

# Métodos cuantitativos II: Modelos de ecuaciones estructurales

Jose M Sallan jose.maria.sallan@upc.edu

## Contenidos



Elementos de un modelo de ecuaciones estructurales

Ajustando un modelo SEM



## Elementos de un modelo de ecuaciones estructurales

Ajustando un modelo SEM

## Modelos de ecuaciones estructurales



El objetivo de los modelos de ecuaciones estructurales (SEM) es evaluar el ajuste con los datos de un modelo estructural de relaciones entre variables

#### SEM permite:

- Trabajar con variales latentes y observables
- Determinar indices de ajuste (fit indices) de significación global del modelo
- Obtener intervalos de confianza y niveles de significación de los parámetros del modelo

## Variables latentes y observables



- ► En investigación sobre organizaciones es habitual representar constructos como variables latentes
- Las variables latentes se miden usando dos o másvariables observables
- Es habitual que las variables observables incorporen errores de medida
- Notación:
  - ▶ Variables observables: □
  - ► Variables latentes: ○

## Ejemplo: modelo con dos variables latentes



#### Supongamos que queremos medir:

- ightharpoonup Compromiso organizativo con los ítemes  $v_1$ ,  $v_2$  y  $v_3$
- Satisfacción laboral con los ítemes v<sub>4</sub> y v<sub>5</sub>

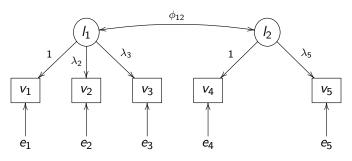
#### Entonces tenemos:

- ▶ Variables observables  $v_1$  a  $v_5$
- ▶ Variables latentes l₁ (compromiso organizativo) y l₂ (satisfacción laboral)

## Ejemplo: modelo con dos variables latentes



Repesentación del modelo estructural



Los parámetros iguales a 1 escalan la variable latente Este modelo es un ejemplo de análisis factorial confirmatorio

## Ejemplo: modelo con tres variables latentes

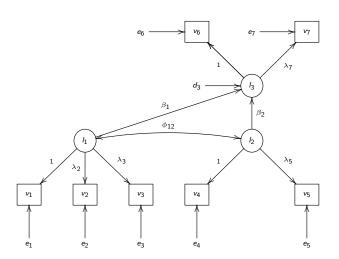


#### Incrementemos la complejidad del modelo añadiendo:

- ► El constructo intención de abandonar, medido con los ítemes v<sub>6</sub> y v<sub>7</sub> y la variable latente l<sub>3</sub>
- Definiendo relaciones directas entre la intención de abandonar l<sub>3</sub> y las otras dos variables latentes (compromiso organizativo l<sub>1</sub> y satisfacción laboral l<sub>2</sub>)

# Ejemplo: modelo con tres variables latentes





## Ejemplo: modelo con tres variables latentes



Modelo de medida y modelo estructural

## El ejemplo tiene dos submodelos:

- Modelo de medida: relaciones entre variables latentes y observables
- ► Modelo estructural: relaciones entre variables latentes

Si todas las variables son observables tenemos un modelo de análisis de caminos (path analysis)

Si el modelo estructural se reduce a covarianzas entre variables latentes tenemos un modelo de análisis factorial confirmatorio

#### Parámetros de un modelo SEM



#### Dos tipos de variables:

- Exógenas: no dependen de otras variables (e.g., l<sub>1</sub> y l<sub>2</sub>, y los términos de error e del modelo anterior)
- ► Endógenas: dependen de otras variables

#### Parámetros del modelo:

- $\blacktriangleright$  Coeficientes de las relaciones en los modelos de medida  $\lambda$  y estructural  $\beta$
- Varianzas de términos de error en las variables endógenas
- Varianza de las variabes exógenas
- Covarianzas existentes entre variables exógenas

## Grados de libertad de un modelo estructural



Los grados de libertad df de un modelo es la diferencia entre:

- ▶ El número de elementos únicos de la matriz de covarianzas (p+1) p/2, donde p es el número de variables observables
- ► El número de parámetros del modelo

Para que pueda ser estinado, el modelo ha de tener grados de libertad positivos (modelo especificado)

# Grados de libertad para el modelo de dos vari

Elementos diferentes de la matriz de covarianzas: (5+1)5/2 = 15 Parámetros del modelo:

- ▶ Varianzas de variables exógenas  $l_1$  y  $l_2$ , y  $e_1$  a  $e_5$ : 2 + 5 = 7
- ▶ Coeficientes  $\lambda_2$ ,  $\lambda_3$  y  $\lambda_5$ : 3
- ► Covarianzas entre *l*<sub>1</sub> y *l*<sub>2</sub>: 1

Grados de libertad: 15 - (7 + 3 + 1) = 4

# Grados de libertad para el modelo con tres va

Elementos diferentes de la matriz de covarianzas: (7 + 1)7/2 = 28 Parámetros del modelo:

- ▶ Varianzas de las variables exógenas  $l_1$  y  $l_2$ ,  $e_1$  a  $e_7$  y  $d_3$ : 2 + 7 + 1 = 10
- ▶ Coeficientes  $\lambda_2$ ,  $\lambda_3$ ,  $\lambda_5$  and  $\lambda_7$ : 4
- ▶ Coeficientes  $\beta_1$  y  $\beta_2$ : 2
- ightharpoonup Covarianza entre  $l_1$  y  $l_2$ : 1

Grados de libertad: 28 - (10 + 4 + 2 + 1) = 11



Elementos de un modelo de ecuaciones estructurales

Ajustando un modelo SEM

## Software para ajustar modelos SEM



- Software comercial: AMOS (extensión de SPSS), EQS, Mplus
- Software de código abierto: paquetes sem y lavaan de R
- Gratuito (pero no abierto): Mx (integrado en R mediante OpenMx)

Usaremos el paquete lavaan de R http://lavaan.ugent.be/

# Ejemplo: datos de Holzinger y Swineford



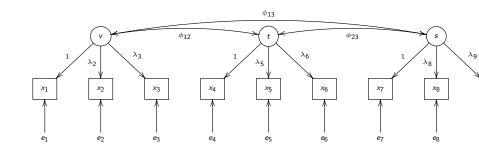
El dataset HolzingerSwineford1939 consiste en puntuaciones de test de capacidad mental de niños de séptimo y octavo grado de dos escuelas diferentes (Pasteur and Grant-White):

- Los ítemes  $x_1$ ,  $x_2$  y  $x_3$  miden la capacidad visual v
- ▶ Los ítemes x<sub>4</sub>, x<sub>5</sub> y x<sub>6</sub> miden la capacidad textual t
- Los ítemes  $x_7$ ,  $x_8$  y  $x_9$  miden la habilidad de velocidad s

Realicemos una prueba para un modelo de análisis factorial confirmatorio CFA para evaluar el ajuste entre variables latentes y observables

# Ejemplo: datos de Holzinger y Swineford





# Especificando el modelo en lavaan



- ► Las relaciones del modelo de medida se especifican con el símbolo =~
- Las convarianzas se definen como parámetros por defecto

- ▶ Varianzas y covarianzas se representan con ~~
- lacktriangle Las relaciones estructurales se representan con  $\sim$

# Especificando el modelo en lavaan



#### El mismo modelo, igualando las covarianzas a cero:



- En este caso, el modelo puede ajustarse usando la función cfa
- Por defecto, lavaan estima modelos igual que Mplus. Para obtener la estimación de AMOS, EQS o LISREL, se ha de introducir el parámetro likelihood=wishart en la función
- > library(lavaan)
- > holz01.fit <- cfa(holz01, data = HolzingerSwineford1939,
- + likelihood="wishart")

#### Resultados del modelo:

- summary: estimadores de parámetros y una selección de índices de ajuste
- fitMeasures: un vector con una colección de índices de ajuste
- parameterEstimates: un data frame con estimadores de los parámetros

#### índices de ajuste con lavaan



#### > fitMeasures(holz01.fit)

npar	fmin	chisq	df
21.000	0.142	85.022	24.000
pvalue	baseline.chisq	baseline.df	baseline.pvalue
0.000	915.799	36.000	0.000
cfi	tli	nnfi	rfi
0.931	0.896	0.896	0.861
nfi	pnfi	ifi	rni
0.907	0.605	0.932	0.931
logl	unrestricted.logl	aic	bic
-3742.252	-3699.600	7526.505	7604.284
ntotal	bic2	rmsea	rmsea.ci.lower
301.000	7537.685	0.092	0.071
rmsea.ci.upper	rmsea.pvalue	rmr	rmr_nomean
0.114	0.001	0.082	0.082
srmr	srmr_bentler	srmr_bentler_nomean	srmr_bollen
0.065	0.065	0.065	0.065
srmr_bollen_nomean	srmr_mplus	srmr_mplus_nomean	cn_05
0.065	0.065	0.065	129.490
cn_01	gfi	agfi	pgfi
152.654	0.943	0.894	0.503
mfi	ecvi		
0.903	0.423		





#### > head(parameterEstimates(holz01.fit))

	lhs	ор	rhs	est	se	z	pvalue	ci.lower	ci.upper
1	v	=~	x1	1.000	0.000	NA	NA	1.000	1.000
2	v	=~	x2	0.554	0.100	5.544	0	0.358	0.749
3	v	=~	x3	0.729	0.109	6.674	0	0.515	0.944
4	t	=~	x4	1.000	0.000	NA	NA	1.000	1.000
5	t	=~	x5	1.113	0.066	16.986	0	0.985	1.242
6	t	=~	x6	0.926	0.056	16.675	0	0.817	1.035



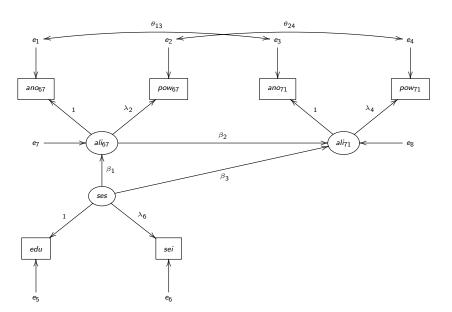
Algunos valores de índices de ajuste en SEM

Chi al cuadradao $\chi^2$	chisq	_
Ratio chi al cuadradao entre grados de libertad $\chi^2/df$	_	$\sim 2$
Índice de ajuste no-normado (NNFI) or índice Tucker-Lewis (TLI)	nnfi, tli	> 0,85
Raíz cuadrada del error de aproximación (RMSEA)	rmsea	< 0,07
Raźi cuadrada media de residuos estandarizados (SRMR)	srmr	< 0,08

Para más información:

http://davidakenny.net/cm/fit.htm





# UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH Fundació Politècnica de Catalunya

Especificación con lavaan

```
> wheaton.model <-'
+ ali67 = ano67 + pow67
+ ali71 = ano71 + pow71
+ ses = edu + sei
+ ali71 ali67 + ses
+ ali67 ses
+ ano67 ano71
+ pow67 pow71'</pre>
```

#### UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH Fundació Politècnica de Catalunya

#### Entrada de datos y ajuste del modelo

#### UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH Fundació Politècnica de Catalunya

#### Valores de índices de ajuste

#### > fitMeasures(fit.wheaton)

npar	fmin	chisq	df
17.000	0.003	4.735	4.000
pvalue	baseline.chisq	baseline.df	baseline.pvalue
0.316	2133.722	15.000	0.000
cfi	tli	nnfi	rfi
1.000	0.999	0.999	0.992
nfi	pnfi	ifi	rni
0.998	0.266	1.000	1.000
logl	unrestricted.logl	aic	bic
-15213.274	-15210.906	30460.548	30542.783
ntotal	bic2	rmsea	rmsea.ci.lower
932.000	30488.792	0.014	0.000
rmsea.ci.upper	rmsea.pvalue	rmr	rmr_nomean
0.053	0.930	0.498	0.498
srmr	srmr_bentler	srmr_bentler_nomean	srmr_bollen
0.007	0.007	0.007	0.007
srmr_bollen_nomean	srmr_mplus	srmr_mplus_nomean	cn_05
0.007	0.007	0.007	1868.387
cn_01	gfi	agfi	pgfi
2614.138	0.998	0.991	0.190
mfi	ecvi		
1.000	0.042		



#### Estimadores de parámetros

#### > head(parameterEstimates(fit.wheaton))

	lhs	op	rhs	est	se	z	pvalue	<pre>ci.lower</pre>	ci.upper
1	ali67	=~	ano67	1.000	0.000	NA	NA	1.000	1.000
2	ali67	=~	pow67	0.979	0.062	15.895	0	0.858	1.099
3	ali71	=~	ano71	1.000	0.000	NA	NA	1.000	1.000
4	ali71	=~	pow71	0.922	0.059	15.498	0	0.805	1.039
5	ses	=~	edu	1.000	0.000	NA	NA	1.000	1.000
6	ses	=~	sei	5.219	0.422	12.364	0	4.392	6.047

## Índices de modificación



- Los índices de modificación nos indican cómo afecta al ajuste del modelo el dejar libre un parámetro fijo
- Al añadir un parámetro, se reducen los grados de libertad, pero también puede reducirse la chi cuadrada

Por ejemplo, para el modelo de Holzinger tenemos:

```
> holz01.mi <- modindices(holz01.fit)
> tail(holz01.mi)
```

```
lhs op rhs mi epc sepc.lv sepc.all sepc.nox
73 x6 ~~
         x7 0.259 -0.020 -0.020
                                 -0.017
                                        -0.017
74
   x6 ~~
         x8 0.275 0.018 0.018 0.016 0.016
75
   x6 ~~
         x9 0.097 -0.011 -0.011 -0.010 -0.010
76
   x7 ~~
         x8 34.145 0.538 0.538 0.488 0.488
   x7 ~~
77
         x9 5.183 -0.187 -0.187 -0.170 -0.170
78
   x8 ~~ x9 14.946 -0.425 -0.425
                               -0.415 -0.415
```

Añadiendo un parámetro de covarianza entre los errores de  $x_7$  y  $x_8$ , se mejora significativamente el ajuste del modelo