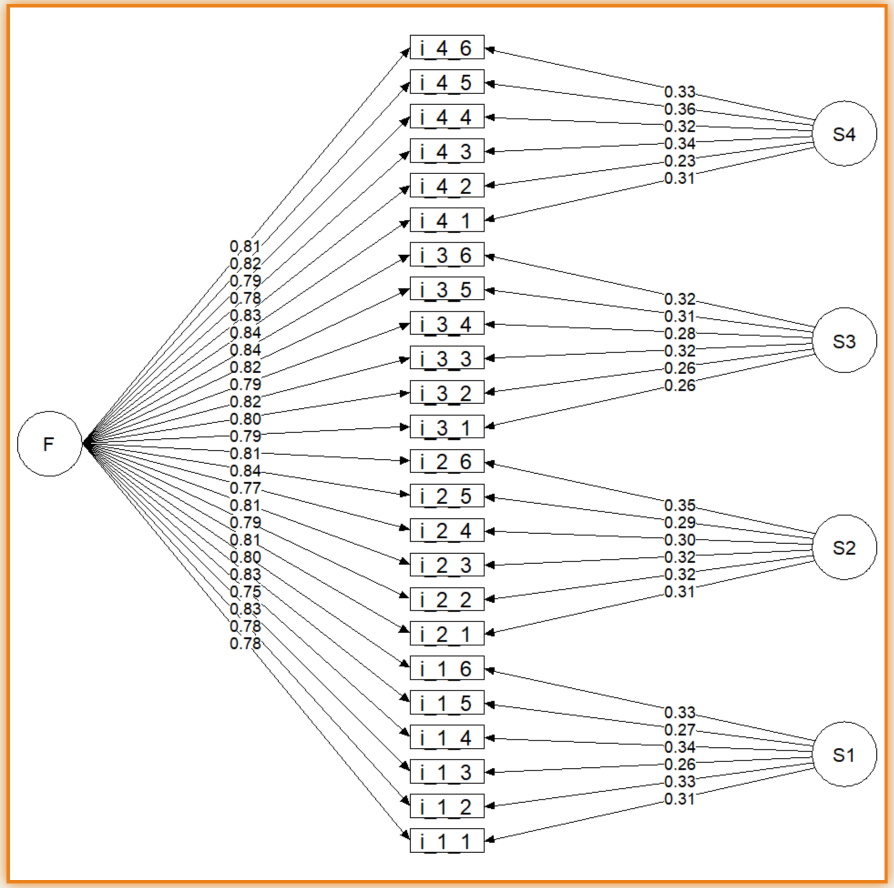
- En contexto SEM, algunos instrumentos pueden tratarse/especificarse de forma unidimensional sin cometer un sesgo relevante en las estimaciones de las cargas factoriales. Esto se ve apoyado cuando ECV y PUC > .70.
- En la revisión de Rodriguez, Reise, y Haviland (2016) observan que incluso cuando ECV se encuentra entorno a .50, si PUC es alto, podría tratarse el instrumento de forma unidimensional sin cometer un sesgo importante.
- PUC puede sobre-estimarse cuando se tiene muchos FE’s y pocos ítems en cada uno

Ejemplificación de un Bifactor Perfecto

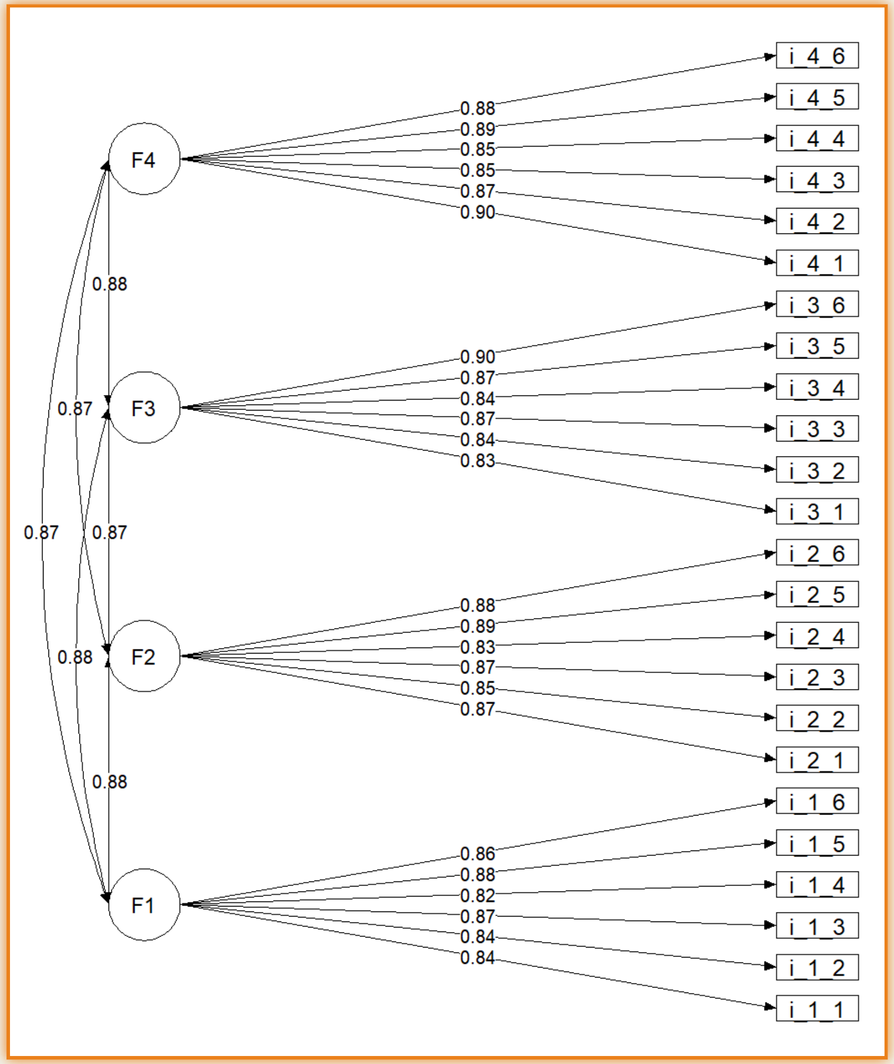
Ejemplificación

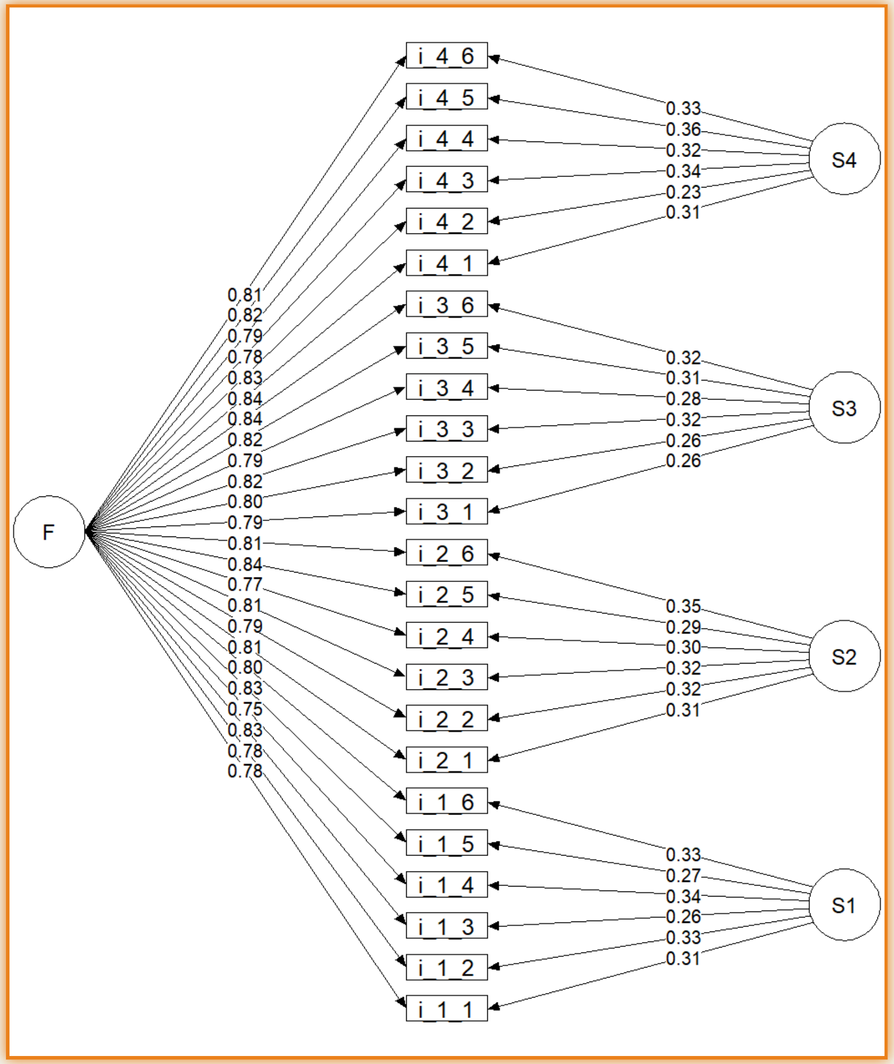
Con una muestra suficiente, 4 factores con 6 ítems en cada factor y un buen comportamiento bifactorial.

```r
library(lavaan)
library(semPlot)
library(BifactorIndicesCalculator)
source("simulaBifactor.R")

result_ideal <- simulaBifactor(
  sampleSize = 1000,
  nFactors = 4,
  itemsPerFactor = 6,
  loadingGeneral = 0.8,
  loadingSpecific = 0.3,
  fluctuation = 0.05,
  type_problematic = "perfect",
  estimator = "MLR"
)
```



¿Y si fuera multifactorial?





Solución multifactorial

```r
summary(result_ideal$fit_multi, fit.measures = TRUE, standardized = TRUE)
```

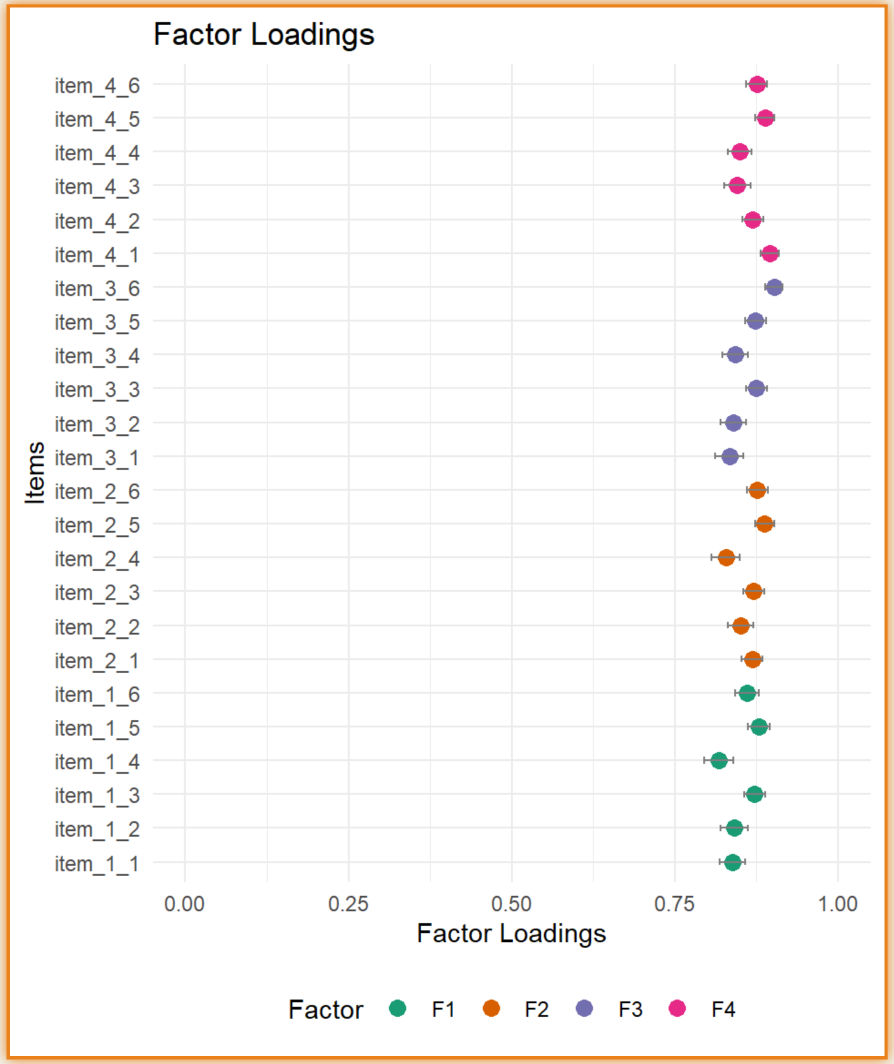
```
lavaan 0.6-19 ended normally after 39 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        54

  Number of observations                          1000

Model Test User Model:
                                              Standard      Scaled
  Test Statistic                               259.862     261.970
  Degrees of freedom                               246         246
  P-value (Chi-square)                           0.260       0.231
  Scaling correction factor                                  0.992
    Yuan-Bentler correction (Mplus variant)
```

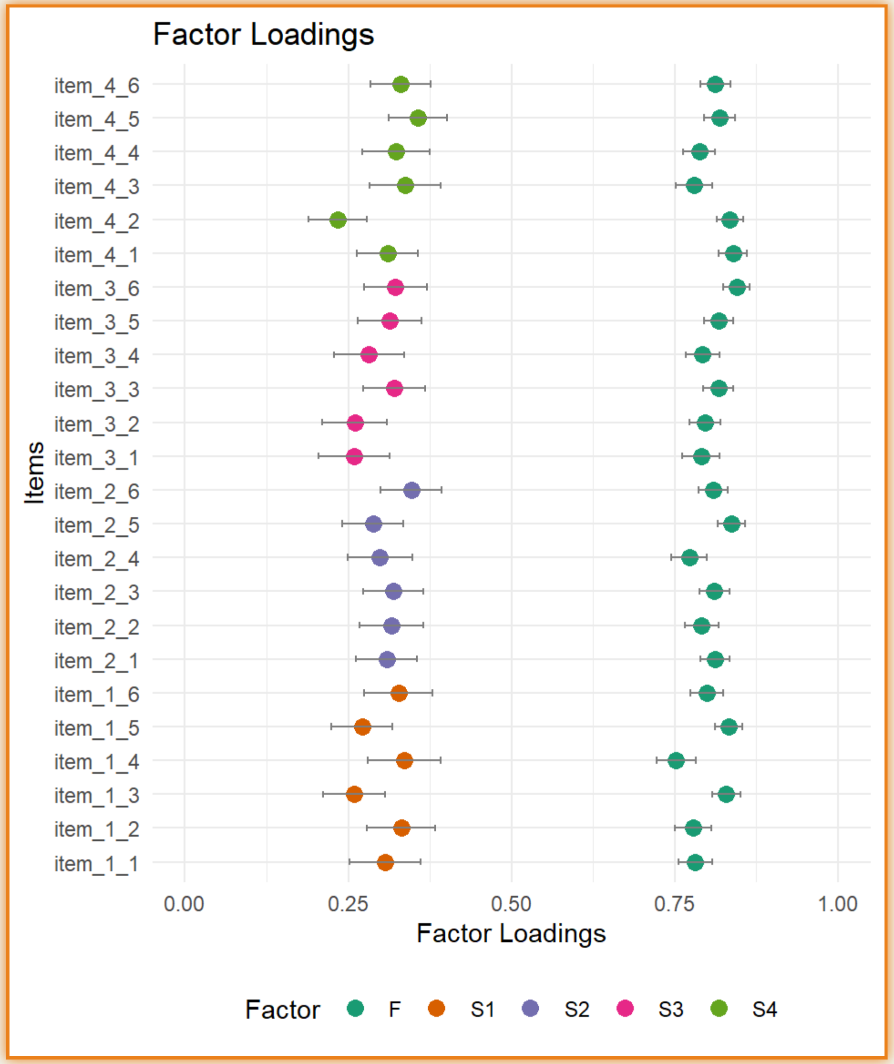
Solución multifactorial

```
summary(result_ideal$fit_bifactor, fit.measures = TRUE, standardized = TRUE)
```

```
lavaan 0.6-19 ended normally after 35 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        72

  Number of observations                          1000

Model Test User Model:
                                              Standard      Scaled
  Test Statistic                               218.905     221.095
  Degrees of freedom                               228         228
  P-value (Chi-square)                           0.656       0.616
  Scaling correction factor                                  0.990
    Yuan-Bentler correction (Mplus variant)
```

Organización de las cargas factoriales





Obtención de Índices

```r
bifactorIndices(result_ideal$fit_bifactor, UniLambda = result_ideal$fit_uni)
```

```
$ModelLevelIndices
     ECV.F        PUC    Omega.F   OmegaH.F       ARPB 
0.87250776 0.78260870 0.98438334 0.94996921 0.01564634 

$FactorLevelIndices
      ECV_SS     ECV_SG    ECV_GS     Omega    OmegaH         H        FD
F  0.8725078 0.87250776 0.8725078 0.9843833 0.9499692 0.9783633 0.9753907
S1 0.1292968 0.03159233 0.8707032 0.9410352 0.1208026 0.3847117 0.7239138
S2 0.1314340 0.03298034 0.8685660 0.9464750 0.1240935 0.3954727 0.7380738
S3 0.1164400 0.02903263 0.8835600 0.9451736 0.1092838 0.3628306 0.7176394
S4 0.1326829 0.03388694 0.8673171 0.9499120 0.1244707 0.4037519 0.7475393

$ItemLevelIndices
              IECV  RelParBias
item_1_1 0.8663711 0.014460658
```

¿Unidimesionalidad tiene sentido?

```r
summary(result_ideal$fit_uni, fit.measures = TRUE, standardized = TRUE)
```

```
lavaan 0.6-19 ended normally after 19 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        48

  Number of observations                          1000

Model Test User Model:
                                              Standard      Scaled
  Test Statistic                              2606.856    2634.837
  Degrees of freedom                               252         252
  P-value (Chi-square)                           0.000       0.000
  Scaling correction factor                                  0.989
    Yuan-Bentler correction (Mplus variant)
```

¿Unidimesionalidad tiene sentido?

```r
modificationindices(result_ideal$fit_uni,
                    maximum.number = 10,
                    sort. = TRUE)
```

```
          lhs op     rhs      mi   epc sepc.lv sepc.all sepc.nox
314 item_4_1 ~~ item_4_5 105.803 0.098   0.098    0.344    0.344
283 item_3_3 ~~ item_3_6 105.581 0.097   0.097    0.344    0.344
325 item_4_5 ~~ item_4_6  94.935 0.101   0.101    0.324    0.324
298 item_3_5 ~~ item_3_6  85.588 0.089   0.089    0.309    0.309
235 item_2_5 ~~ item_2_6  84.443 0.090   0.090    0.306    0.306
322 item_4_3 ~~ item_4_6  79.879 0.101   0.101    0.295    0.295
177 item_2_1 ~~ item_2_6  79.558 0.092   0.092    0.296    0.296
208 item_2_3 ~~ item_2_6  78.026 0.092   0.092    0.293    0.293
321 item_4_3 ~~ item_4_5  76.247 0.097   0.097    0.289    0.289
313 item_4_1 ~~ item_4_4  75.223 0.088   0.088    0.288    0.288
```

Comparación de ajustes

```r
psymetrics::compare_model_fit(
  result_ideal$fit_uni,
  result_ideal$fit_multi,
  result_ideal$fit_bifactor
)
```

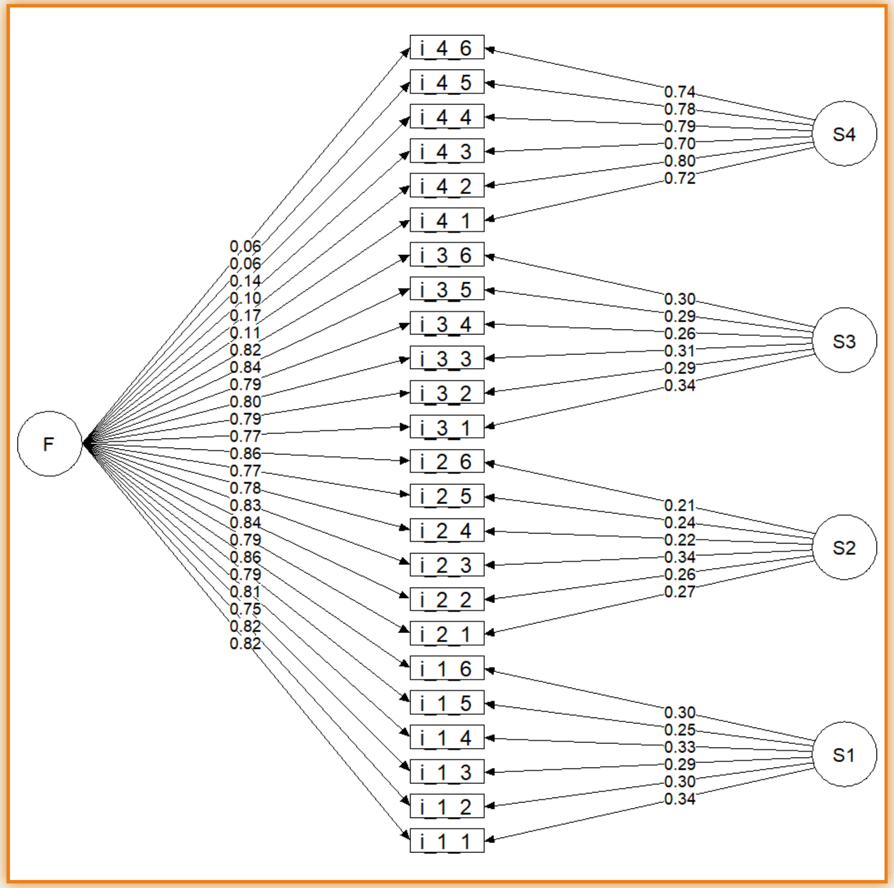
```
MODEL                     | NOBS | ESTIMATOR | NPAR |   Chi2    | Chi2_df
-------------------------------------------------------------------------
result_ideal$fit_uni      | 1000 |    MLR    |  48  | 2634.837 |   252
result_ideal$fit_multi    | 1000 |    MLR    |  54  | 261.970  |   246
result_ideal$fit_bifactor | 1000 |    MLR    |  72  | 221.095  |   228

MODEL                     | p (Chi2) |  CFI  |  TLI  | RMSEA |   RMSEA  CI    | SRMR
------------------------------------------------------------------------------------
result_ideal$fit_uni      |  < .001  | 0.906 | 0.897 | 0.097 | [0.094, 0.101] | 0.038
result_ideal$fit_multi    |  0.231   | 0.999 | 0.999 | 0.008 | [0.000, 0.016] | 0.012
result_ideal$fit_bifactor |  0.616   | 1.000 | 1.000 | 0.000 | [0.000, 0.012] | 0.009
```

Ejemplificación de un Bifactor con FE dominante

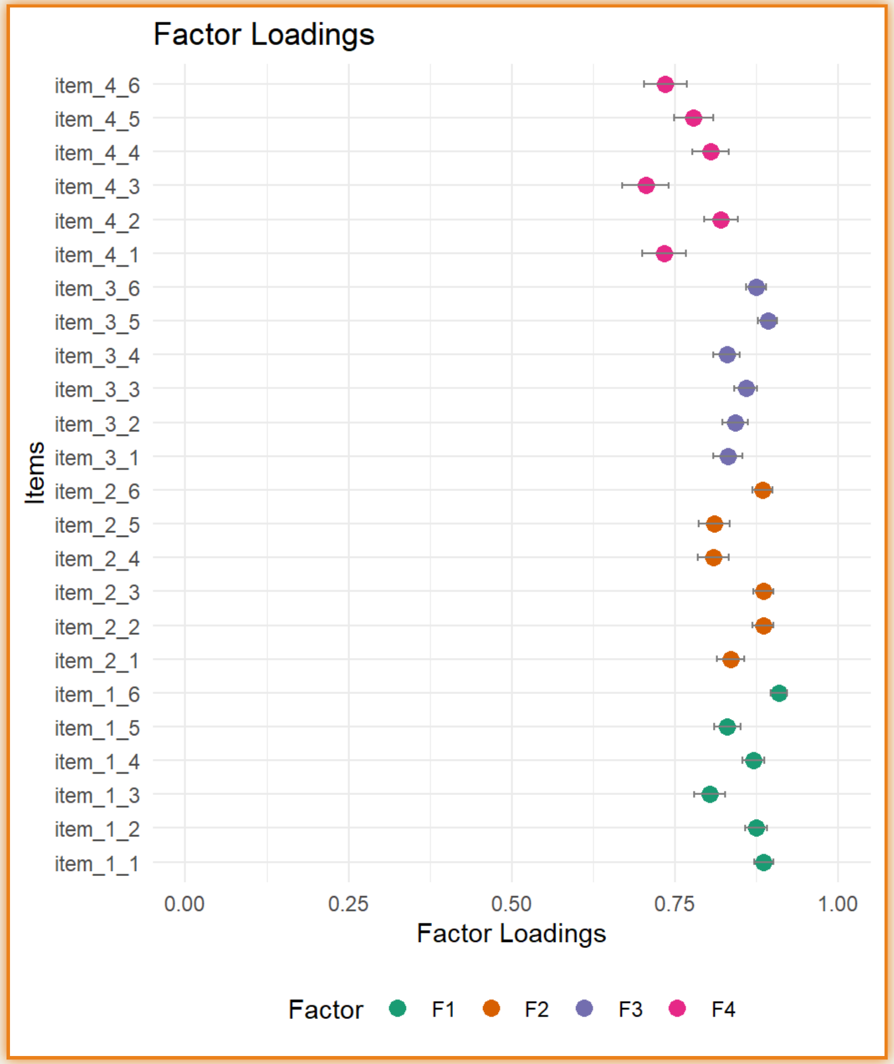
Ejemplificación

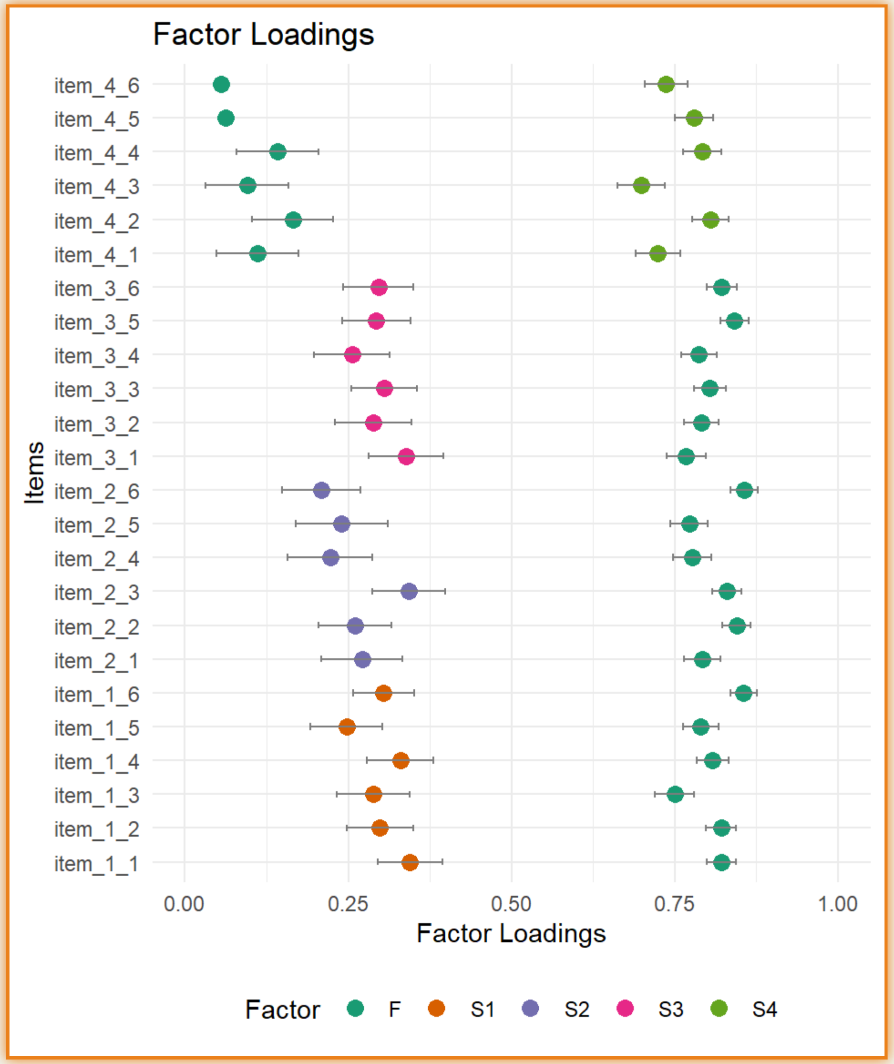
Con una muestra suficiente, 4 factores con 6 ítems en cada factor y un buen comportamiento bifactorial.

```r
result_1f <- simulaBifactor(
  sampleSize = 1000,
  nFactors = 4,
  itemsPerFactor = 6,
  loadingGeneral = 0.8,
  loadingSpecific = 0.3,
  fluctuation = 0.05,
  type_problematic = "1f",
  mod_gen_factor = 0.15,
  mod_spec_factor = 2.5,
  estimator = "MLR"
)
```



Organización de las cargas factoriales





Obtención de Índices

```
bifactorIndices(result_1f$fit_bifactor, UniLambda = result_1f$fit_uni)
```

```
$ModelLevelIndices
     ECV.F        PUC    Omega.F   OmegaH.F       ARPB
0.70564532 0.78260870 0.97275756 0.86241389 0.03276995

$FactorLevelIndices
       ECV_SS     ECV_SG     ECV_GS     Omega     OmegaH         H        FD
F  0.70564532 0.70564532 0.70564532 0.9727576 0.86241389 0.9721486 0.9705498
S1 0.12377861 0.03305574 0.87622139 0.9459541 0.11618284 0.3798054 0.7092911
S2 0.09359639 0.02442729 0.90640361 0.9412326 0.08600959 0.3071786 0.6467970
S3 0.12061010 0.03162232 0.87938990 0.9423015 0.11307017 0.3680375 0.6925122
S4 0.97843528 0.20524934 0.02156472 0.8939206 0.87694498 0.8925204 0.9461048

$ItemLevelIndices
                IECV  RelParBias
item_1_1 0.850379317 0.028620201
```

Desajuste del unidimensional

```r
summary(result_1f$fit_uni, fit.measures = TRUE, standardized = TRUE)
```

```
lavaan 0.6-19 ended normally after 27 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        48

  Number of observations                          1000

Model Test User Model:
                                              Standard      Scaled
  Test Statistic                              4604.696    4604.680
  Degrees of freedom                               252         252
  P-value (Chi-square)                           0.000       0.000
  Scaling correction factor                                  1.000
    Yuan-Bentler correction (Mplus variant)
```

Comparación de ajustes

```r
psymetrics::compare_model_fit(
  result_1f$fit_uni,
  result_1f$fit_multi,
  result_1f$fit_bifactor
)
```

```
MODEL                  | NOBS | ESTIMATOR | NPAR |   Chi2   | Chi2_df
----------------------------------------------------------------------
result_1f$fit_uni      | 1000 |    MLR    |  48  | 4604.680 |   252
result_1f$fit_multi    | 1000 |    MLR    |  54  | 284.945  |   246
result_1f$fit_bifactor | 1000 |    MLR    |  72  | 245.710  |   228

MODEL                  | p (Chi2) |  CFI  |  TLI  | RMSEA |   RMSEA  CI    | SRMR
----------------------------------------------------------------------------------
result_1f$fit_uni      |  < .001  | 0.794 | 0.775 | 0.131 | [0.128, 0.135] | 0.131
result_1f$fit_multi    |  0.045   | 0.998 | 0.998 | 0.013 | [0.002, 0.019] | 0.022
result_1f$fit_bifactor |  0.201   | 0.999 | 0.999 | 0.009 | [0.000, 0.016] | 0.015
```

References

Rodriguez, Anthony, Steven P. Reise, y Mark G. Haviland. 2016. «Applying Bifactor Statistical Indices in the Evaluation of Psychological Measures». *Journal of Personality Assessment* 98 (3): 223-37. https://doi.org/10.1080/00223891.2015.1089249.