Análisis Factorial Confirmatorio y Confiabilidad en `R`

3er CIEP - Usil

Brian N. Peña-Calero y Arnold Tafur-Mendoza

Avances en Medición Psicológica (AMP) - UNMSM

12/10/2019

Temario

Temario

1. Modelos Factoriales
    - Definición y utilidad
    - Conceptos: cargas factoriales, errores, variables observables, variables latentes.
2. Análisis Factorial Confirmatorio
    - Diferencias y usos entre el análisis factorial exploratorio y confirmatorio
    - Concepto y utilidad
    - Términos básicos: Estimación restringidas (correlación de errores), liberadas, fijas
    - Modelos AFC conocidos: Primer orden (ortogonales y oblicuos), segundo orden, bifactor
    - Procedimientos de análisis: Especificación, Identificación, Estimación, evaluación y re-especificación
    - Ejemplificación con caso
3. Fiabilidad por consistencia interna
    - Problemática de Alfa y la tau-equivalencia
    - Ventajas de la fiabilidad basado en modelos factoriales
    - Procedimiento en software

Modelos Factoriales

Modelos Factoriales

Definición y Utilidad

Conceptos

- Cargas Factoriales
- Errrores
- Variables Observables
- Variables Latentes

Análisis Factorial Confirmatorio

Análisis Factorial Confirmatorio

- Diferencias y usos entre el análisis factorial exploratorio y confirmatorio
- Concepto y utilidad
- Términos básicos: Estimación restringidas (correlación de errores), liberadas, fijas
- Modelos AFC conocidos: Primer orden (ortogonales y oblicuos), segundo orden, bifactor

Procedimientos de análisis

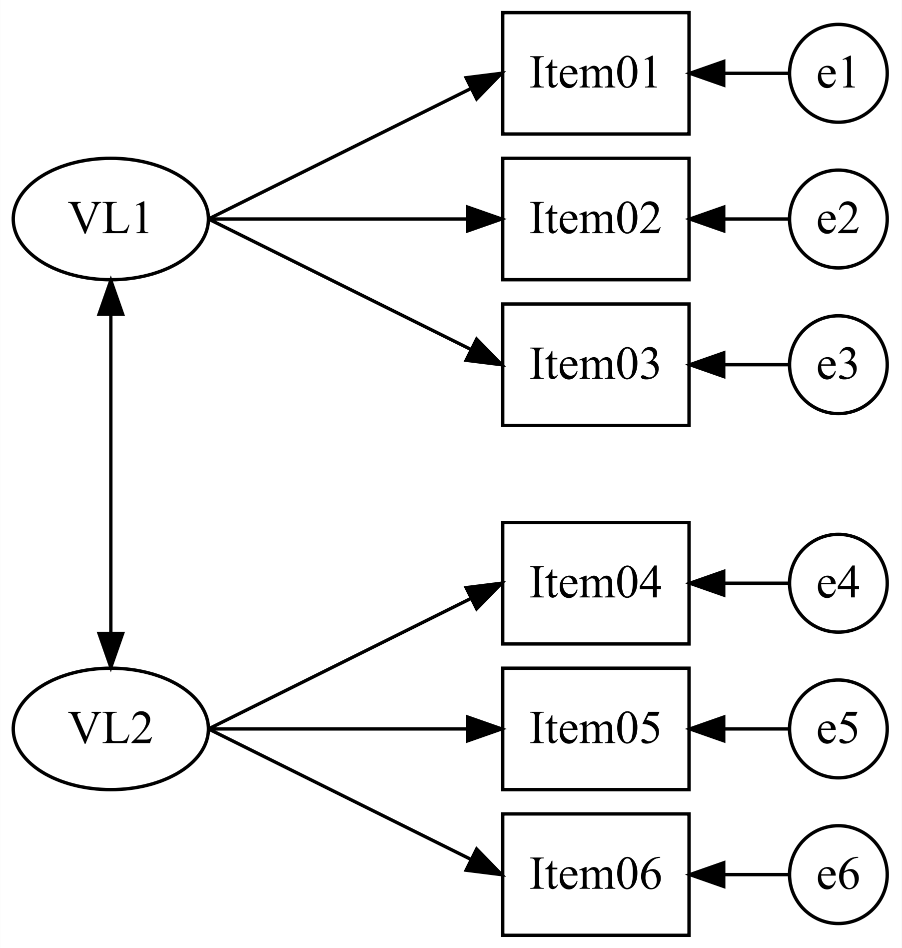
1. Especificación
2. Identificación
3. Estimación
4. Evaluación
5. Re-especificación

Procedimientos de análisis

1. Especificación

Procedimientos de análisis

1. Especificación



El símbolo `↔` hace referencia a la ***covarianza*** entre los factores.

```
model ← " # Modelo de Medición
          VL1 =~ Item01 + Item02 + Item03
          VL2 =~ Item04 + Item05 + Item06
          # VL1 ~~ VL2
         "
```

No es necesario especificar `VL1 ~~ VL2`. `lavaan` supone que siempre trabajamos con modelos oblicuos.

Nota: *Más adelante veremos modelos ortogonales.*

Procedimientos de análisis

1. Especificación

Hay más cosas que podemos especificar en nuestros modelos de medición:

- Igualar las cargas factoriales de alguno o varios ítems

```
VL1 =~ a*Item01 + a*Item02 + a*Item03
```

- Especificar modelos ortogonales

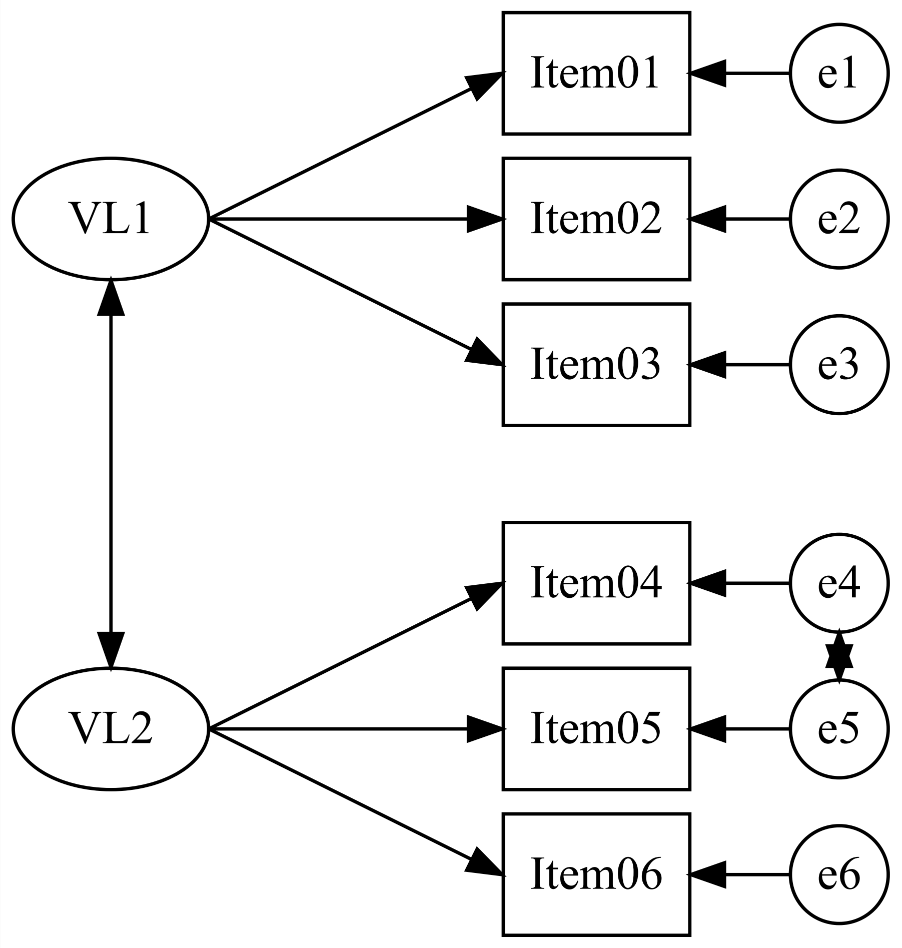
```
VL1 ~~ 0*VL2
```

- Realizar correlación entre errores (covarianza de varianza específica de los ítems)

```
# Correlación de errores
  Item01 ~~ Item02
```

Procedimientos de análisis

1. Especificación



¡ A resolverlo!

```
model ← " # Modelo de Medición
          VL1 =~ Item01 + Item02 + Item03
          VL2 =~ Item04 + Item05 + Item06

          # Correlación de errores
          Item04 ~~ Item05
         "
```

La varianza específica del `Item04` y del `Item05` son los que se relacionan de alguna manera. Existe algo que distinto a `VL2` que está explicando el comportamiento de las puntuaciones de esos ítems.

Procedimientos de análisis

2. Identificación

Este procedimiento hace referencia a la suficiencia de información para el análisis. Dependiendo de la cantidad de información que tengamos y que solicitemos, podremos encontrar una solución satisfactoria. Por ejemplo:

> **X + Y = 20**
> **2X + Y = 28**

¿Cuánto vale `X` y cuánto vale `Y`?

`X = 8`

`Y = 12`

¿Había suficiente información?

Procedimientos de análisis

2. Identificación

Imaginemos ahora la siguiente situación:

> **X + Y = 25**
>
> **2X + 2Y = 50**

¿Cuánto vale `X` y cuánto vale `Y`?

-- Las soluciones pueden ser infinitas

`X = 10; Y = 15`

`X = 8; Y = 17`

`X = 15; Y = 10`

¡No tenemos suficiente información!

Procedimientos de análisis

2. Identificación

En el caso de un `AFC` la suficiencia de información hace referencia a la cantidad de correlaciones que existe en una matriz de las variables a analizar (en nuestros casos, los ítems). Esto se contrasta con la información solicitada (`Especificaciòn`).

```
##
## Correlation method: 'pearson'
## Missing treated using: 'pairwise.complete.obs'

##   rowname   x1   x2   x3
## 1      x1 1.00
## 2      x2  .30 1.00
## 3      x3  .44  .34 1.00
```

¿Cuántas correlaciones tenemos?

`6`

Procedimientos de análisis

2. Identificación

Veamos otro ejemplo con una cantidad de ítems mayor

```
##
## Correlation method: 'pearson'
## Missing treated using: 'pairwise.complete.obs'
```

```
##   rowname   x1   x2   x3   x4   x5
## 1      x1 1.00
## 2      x2  .30 1.00
## 3      x3  .44  .34 1.00
## 4      x4  .37  .15  .16 1.00
## 5      x5  .29  .14  .08  .73 1.00
```

¿Cuántas correlaciones tenemos?

`15`

Procedimientos de análisis

2. Identificación

¡Último ejemplo! *Nota: Encuentren la regla*

```
##
## Correlation method: 'pearson'
## Missing treated using: 'pairwise.complete.obs'

##   rowname   x1   x2   x3   x4   x5   x6   x7   x8
## 1      x1 1.00
## 2      x2  .30 1.00
## 3      x3  .44  .34 1.00
## 4      x4  .37  .15  .16 1.00
## 5      x5  .29  .14  .08  .73 1.00
## 6      x6  .36  .19  .20  .70  .72 1.00
## 7      x7  .07 -.08  .07  .17  .10  .12 1.00
## 8      x8  .22  .09  .19  .11  .14  .15  .49 1.00
```

¿Cuántas correlaciones tenemos?

`36`

$$Información = \frac{n(n+1)}{2}$$

Procedimientos de análisis

2. Identificación

2. Identificación

Notas importantes

- Mientras tengamos más de 3 ítems en nuestro modelo, no tendremos problemas de identificación (*sub-identificación* o *no identificado*).
- En los casos que tenemos modelos con solo 3 ítems, se puede calcular siempre y cuando no se soliciten mayor información (*apenas identificado*) como por ejemplo, `correlación de errores entre ítems`. Debido a que con 3 ítems tenemos 6 cantidad de informaciones y estaríamos solicitando un total de 7 informaciones.
- No es imposible trabajar con modelos de 2 o 1 ítem (*ítems únicos*), siempre y cuando se empleen algunos artificios. Por ejemplo, igualar cargas factoriales o ingresar información previa sobre algún parámetro (de esta manera no se volverá a calcular).

Procedimientos de análisis

3. Estimación

Tabla Resumen de Estimadores

Estimadores	Estimadores Robustos	Descripciones
ML (Máxima Verosimilitud)	MLM, MLR, MLMVS, MLMV	Datos continuos
ULS (Mínimos cuadrados no ponderados)	ULSM, ULSMVS, ULSMV	Variables Categóricas
WLS (Mínimos cuadrados ponderados)	-	Variables Categóricas
DWLS (Mínimos cuadrados con diagonal ponderada	WLSM, WLSMVS, WLSMV	Variables Categóricas. Es el más recomendado en la actualidad

Procedimientos de análisis

3. Estimación

Aclaración sobre los sufijos de las nomenclaturas para los **estimadores robustos**:

- ABC`M`: Trabaja con errores robustos y corrección para chi-cuadrado *Satorra-Bentler*
- ABC`MVS`: Trabaja con errores robustos y corrección para media-varianza y chi-cuadrado *Satterthwaite*
- ABC`MV`: Trabaja con errores robustos y corrección para media-varianza y chi-cuadrado *scale-shifted*

Los estudios de simulación coinciden en que la corrección **scale-shifted** es quien brinda mejores resultados siempre y cuando se tenga una cantidad de datos *suficiente* (n > 250, dependiendo de la cantidad de ítems).
Para nuestros casos: Los mejores estimadores a utilizar cuando trabajemos con datos contínuos será `MLR` (en caso de no-normalidad) y `WLSMV` (en caso de datos categóricos).

4. Evaluación

Este procedimiento hace referencia al cálculo y valoración de los ***índices de ajuste*** de los modelos estimados, así como a las cargas factoriales calculadas.

- Estos índices son derivados del test *chi-cuadrado* por lo que a medida que cambie este, el resto de los índices cambiará.
- A medida que se emplee un estimador diferente (*recordar que hace correcciones a esta prueba*) y/o aumenten o desciendan el número de ítems/factores, los índices de ajuste cambiarán.
- Por último a medida que se ingresen o resten especificaciones al modelo, los **índices de ajuste** cambiarán.

Este procedimiento nos permitirá tomar la decisión de finalizar el análisis en este punto o ir al siguiente paso ***Re-Especificación***.

Procedimientos de análisis

4. Evaluación

Índices de ajuste	Descriptivos	Puntos de Corte
χ^2	Test Chi-Cuadrado (bondad de ajuste)	No estadísticamente significativo
χ^2/gl	Medida de parsimonia	Menor a 3 o 2
CFI (Comparative Fit Index) TLI (Tucker Lewis Index) o NNFI (Non-Normed Fit Index)	Medida de ajuste independiente o incremental	$\geq .90$ = ajuste adecuado; $\geq .95$ = buen ajuste
RMSEA (Root Mean Square Error of Approximation)	Evalúa que tan lejos está de un modelo perfecto	$\geq .10$ = ajuste pobre; $\leq .08$ = adecuado ajuste, $\leq .05$ = buen ajuste
SRMR (Standardized Root Mean Residual)	Evalúa que tan grande es el error de reproducir el modelo	$\leq .08$ = buen ajuste, $\leq .06$: ideal

Procedimientos de análisis

5. Re-Especificación

Este procedimiento hace referencia a empezar nuevamente el flujo del análisis, quitar, aumentar algo en el proceso de `Espeficación` que permita tener un modelo factorial idóneo.

Una de las cosas que ayudan en esta etapa es el cálculo de los `índices de modificación`. En el paquete `lavaan` podremos realizar rápidamente con la sintaxis:

```
modificationindices()
```

¿Listos para hacer un ejemplo?

Ejemplificación de caso

Para este primer ejemplo usaremos la BD de Holzinger & Swineford (1939). Se tienen 9 pruebas que se estructuran en 3 factores latentes: Visual, Textual y Velocidad.

```
lavaan::HolzingerSwineford1939
```

Show 4 entries Search:

	id	sex	ageyr	agemo	school	grade	x1	x2	x3	x4	x5	x6	x7	x8	x9
1	1	1	13	1	Pasteur	7	3.33	7.75	0.38	2.33	5.75	1.29	3.39	5.75	6.36
2	2	2	13	7	Pasteur	7	5.33	5.25	2.12	1.67	3	1.29	3.78	6.25	7.92
3	3	2	13	1	Pasteur	7	4.5	5.25	1.88	1	1.75	0.43	3.26	3.9	4.42
4	4	1	13	2	Pasteur	7	5.33	7.75	3	2.67	4.5	2.43	3	5.3	4.86

Showing 1 to 4 of 301 entries

Previous 1 2 3 4 5 … 76 Next

Ejemplificación de caso

1. Especificación

```
model ← " visual  =~ x1 + x2 + x3
          textual =~ x4 + x5 + x6
          speed   =~ x7 + x8 + x9 "
```

2. Identificación y Estimación

Almacenamos la estimación en el objeto `fit` sobre el cuál consultaremos para obtener información.

Recordar: Es importante el almacenar información con `←`

```
library(lavaan)
```

```
## This is lavaan 0.6-6
```

```
## lavaan is BETA software! Please report any bugs.
```

Ejemplificación de caso

4. Evaluación

```
summary(fit)
```

```
## lavaan 0.6-6 ended normally after 35 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of free parameters                         21
##
##   Number of observations                           301
##
## Model Test User Model:
##
##   Test statistic                                85.306
##   Degrees of freedom                                24
##   P-value (Chi-square)                           0.000
##
## Parameter Estimates
```

Ejemplificación de caso

4. Evaluación

```
summary(fit, fit.measures = TRUE, standardized = TRUE)
```

```
## lavaan 0.6-6 ended normally after 35 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of free parameters                         21
##
##   Number of observations                           301
##
## Model Test User Model:
##
##   Test statistic                                85.306
##   Degrees of freedom                                24
##   P-value (Chi-square)                           0.000
##
```

5. Re-Especificación

```r
modindices(fit, sort = TRUE, maximum.number = 10)
```

```
##         lhs op rhs     mi    epc sepc.lv sepc.all sepc.nox
## 30   visual =~  x9 36.411  0.577   0.519    0.515    0.515
## 76       x7 ~~  x8 34.145  0.536   0.536    0.859    0.859
## 28   visual =~  x7 18.631 -0.422  -0.380   -0.349   -0.349
## 78       x8 ~~  x9 14.946 -0.423  -0.423   -0.805   -0.805
## 33  textual =~  x3  9.151 -0.272  -0.269   -0.238   -0.238
## 55       x2 ~~  x7  8.918 -0.183  -0.183   -0.192   -0.192
## 31  textual =~  x1  8.903  0.350   0.347    0.297    0.297
## 51       x2 ~~  x3  8.532  0.218   0.218    0.223    0.223
## 59       x3 ~~  x5  7.858 -0.130  -0.130   -0.212   -0.212
## 26   visual =~  x5  7.441 -0.210  -0.189   -0.147   -0.147
```

Fiabilidad por consistencia interna

Fiabilidad por consistencia interna

- Problemática de Alfa y la tau-equivalencia
- Procedimiento en software

Gracias