KU CENTER FOR RESEARCH METHODS & DATA ANALYSIS

College of Liberal Arts & Sciences

INTRODUCCIÓN AL MODELADO DE ECUACIONES ESTRUCTURALES

MAURICIO GARNIER VILLARREAL

FILOSOFÍA
DOS ESCUELAS DE PENSAMIENTO

Entender las diferencias entre la escuela clásica y la escuela de modelado

Aprender las ventajas y desventajas de cada escuela

DOS ESCUELAS DE INVESTIGACIÓN

Escuela clásica	Escuela de modelado
Diseños sistemáticos (ANOVA)	Diseños representativos (correlaciónales)
Experimental	Ambiental
Enfoca en diferencias de grupo	Diferencias intra e inter sujetos
Prueba relaciones posibles	Prueba relaciones probables
Bien entendido, fácil de usar	Complicado, consume tiempo, costoso

COMPARANDO LAS ESCUELAS

Ambas con importantes y útiles

Las ciencias sociales han favorecido a la escuela clásica

Nuevos avances dan fuerza a la escuela de modelado

- Probar modelos muy sofisticados
- Controlar posibles factores que confundan
- Implicar causalidad (hasta cierto punto)
- Se pueden comparar hipótesis

ESCUELA CLÁSICA: DESVENTAJAS

Presupuestos limitantes (ej., varianzas iguales)

No corrige por el error de medida

- Inconsistente, sesga los resultados
- Error tipo I elevado
- Error tipo II elevado
- Disminuye la posibilidad de generalizar
- Conclusiones sustantivas inexactas

ESCUELA DE MODELADO: DESVENTAJAS

- Complejo de especificar
- No tan comprendido por otros
- Requiere muestras un poco mas grandes
- Los programas no son amigables
- Se enfoca en el ajuste del modelo
- Se puede saber apenas suficiente para ser peligroso

ESCUELA DE MODELADO: VENTAJAS

Se pueden probar efectos directos, indirectos, mediación, moderación, y cualquier combinación

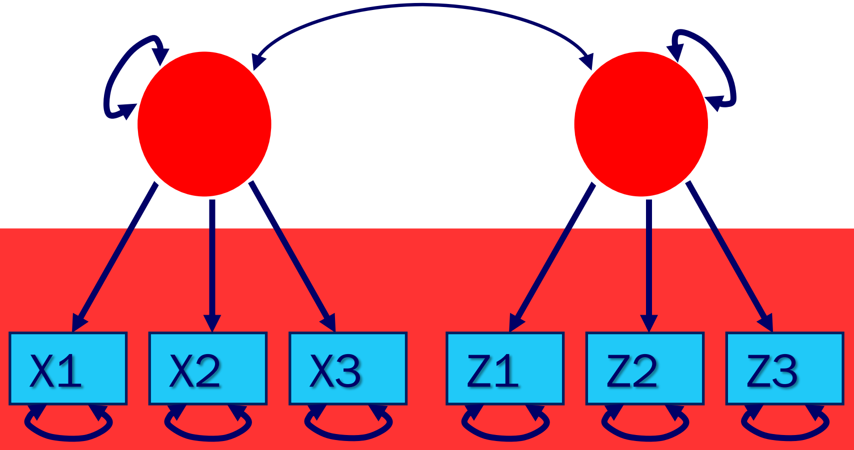
Se puede establecer comparaciones cros-culturales

etc

MODELADO

Conceptos -> Constructos -> Indicadores
Los modelos son dispositivos
de organización constante en
el dialogo entre los datos y
los presupuestos

Los modelos son necesariamente simplificaciones y solo se aproximan a la realidad



MODELOS

Los modelos deben:

Describir el comportamiento transparente y sistemáticamente

Explicar fuerzas que promueven un comportamiento

Predecir cuando y bajo que condiciones ocurre el comportamiento

Mejorar la calidad de vida con el uso del conocimiento

Controlar el comportamiento

Probar todas las anteriores

BUENOS MODELOS

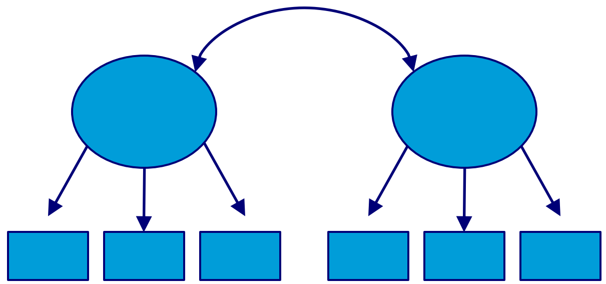
Permiten que el conocimiento se mueva hacia adelante

- Consistente lógica e internamente
- Falseable (se puede probar empíricamente)
- Toma en cuenta los resultados públicos
- Parsimonioso

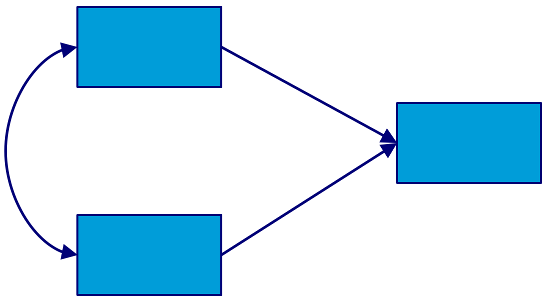
SEM

- Intuición
- Especificación
- Estimación
- Evaluación

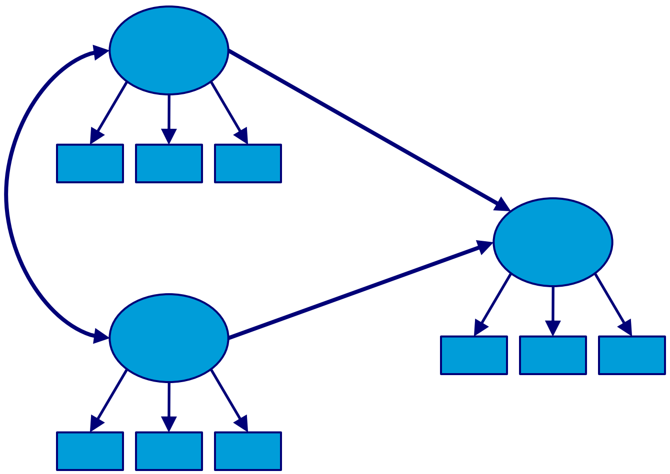
TIPOS BÁSICOS DE MODELOS



Confirmatory Factor Analysis (CFA)



Path Analysis



Structural Equation Model (SEM)

PROGRAMAS

LISREL (Jöreskog & Sörbom; www.ssicentral.com)

M-*plus* (Muthén & Muthén; www.statmodel.com)

R-based SEM packages (e.g., lavaan; OpenMx)

EQS (Bentler; www.mvsoft.com)

PROC CALIS (SAS Institute; www.sas.com)

Amos (Arbuckle; www.spss.com/amos/)

COSAN (Fraser & McDonald)

EzPATH & SEPATH (Steiger)

RAMONA (Browne & Mels)

DEFINICIÓN DE CONSTRUCTOS

Tipos de indicadores y constructos

Convenciones de diagramación

Como se divide la varianza

TIPOS DE INDICADORES

Continuos: mas comunes

Dicotómicos: 2 verdaderas categorías

Dicotomizadas: de un valor continuo dividido en 2 (mala idea)

Politomicas: verdaderas variables ordinales, con mas de 2 categorías

Politomizadas: un valor continuo dividido en varias categorías

VARIANCIA

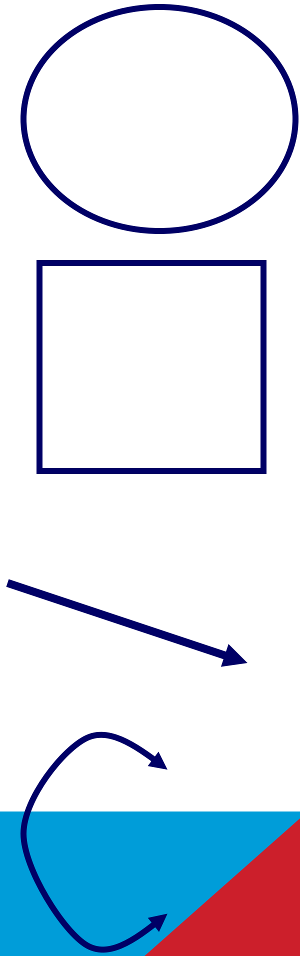
Variables manifiestas:

- $X_i = T_i + S_i + e_i)$

Variables latentes:

- $X_i = C_i + M_i + O_i + (O_{ti} + S_i + e_i)$

CONVENCIONES DE DIAGRAMACIÓN



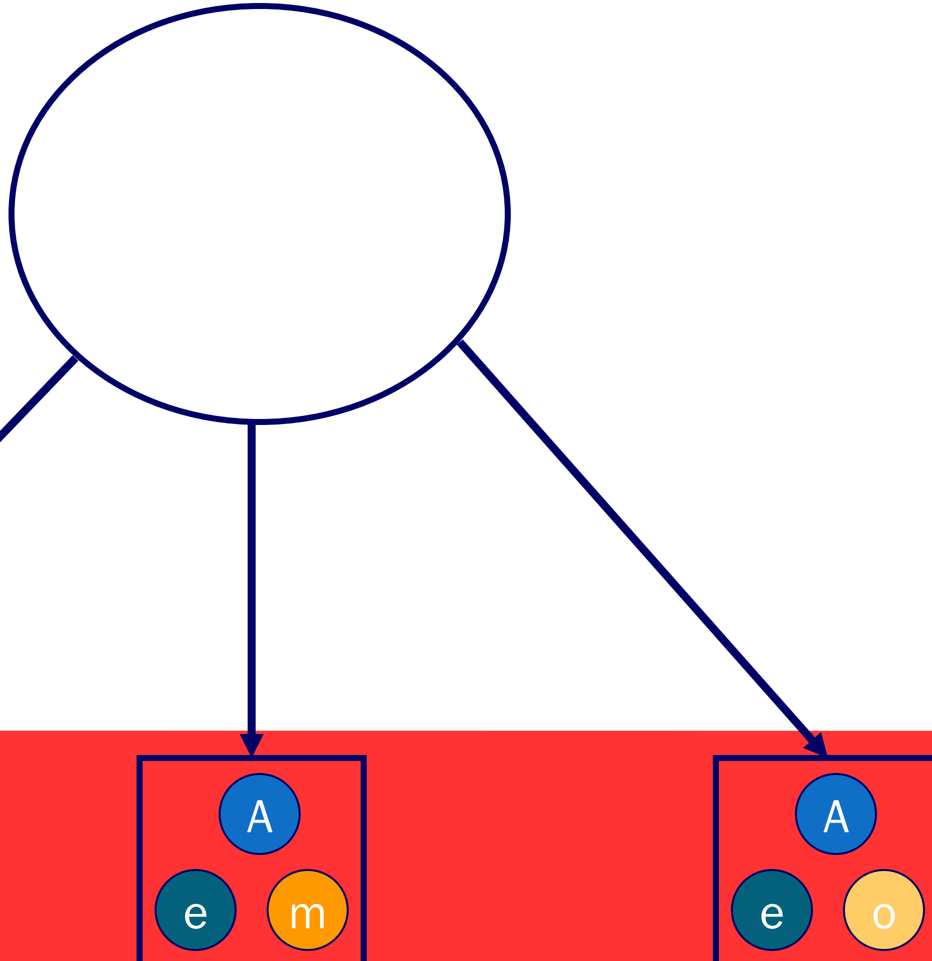
Círculos/óvalos: Constructos latentes

Cuadrados/rectángulo: variables manifiestas (medidas)

Líneas con una flecha: relaciones con una dirección (causal), como regresiones

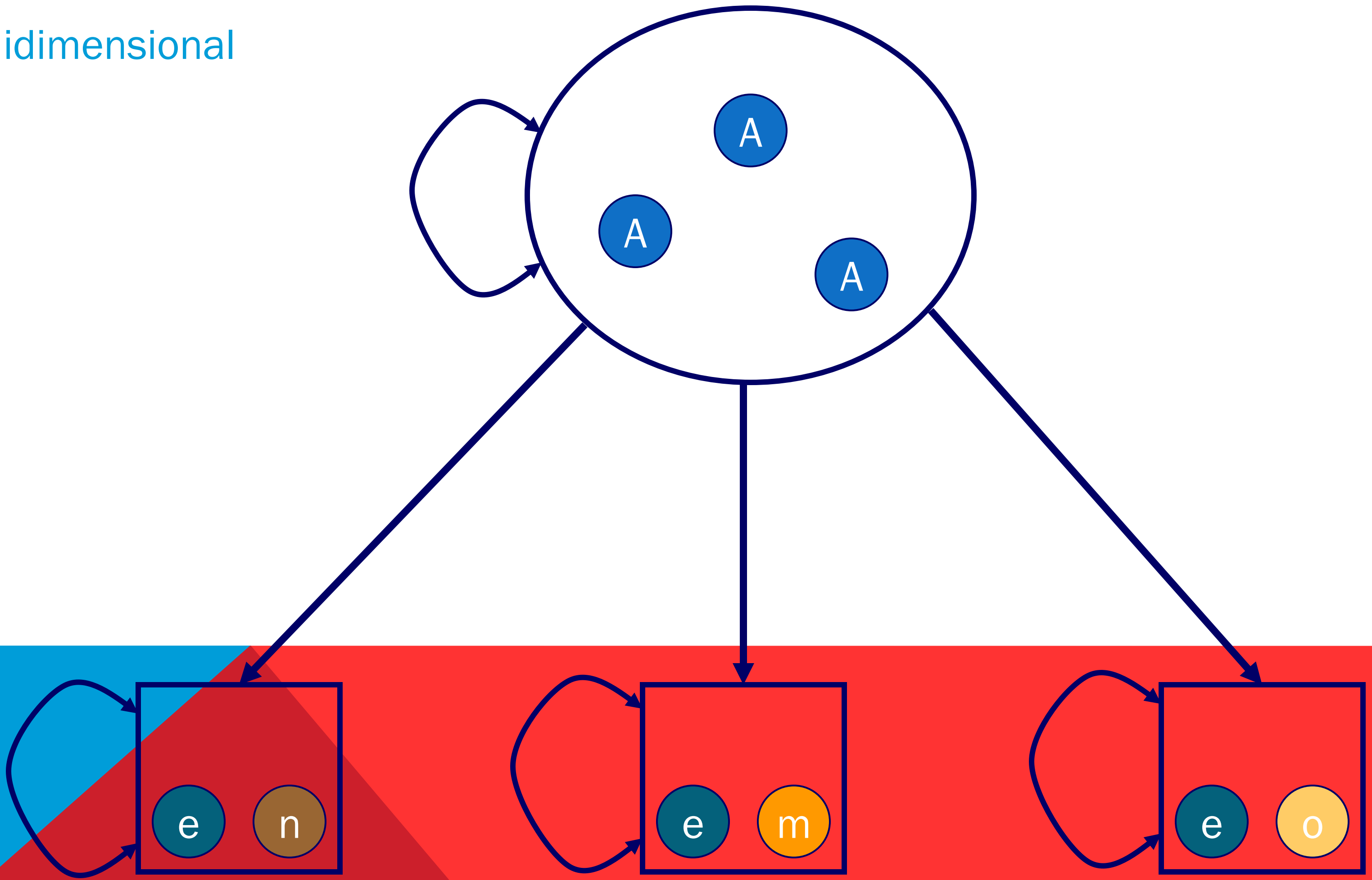
Líneas con doble flecha: relación sin dirección, representa covarianzas y varianzas

DEFINIENDO UN CONSTRUCTO

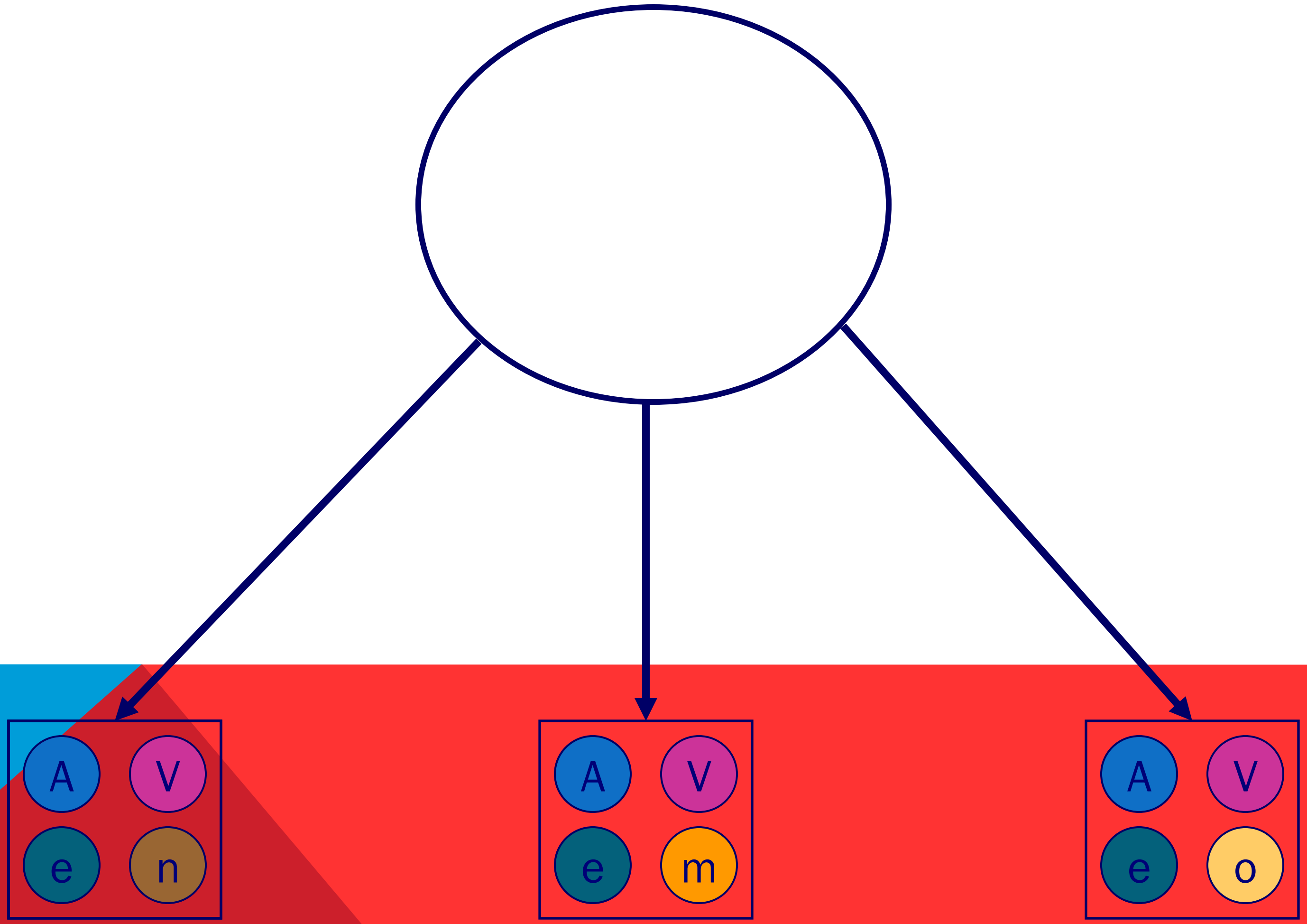


DEFINIENDO UN CONSTRUCTO

Unidimensional

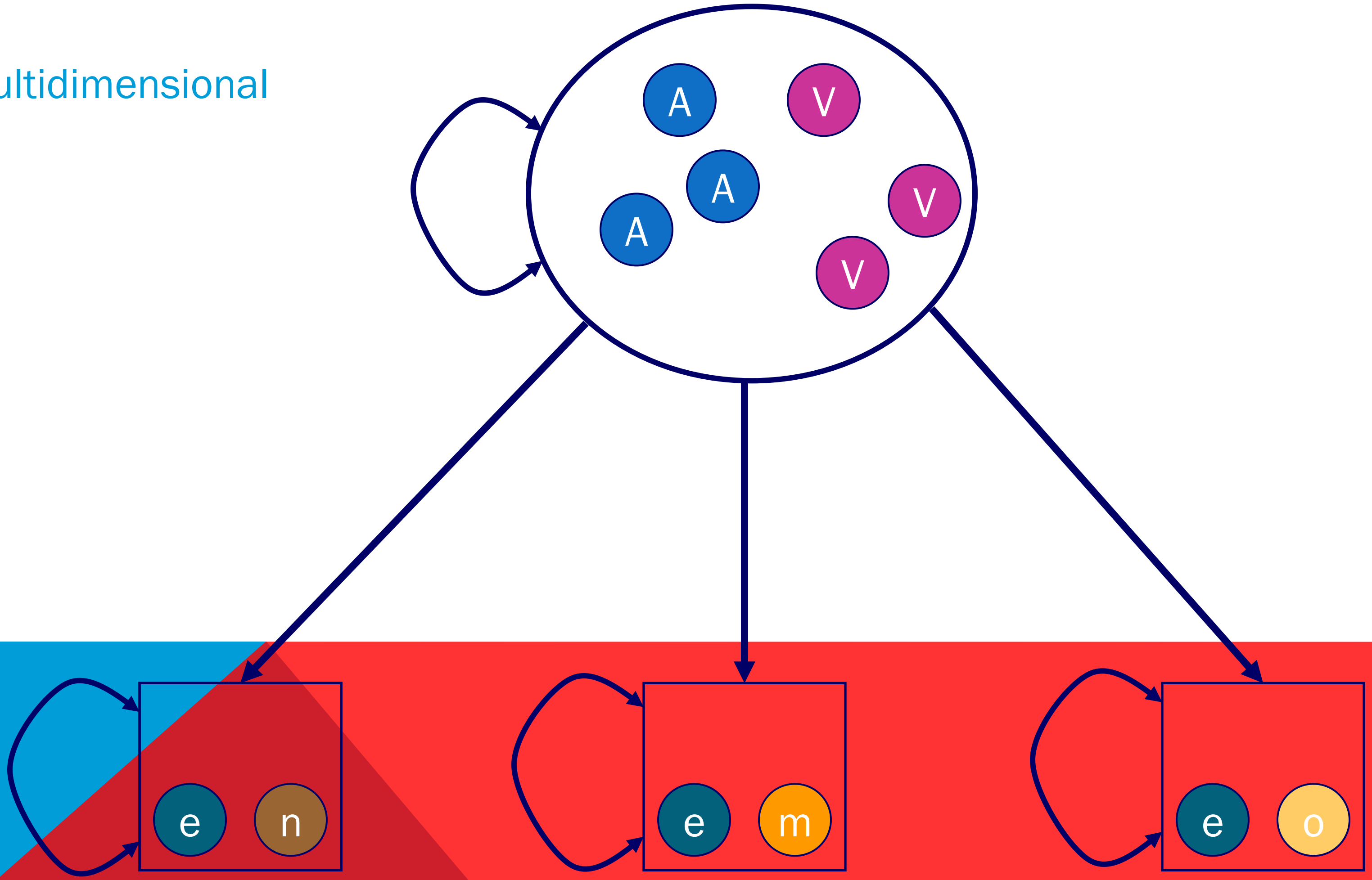


DEFINIENDO UN CONSTRUCTO

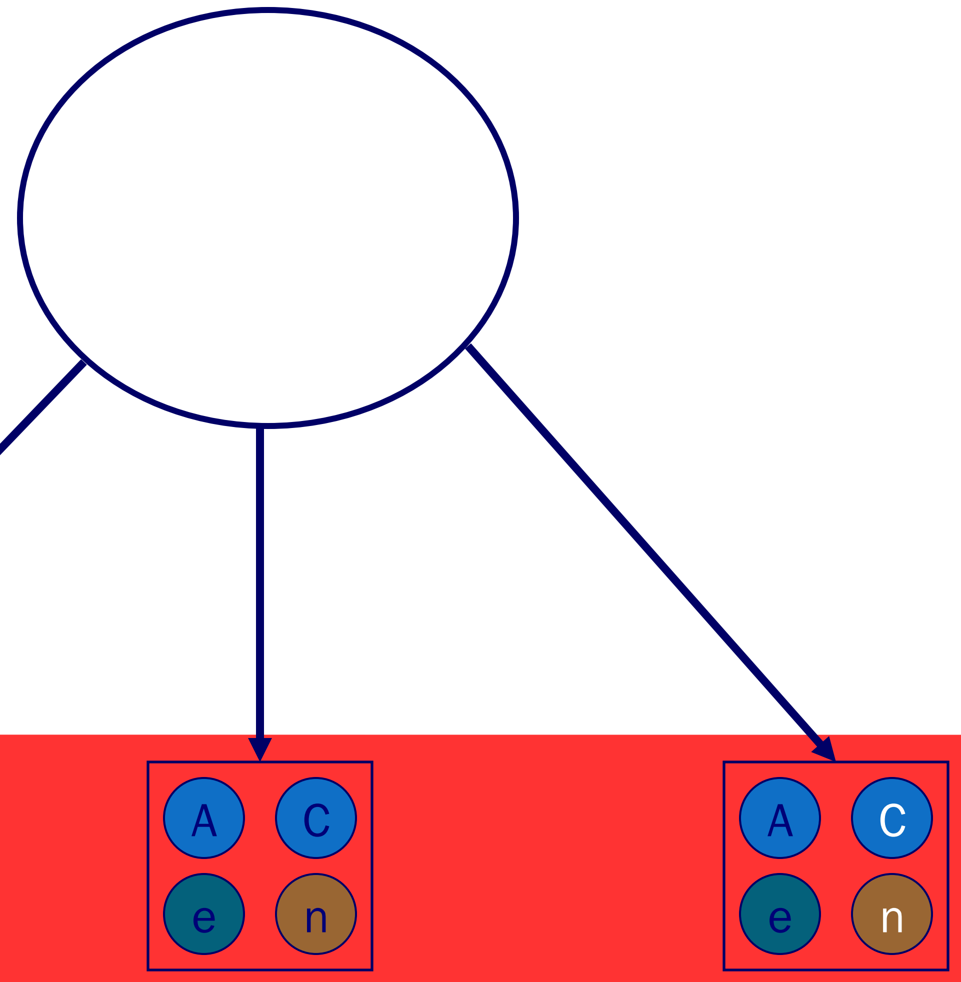


DEFINIENDO UN CONSTRUCTO

Multidimensional

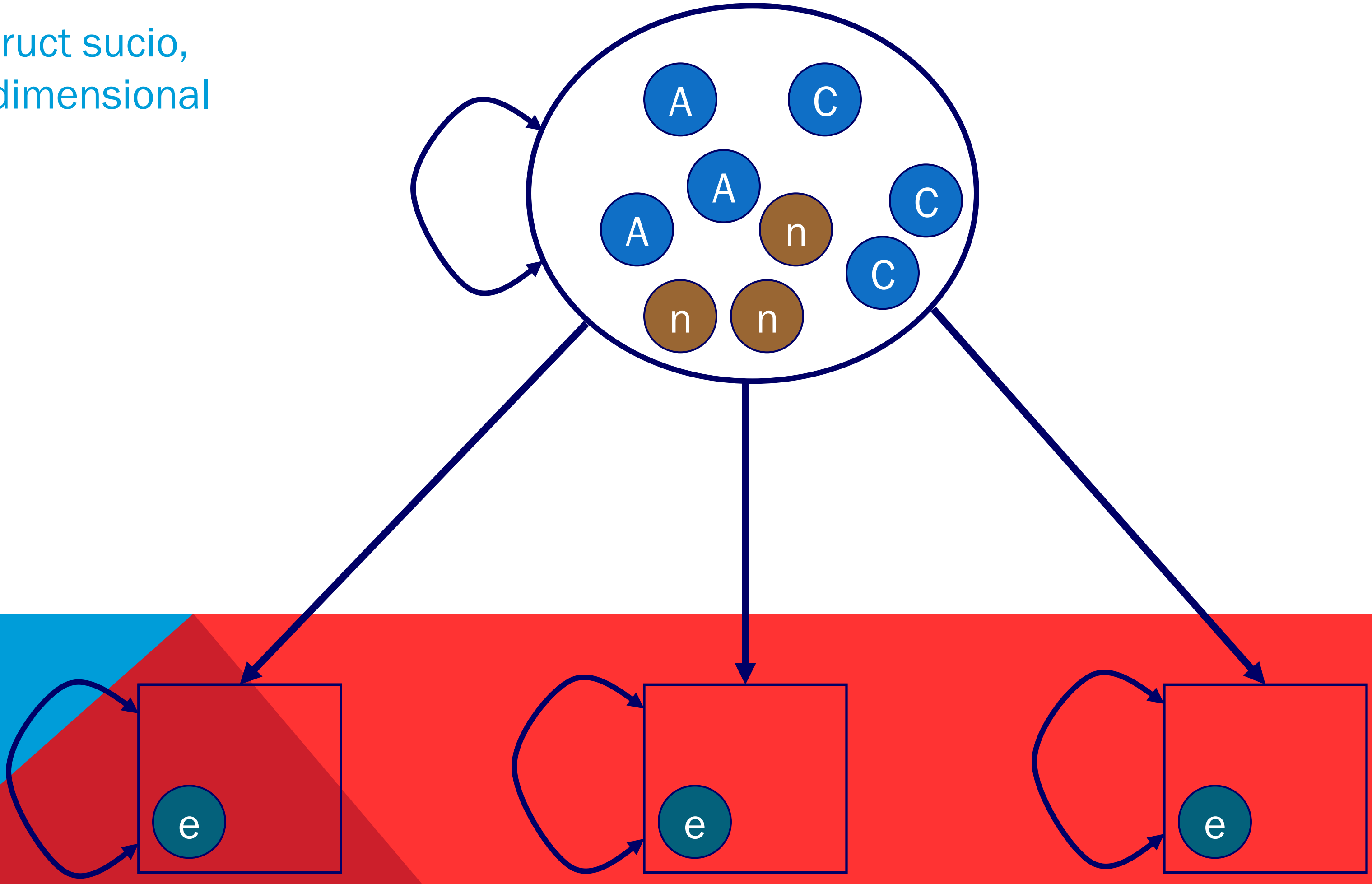


DEFINIENDO UN CONSTRUCTO

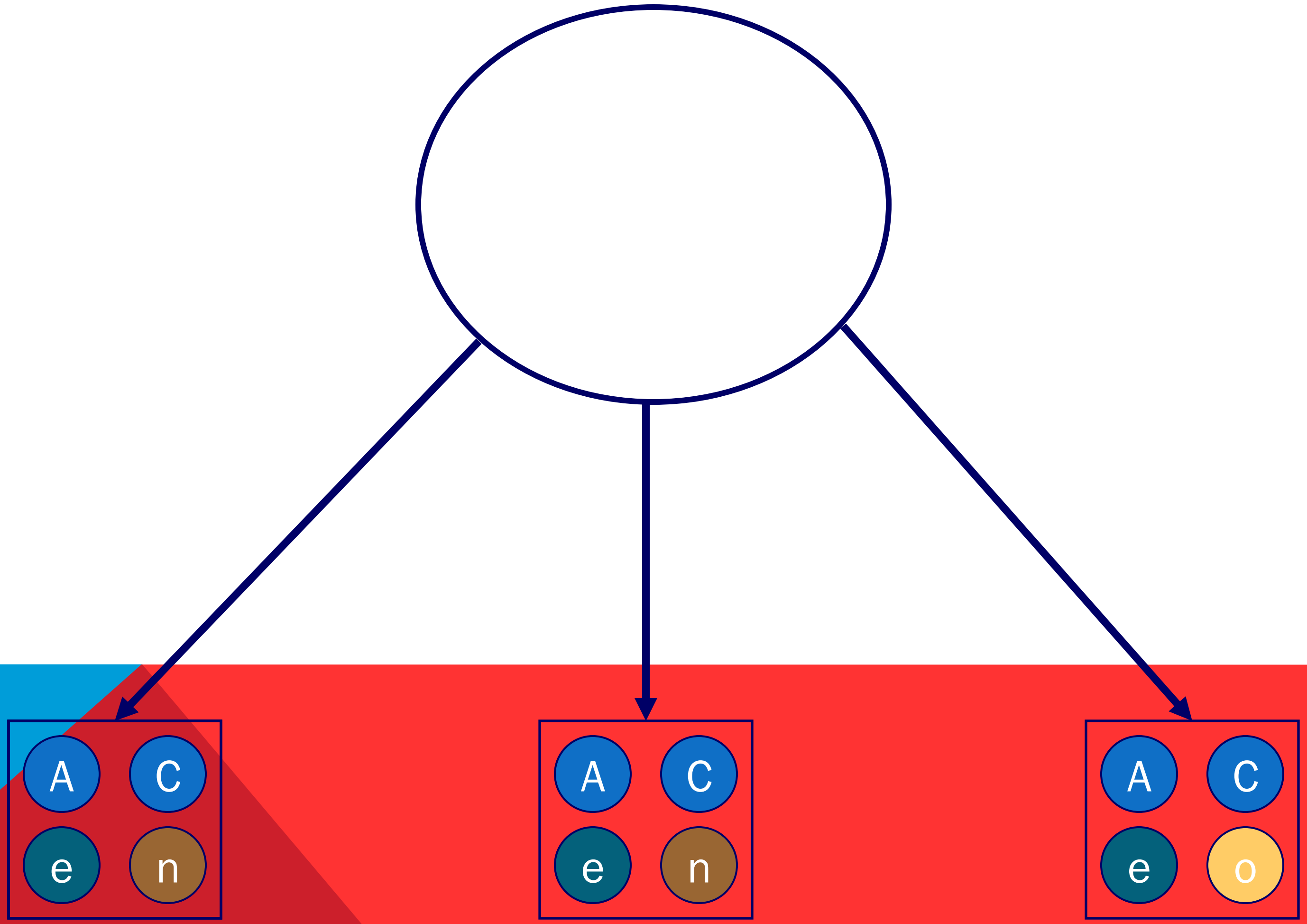


DEFINIENDO UN CONSTRUCTO

Construct sucio,
multidimensional



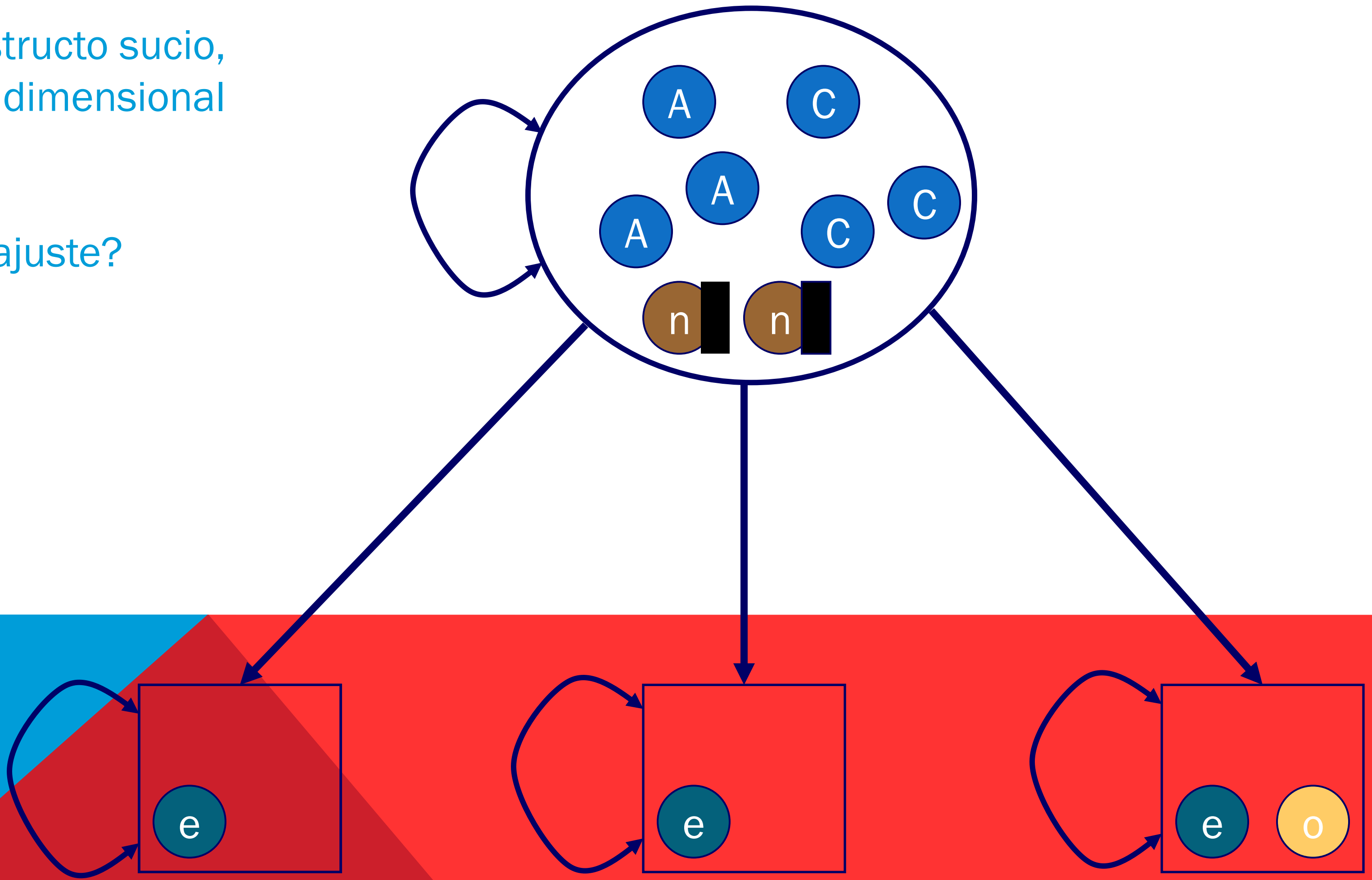
DEFINIENDO UN CONSTRUCTO



DEFINIENDO UN CONSTRUCTO

Constructo sucio, multidimensional

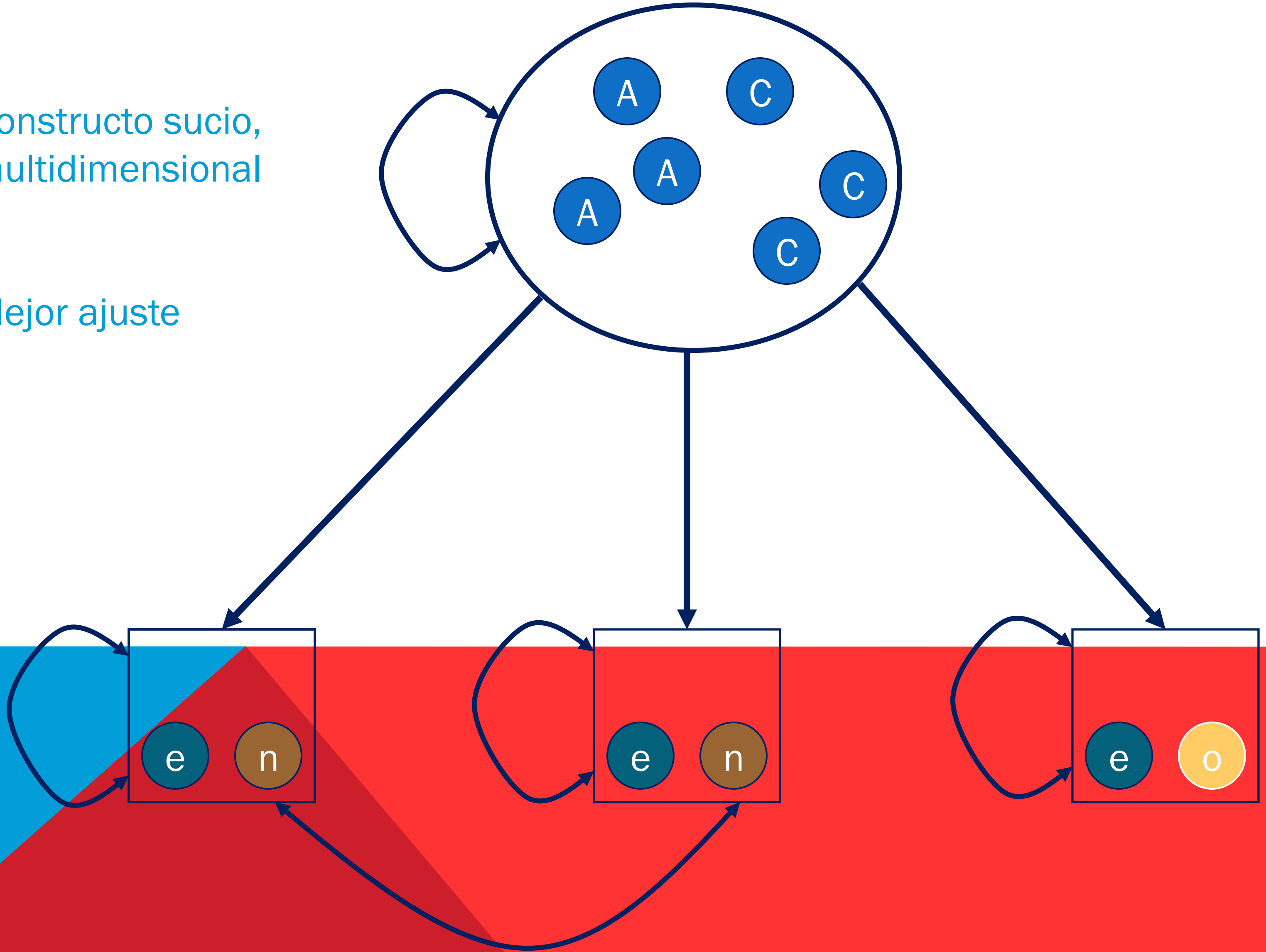
Mal ajuste?



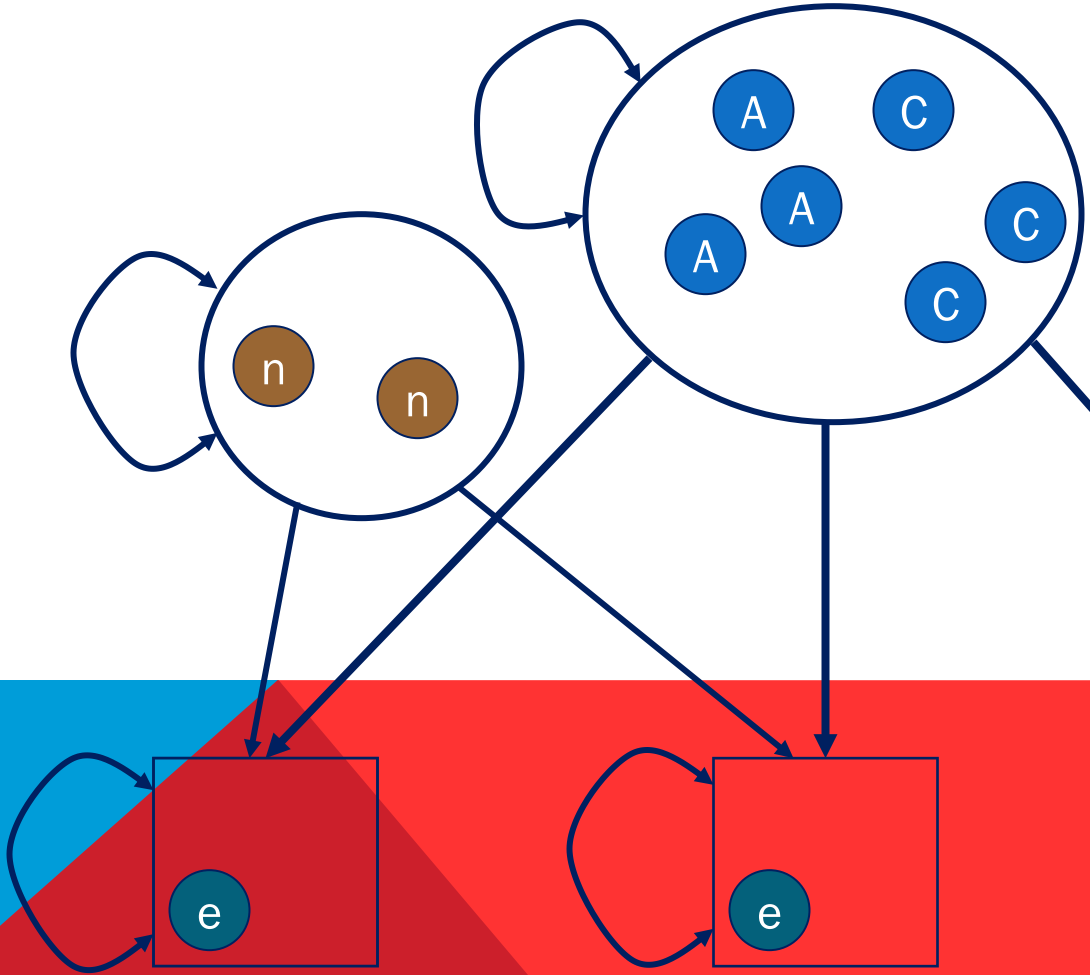
DEFINIENDO UN CONSTRUCTO

Constructo sucio, multidimensional

Mejor ajuste



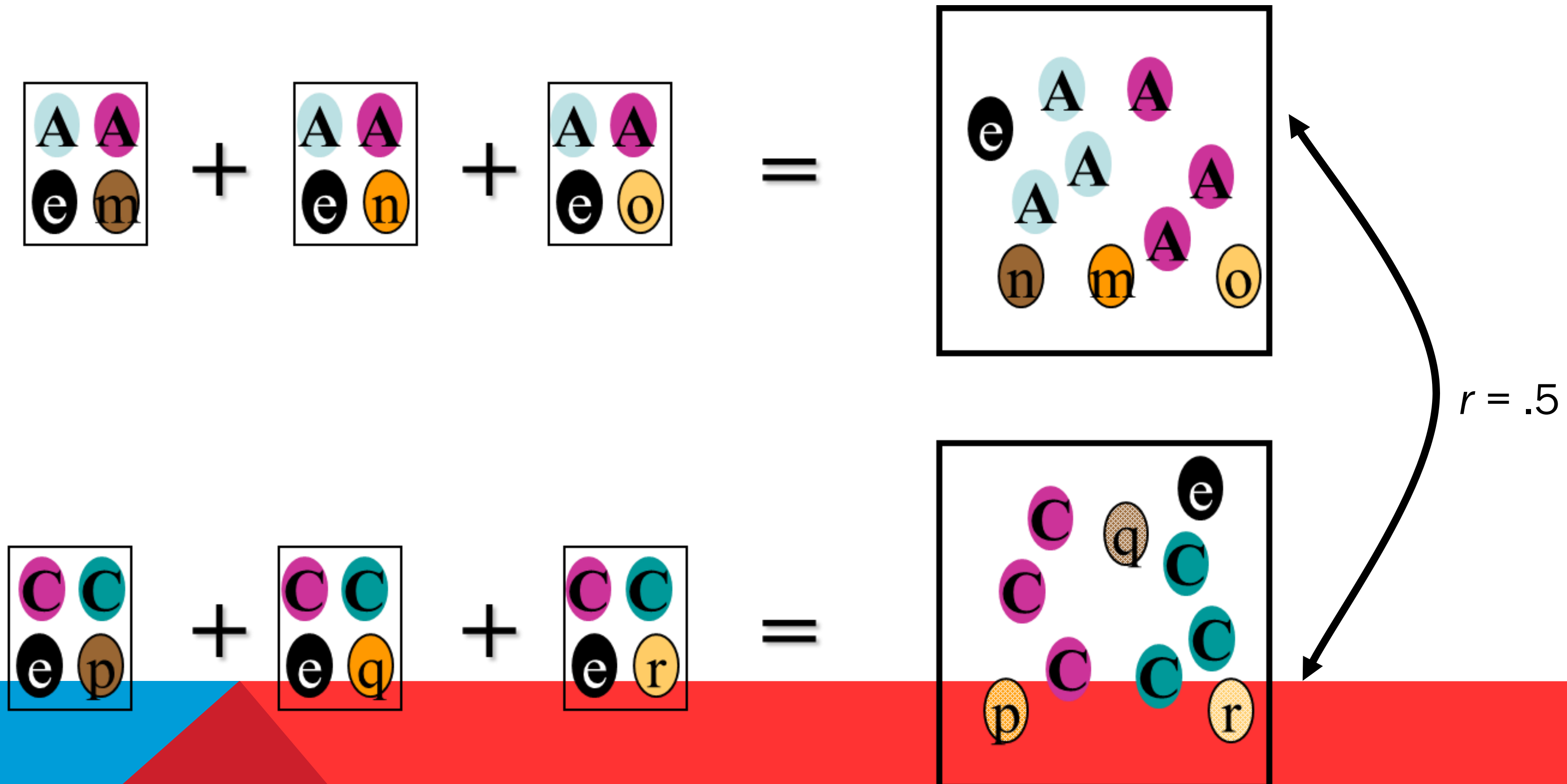
DEFINIENDO UN CONSTRUCTO



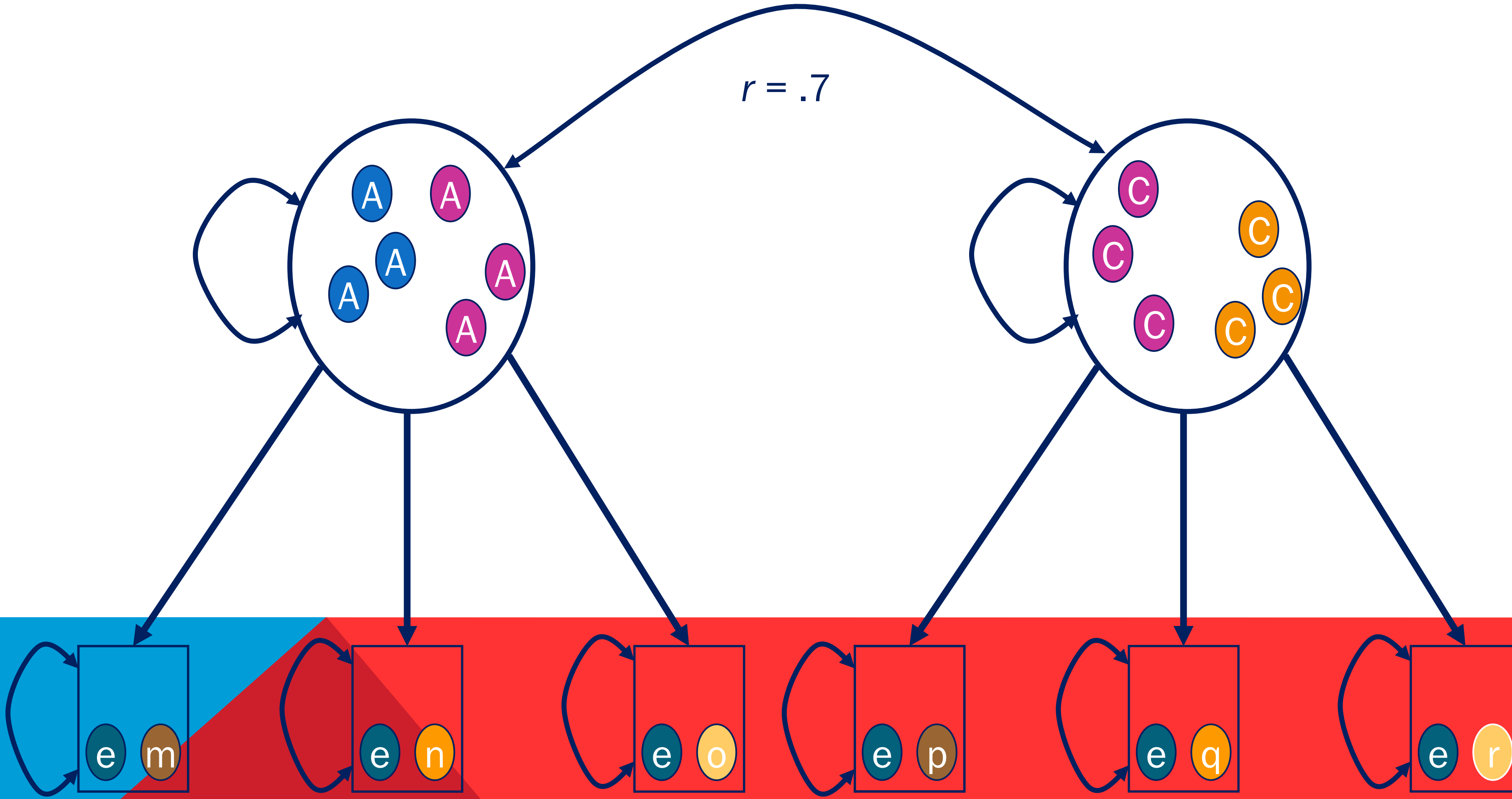
Constructo sucio, multidimensional

Factor separado

RELACIÓN ENTRE VARIABLES MANIFIESTAS



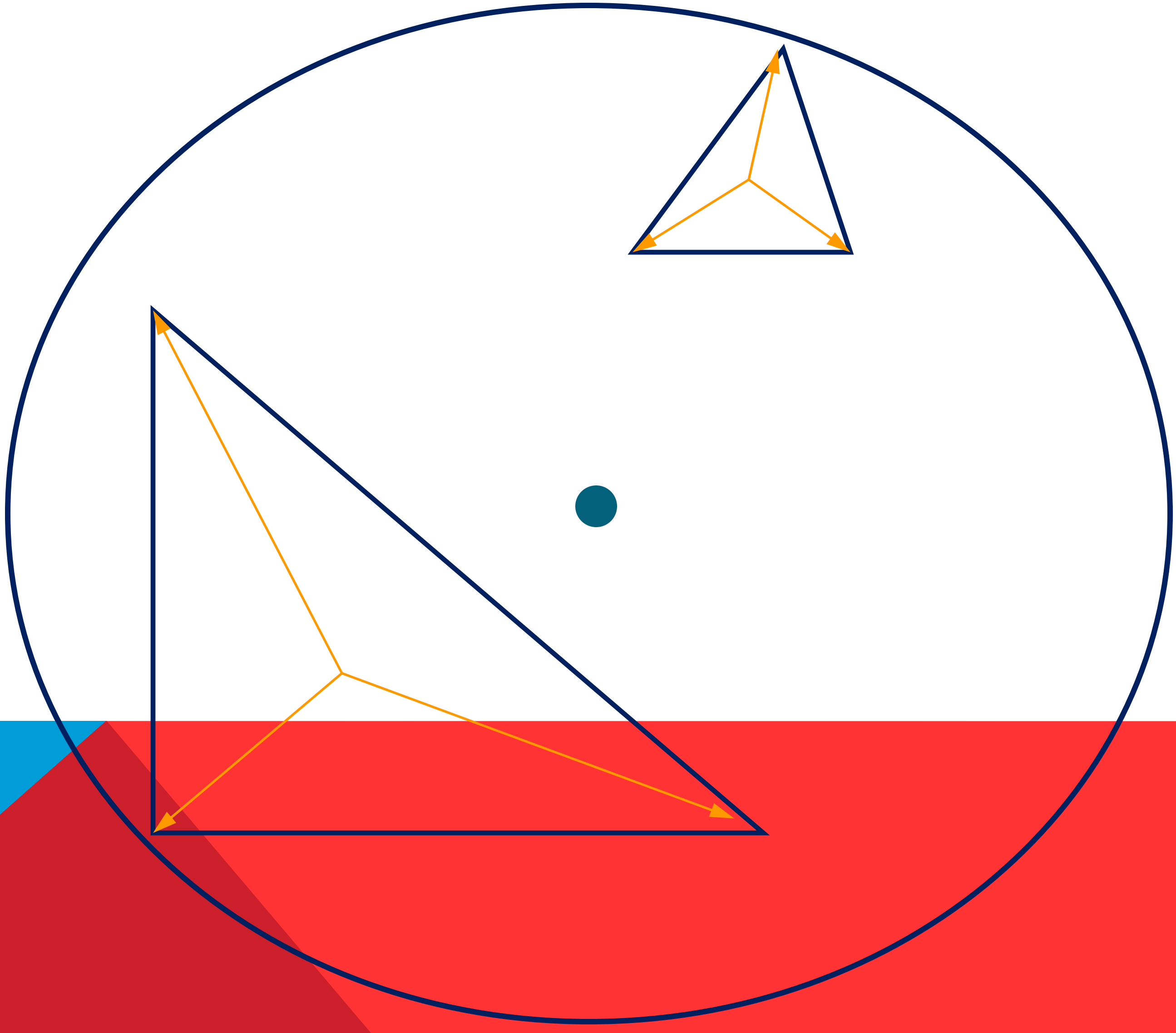
RELACIÓN ENTRE CONSTRUCTOS



SELECCIONAR INDICADORES

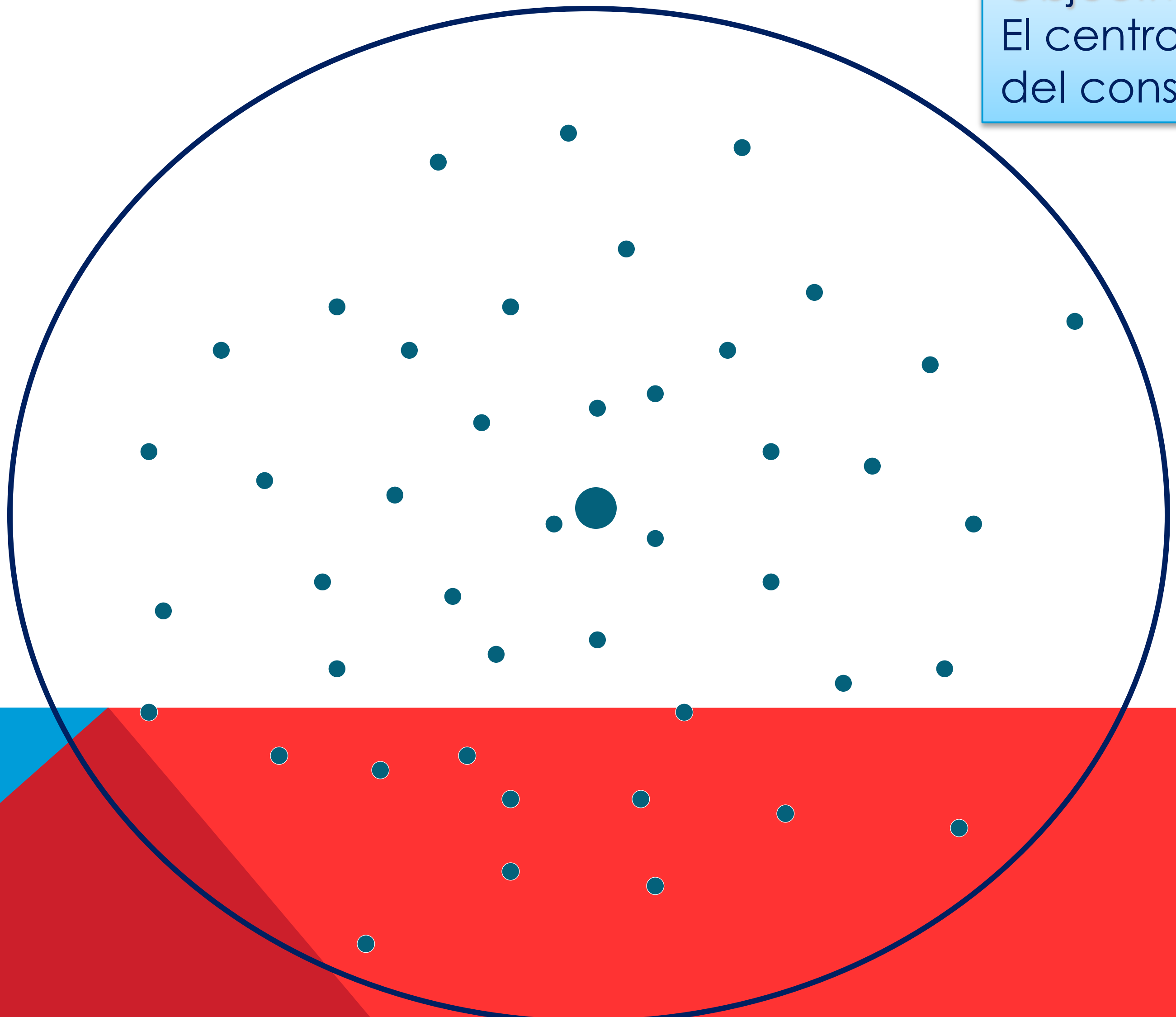
Hay que entender las circunstancias en las que 'buenos' indicadores son malos, y 'malos' indicadores son buenos

CONFIABILIDAD DEL CONSTRUCTO

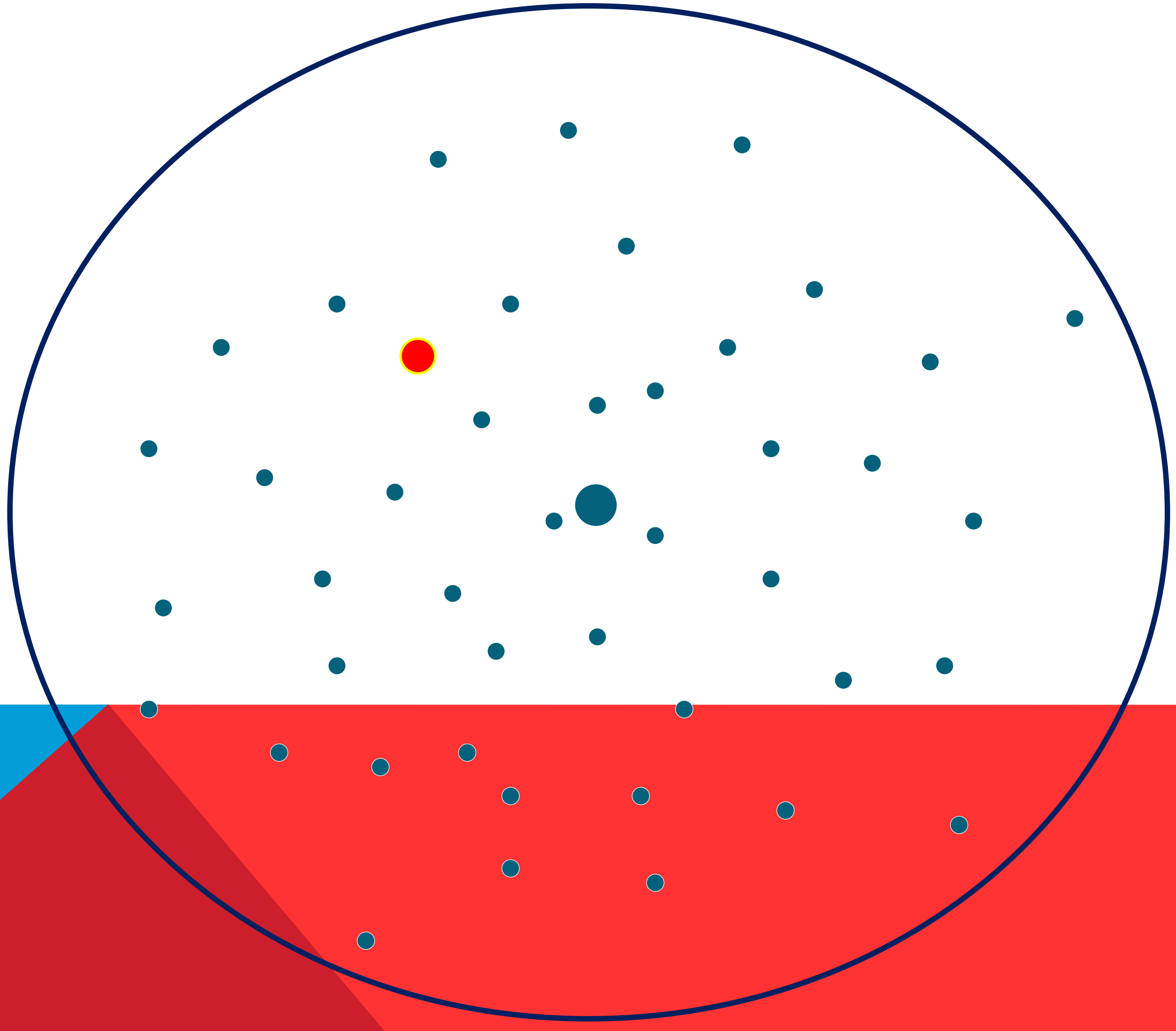


CARACTERÍSTICAS DE INDICADORES

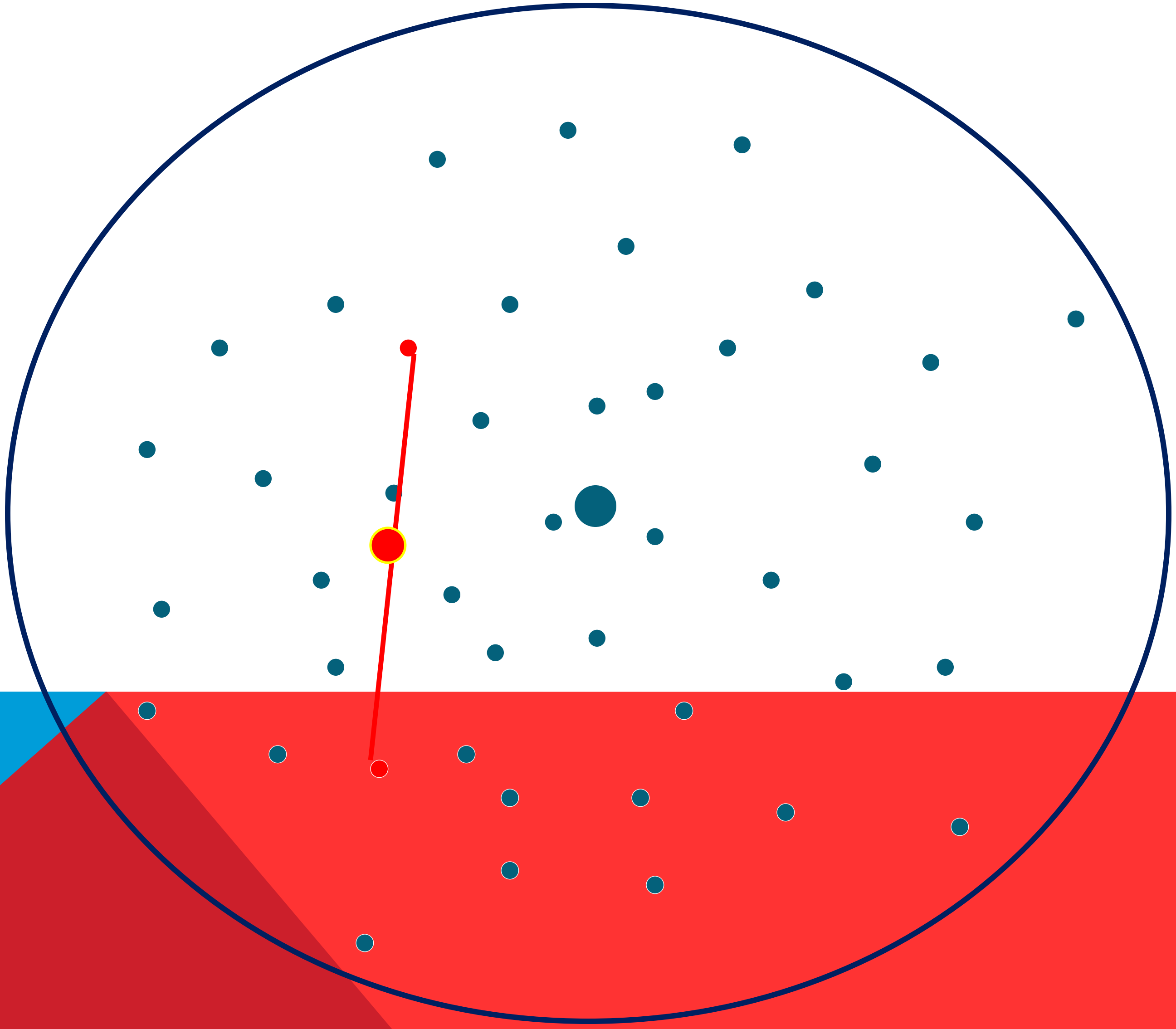
Objectivo:
El centroide
del constructo



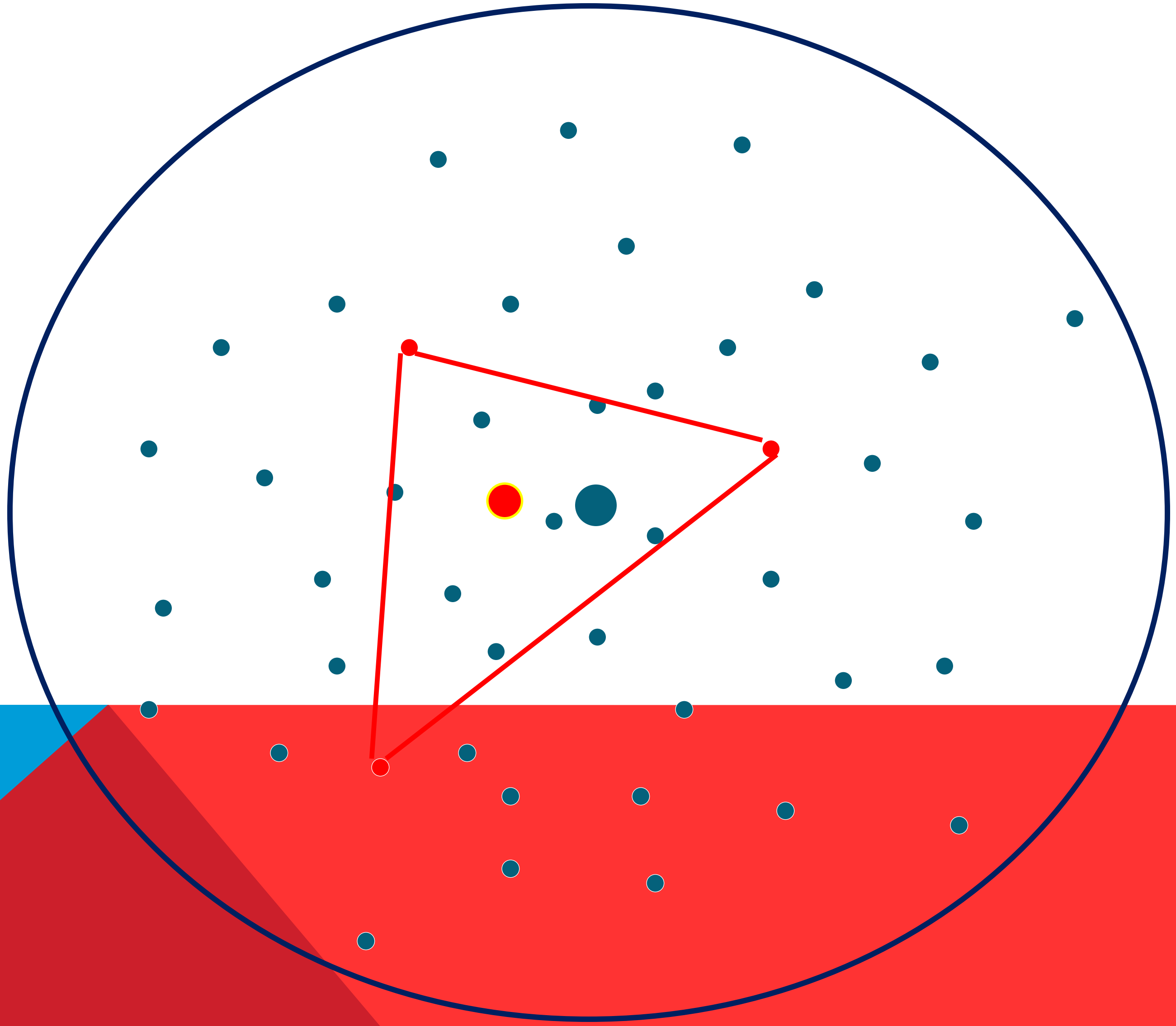
CARACTERÍSTICAS DE INDICADORES



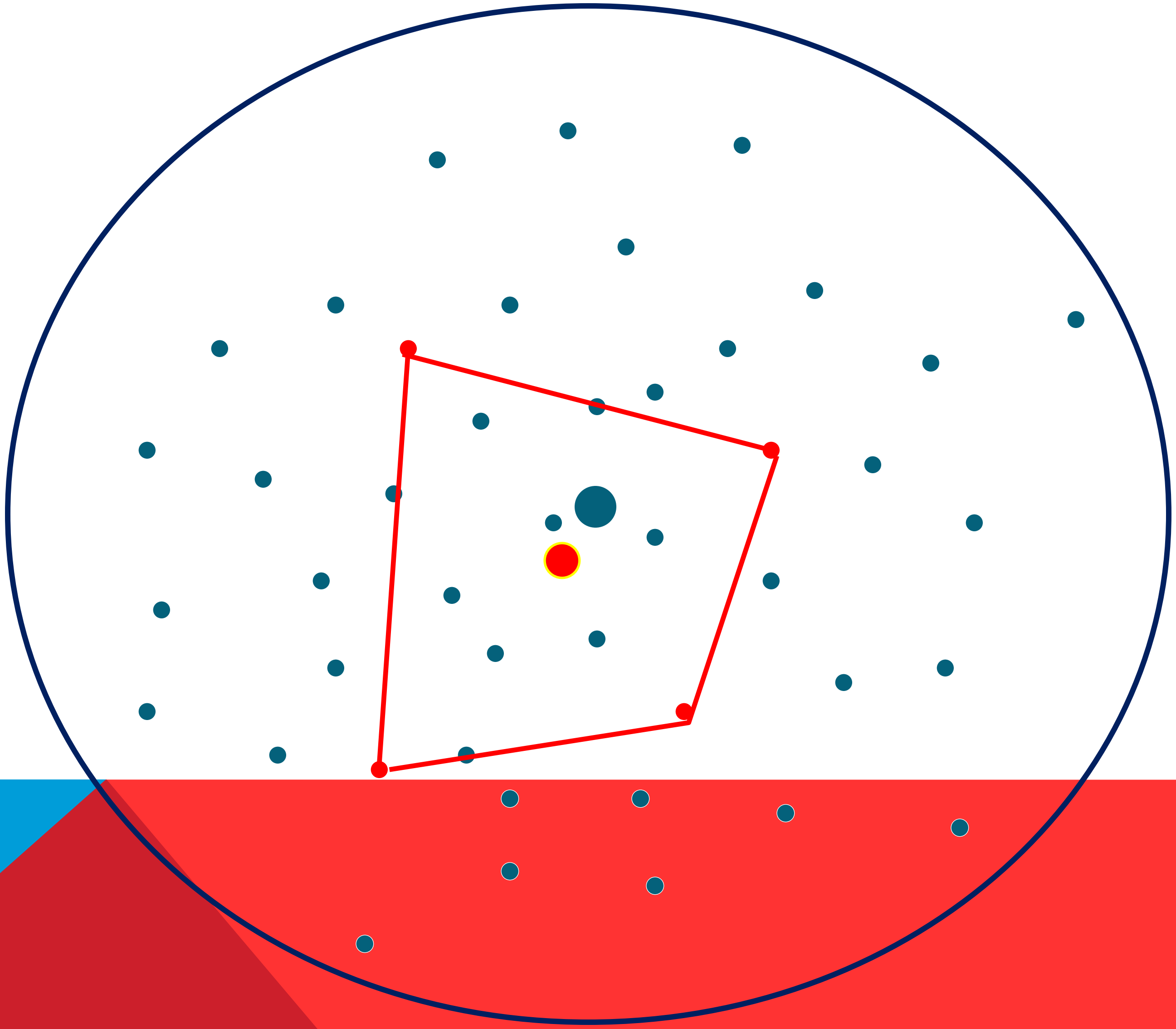
CARACTERÍSTICAS DE INDICADORES



CARACTERÍSTICAS DE INDICADORES

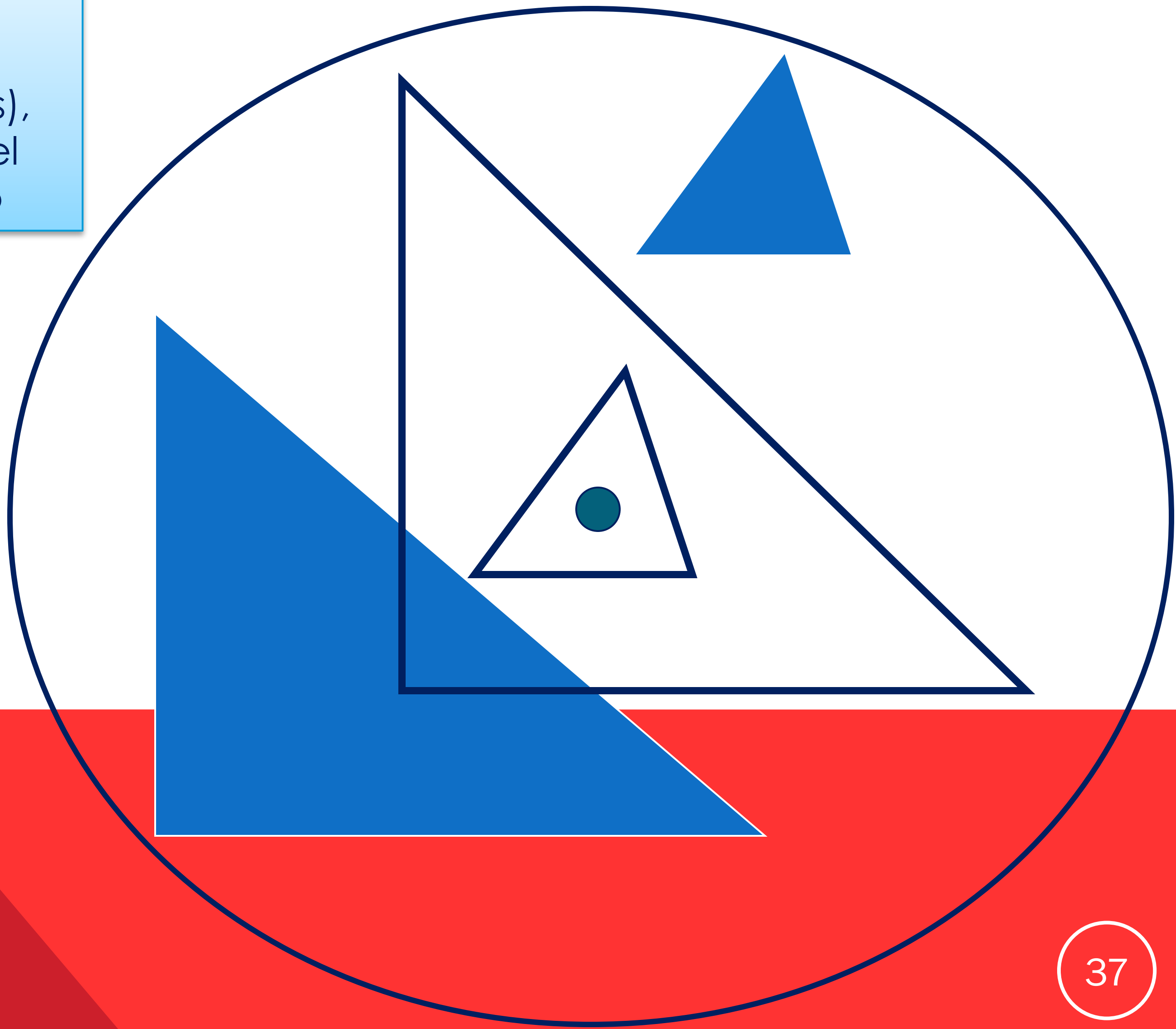


CARACTERÍSTICAS DE INDICADORES



SELECCIONANDO INDICADORES

- Baja consistencia interna (correlaciones débiles), Pero todavía evalúa el constructo verdadero



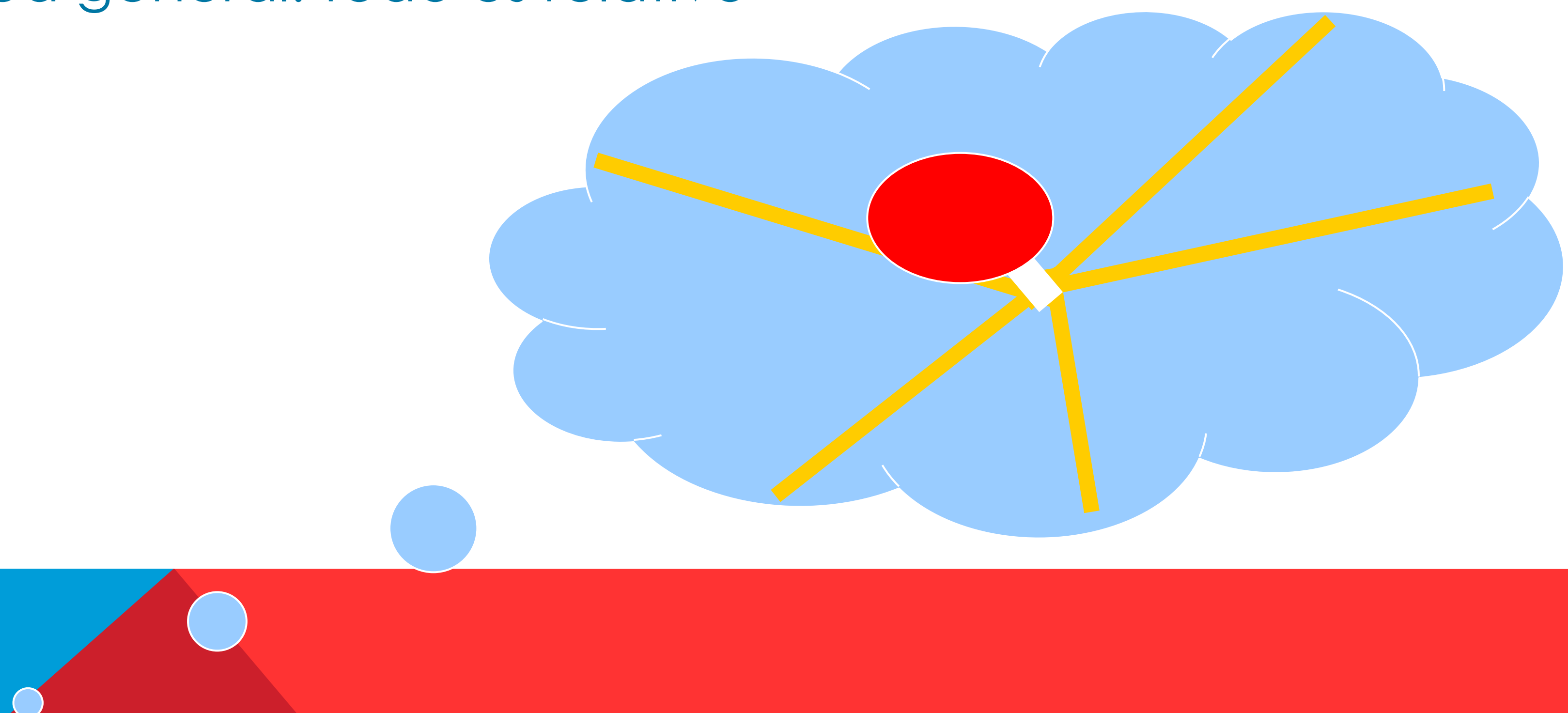
FIJANDO LA ESCALA E IDENTIFICACIÓN

FIJANDO LA ESCALA

(arbitrariamente) Establecer un punto de referencia

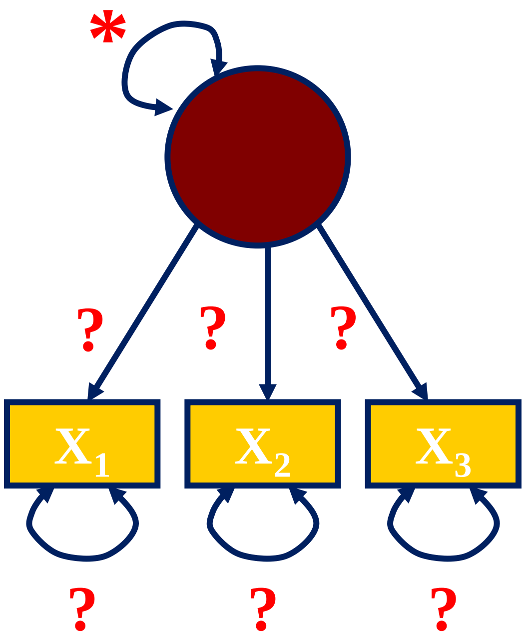
Tome las medidas desde este punto de referencia

Idea general: todo es relativo



IDENTIFICACIÓN

'Apenas' identificado: el numero de varianzas y covarianzas (conocidas) es igual al numero de parámetros que se estiman (desconocidos)



Variance/Covariance Matrix

	X_1	X_2	X_3
X_1	.64		
X_2	.56	.81	
X_3	.48	.53	.77

Number of 'knowns' = 6
= (p * (p+1)) / 2
(where p = # manifest variables

Number of 'unknowns' (λ, θ) = 6

IDENTIFICACIÓN

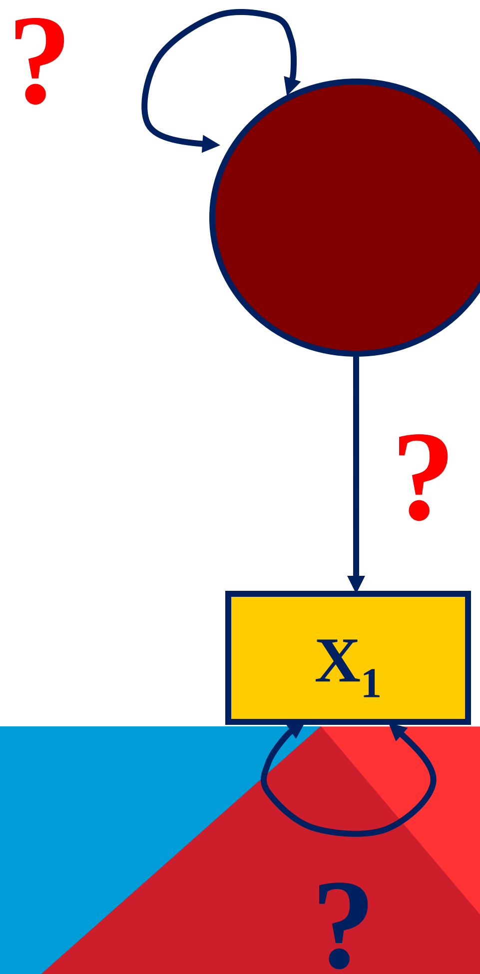
'Sobre' identificados: el numero de varianzas y covarianzas es mayor al numero de parámetros a estimar

- Grados de libertad (df) de sobra

'Sub' identificados: el numero de varianzas y covarianzas es menor que el numero de parámetros a estimar

- Grados de libertad negativos

UN INDICADOR



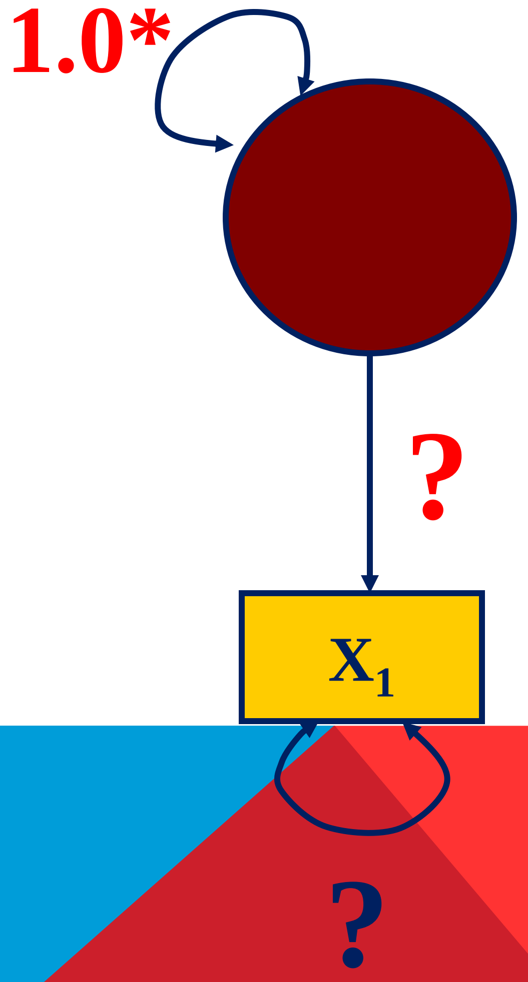
Variance/Covariance Matrix

	X_1
X_1	.64

Parámetros estimados (desconocidos): 3

Información observada (conocida): 1

UN INDICADOR



Variance/Covariance Matrix

	X_1
X_1	.64

Fijar la escala de medida

- Factor ID method:

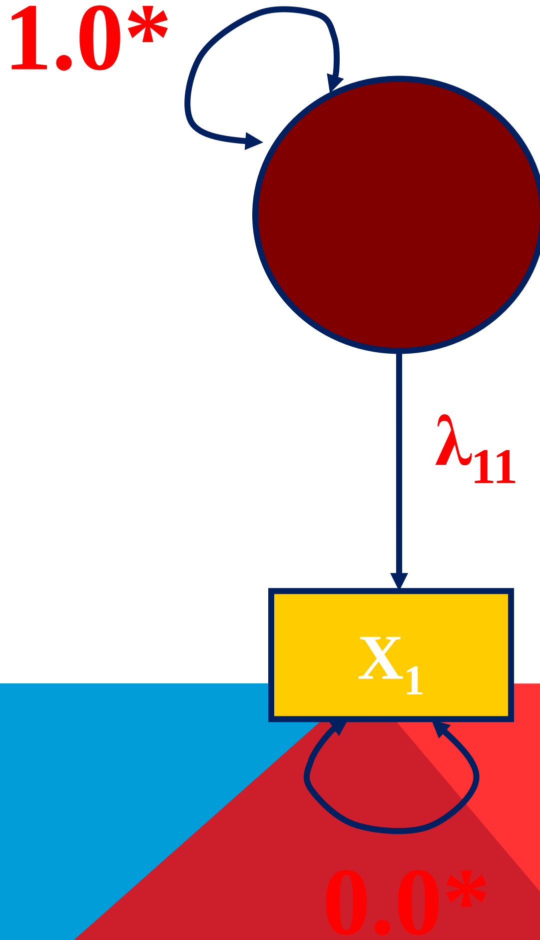
  Fijar la varianza latente (y_{11}) = 1.0

Parámetros estimados (desconocidos): 2

Información observada (conocida): 1

…Problema: 2 unknowns > 1 known

UN INDICADOR



Variance/Covariance Matrix

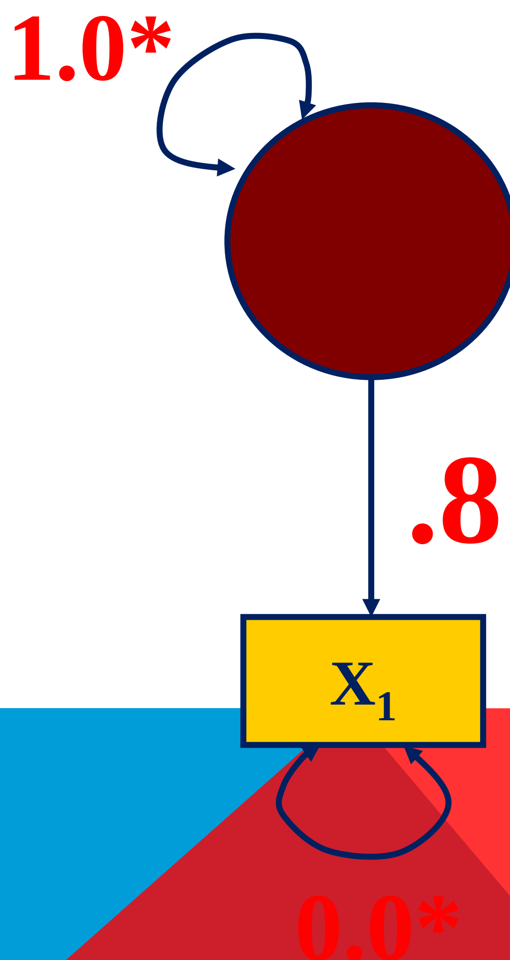
	X_1
X_1	.64

- **Tracing Rules:**

$$\text{Varianza}(X_1) = \lambda_{11}^2 \psi_{11} + \theta_{11}$$

$$= \lambda_{11}^2 (1) + (0)$$

UN INDICADOR



Variance/Covariance Matrix

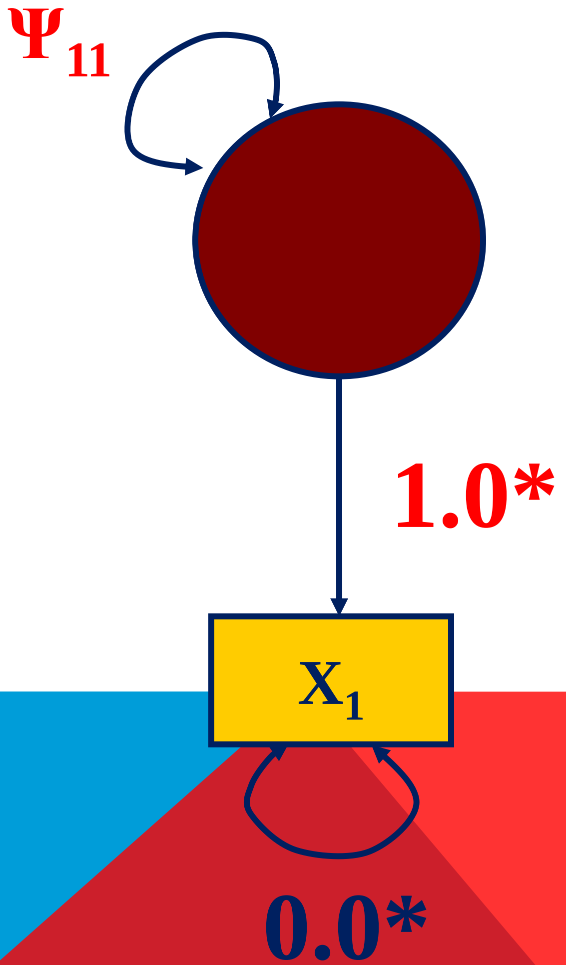
	X_1
X_1	.64

Variance $(X_1) = \lambda_{11}^2$

$.64 = \lambda_{11}^2$

$\lambda_{11} = .8$

UN INDICADOR



Variance/Covariance Matrix

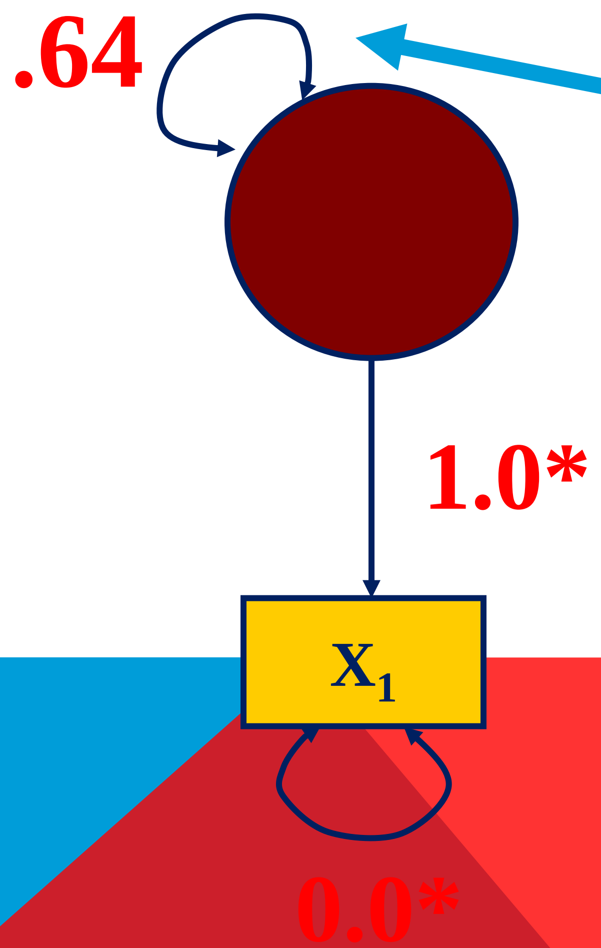
	X_1
X_1	.64

Fijar la escala de medida

- Marker variable method: Fijarlambda (λ) = 1.0

Var(indicador) = Var(constructo)

UN INDICADOR

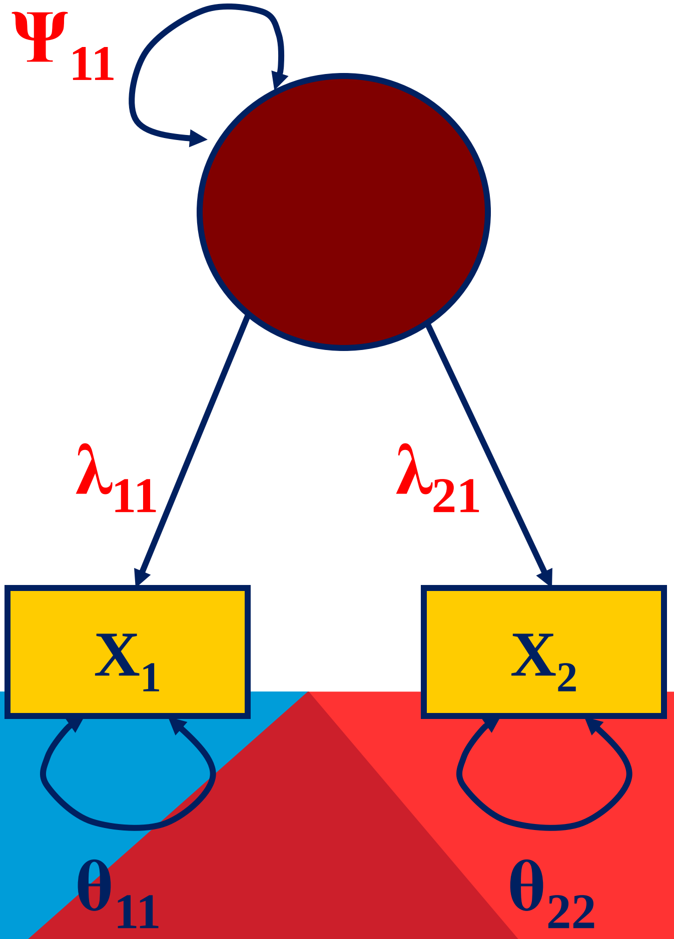


Variance/Covariance Matrix

	X_1
X_1	.64

$$\text{Variancia}(X_1) = \lambda_{11}^2 y_{11} + q_{11}$$
$$= (1)^2 y_{11} + (0)$$
$$= y_{11}$$
$$y_{11} = .64$$

DOS INDICADORES



Variance/Covariance Matrix

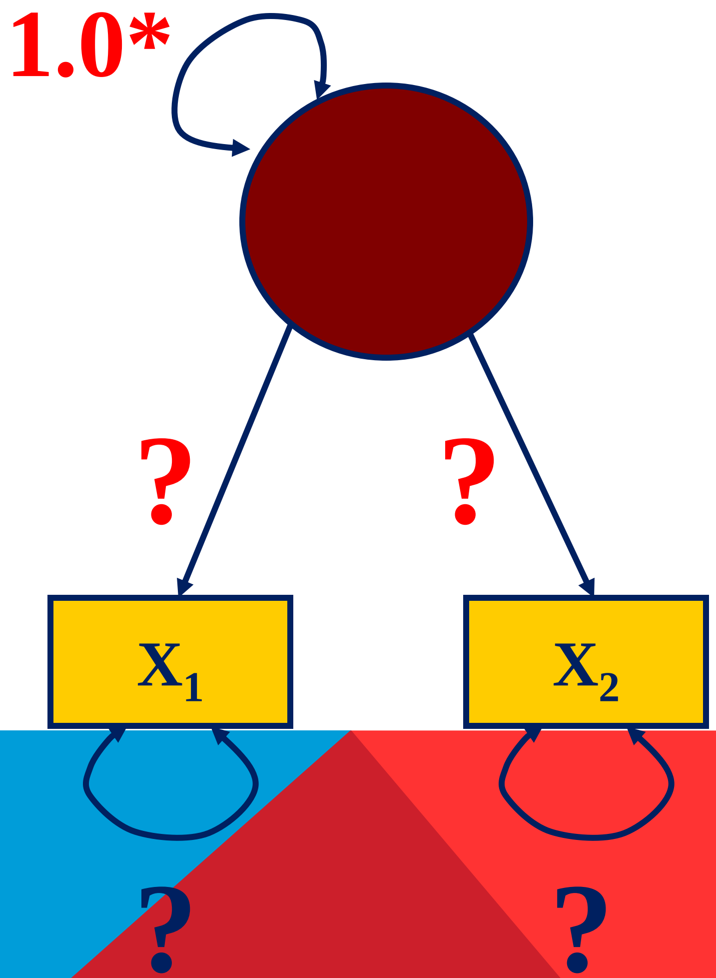
	X_1	X_2
X_1	.64	
X_2	.56	.81

- Diagonal: Variancias
- Off-diagonal: Covariancias

Parámetros estimados (desconocidos): 5

Información observada (conocida): 3

DOS INDICADORES



Variance/Covariance Matrix

	X_1	X_2
X_1	.64	
X_2	.56	.81

Fijar la escala de medida

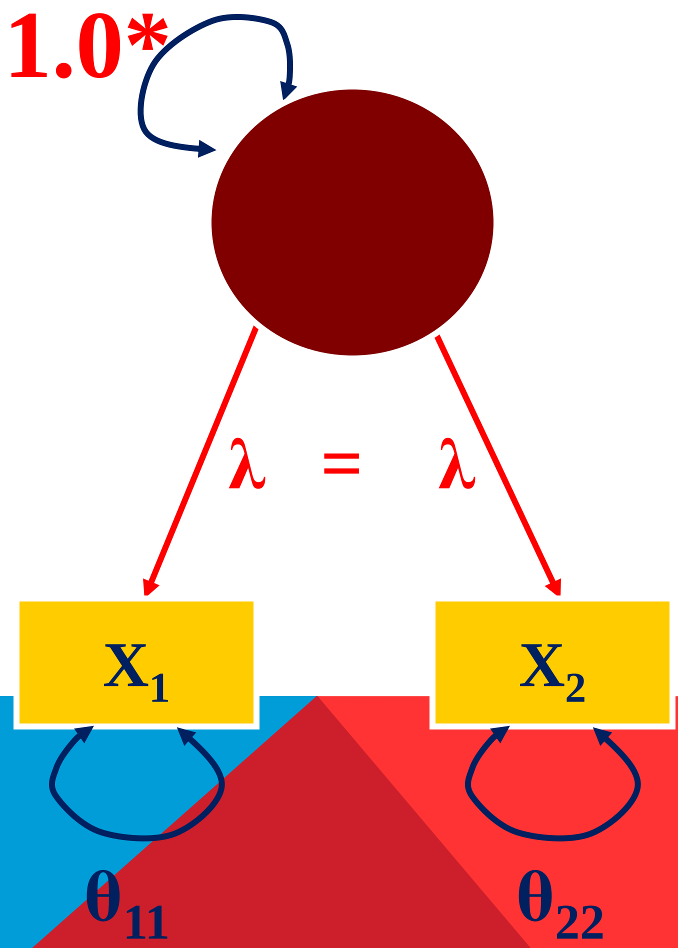
Factor ID method:

Fijar la variancia latente = 1.0

Parámetros estimados (desconocidos): 4

Información observada (conocida): 3

DOS INDICADORES



Variance/Covariance Matrix

	X_1	X_2
X_1	.64	
X_2	.56	.81

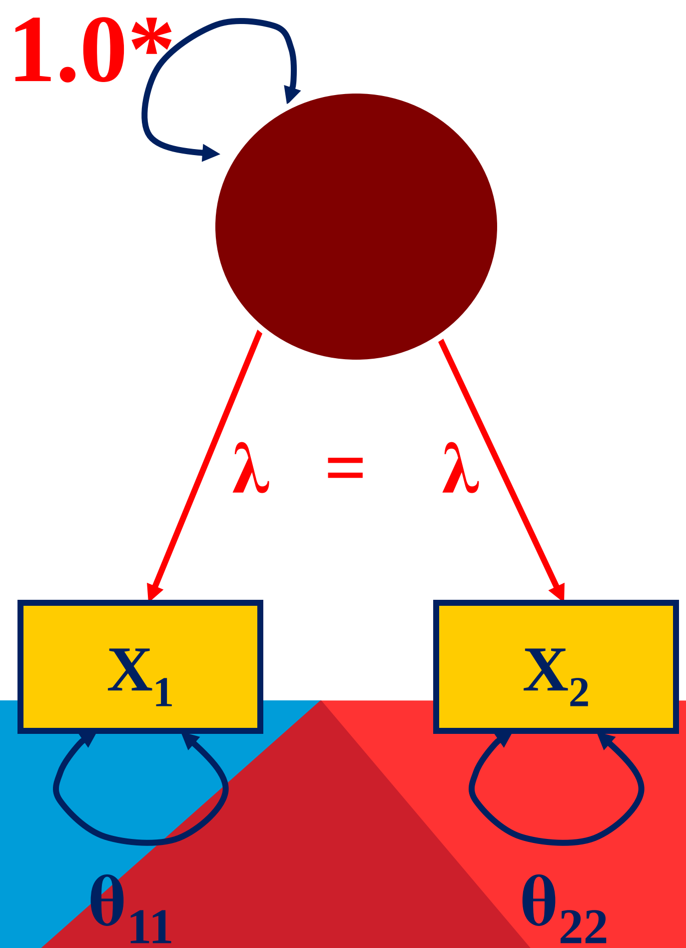
Presupuesto: Tau equivalence

- Igualar los lambda (λ11= λ21)
- Asumir que X1 y X2 contribuyen de manera igual al constructo

Parámetros estimados (desconocidos): 3

Información observada (conocida): 3

DOS INDICADORES



Variance/Covariance Matrix

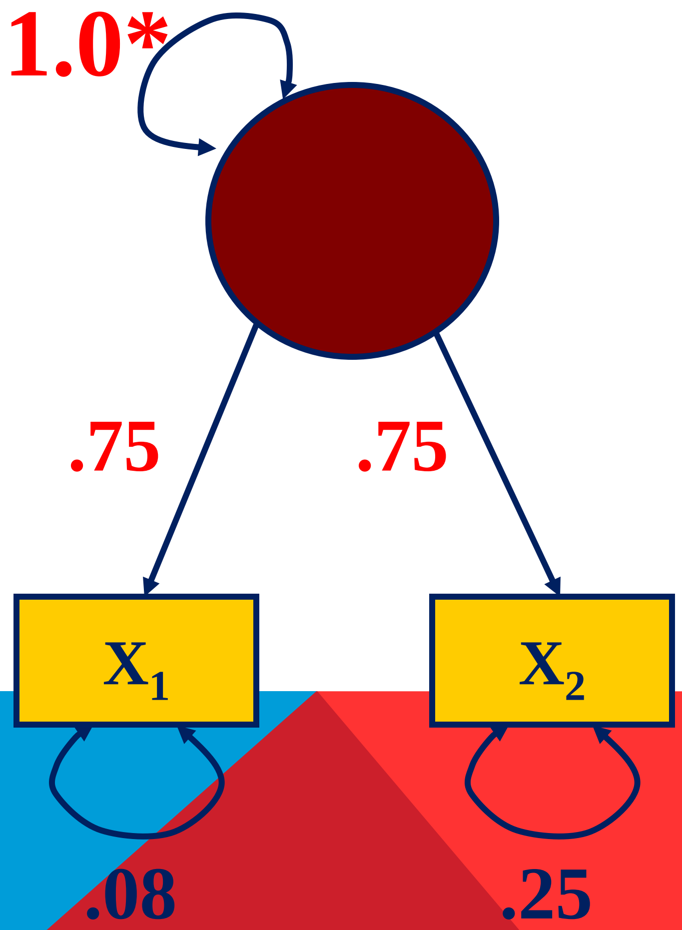
	X_1	X_2
X_1	.64	
X_2	.56	.81

Variancia$(X_1) = \lambda_{11} y_{11} \lambda_{11} + q_{11}$

Variancia$(X_2) = \lambda_{21} y_{11} \lambda_{21} + q_{22}$

Covariancia$(X_1 X_2) = \lambda_{11} \Psi_{11} \lambda_{21}$

DOS INDICADORES



Variance/Covariance Matrix

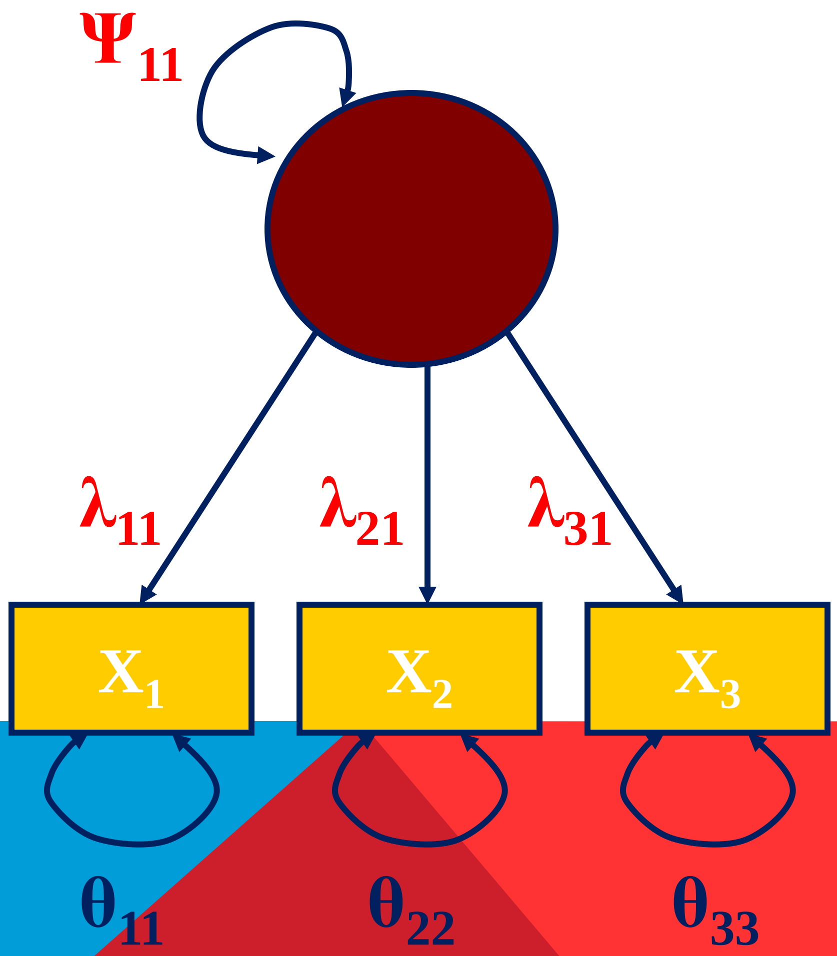
	X_1	X_2
X_1	.64	
X_2	.56	.81

$\lambda_{11} = \lambda_{21} = .75$ (i.e., $\sqrt{.56}$)

$\theta_{11} = .08 = .64 - .75^2$

$\theta_{22} = .25 = .81 - .75^2$

TRES INDICADORES



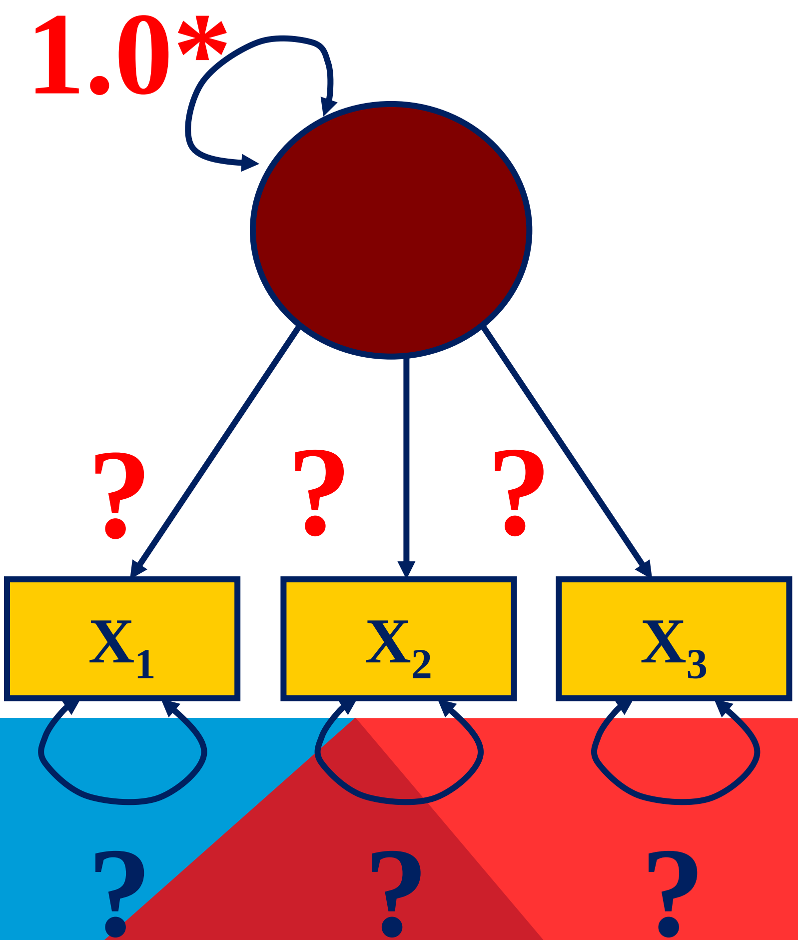
Variance/Covariance Matrix

	X_1	X_2	X_3
X_1	.64		
X_2	.56	.81	
X_3	.48	.53	.77

Parámetros estimados: 7

Información observada: 6

TRES INDICADORES



Variance/Covariance Matrix

	X_1	X_2	X_3
X_1	.64		
X_2	.56	.81	
X_3	.48	.53	.77

Fijar la escala de medida

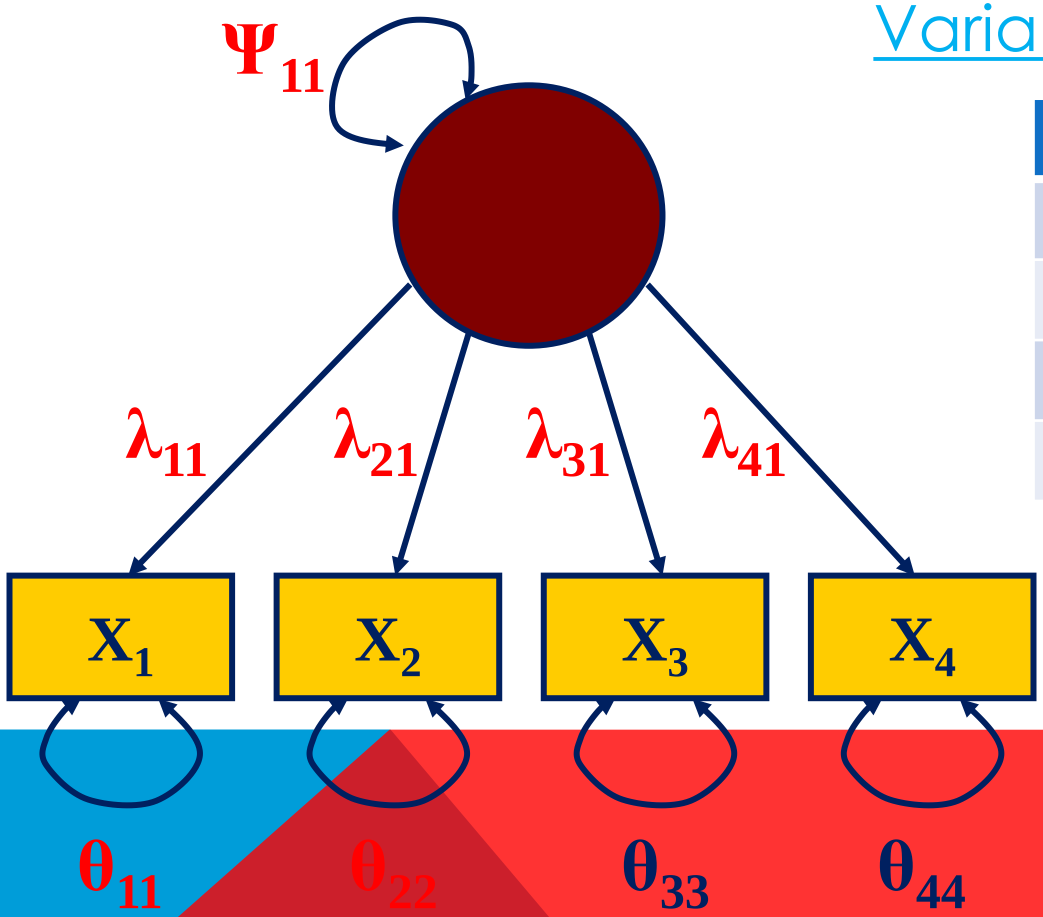
- Factor ID method:

  Fijar la varianza latente= 1.0

Parámetros estimados: 6

Información observada: 6

CUATRO INDICADORES



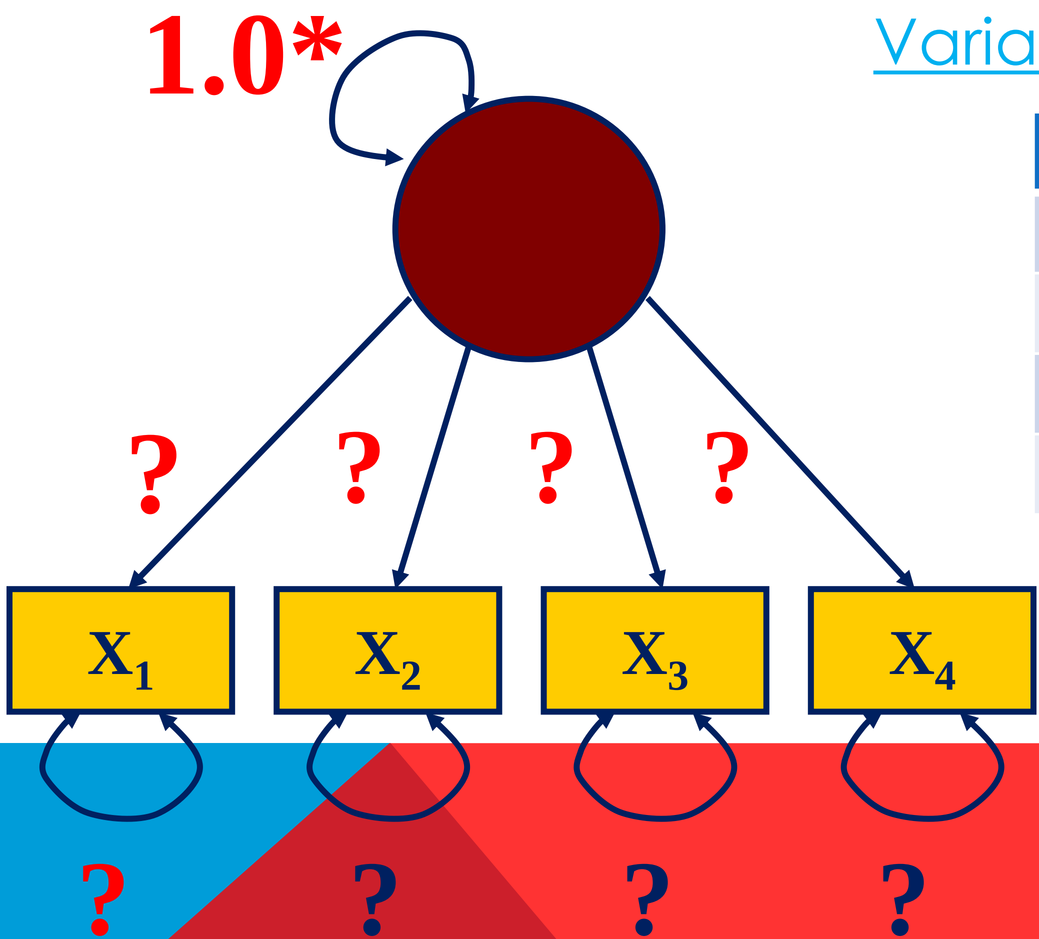
Variance/Covariance Matrix

	X_1	X_2	X_3	X_4
X_1	.64			
X_2	.56	.81		
X_3	.48	.53	.77	
X_4	.41	.49	.50	.70

Parámetros estimados: 9

Información observada: 10

CUATRO INDICADORES



Variance/Covariance Matrix

	X_1	X_2	X_3	X_4
X_1	.64			
X_2	.56	.81		
X_3	.48	.53	.77	
X_4	.41	.49	.50	.70

Fijar la escala:

- Var latente = 1.0

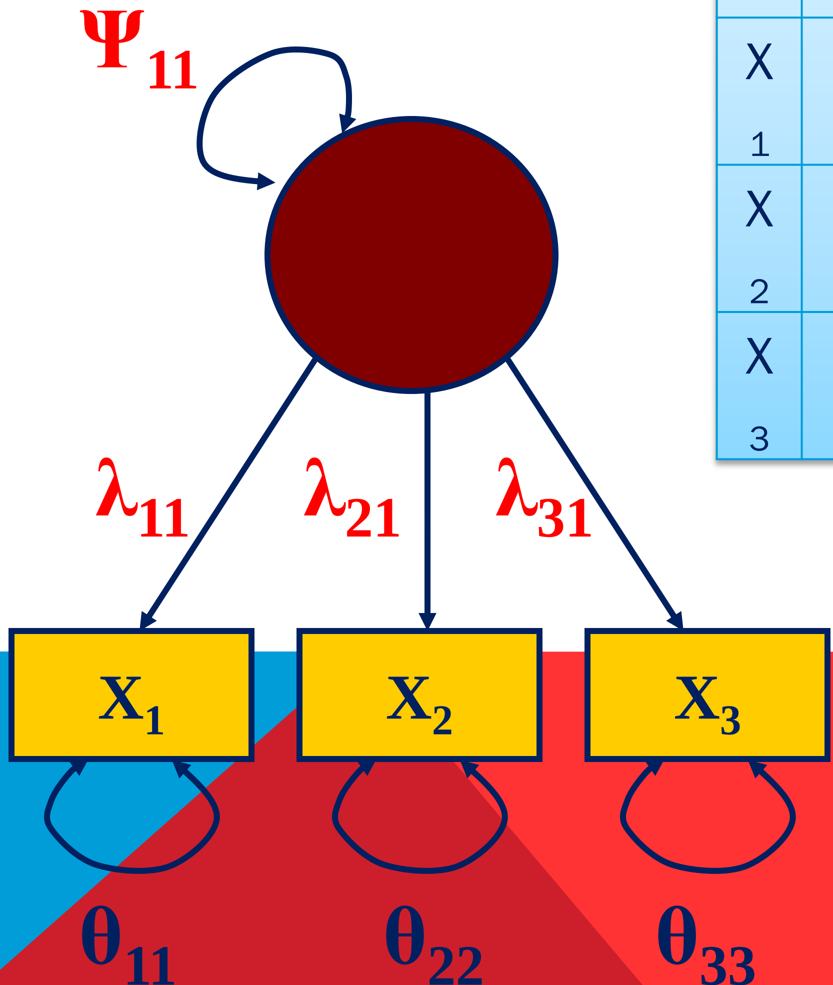
Sobre identificada (no necesariamente bueno) ← Parámetros estimados: 8

Información observada: 10

MÉTODOS DE FIJAR LA ESCALA

1. Fixed factor: fijar la varianza latente (ψ_{11}= 1.0)
2. Marker variable: fijar una carga factorial (λ_{11}= 1.0)
3. Effects coding: limitar a las cargas factoriales que sean en promedio 1.0

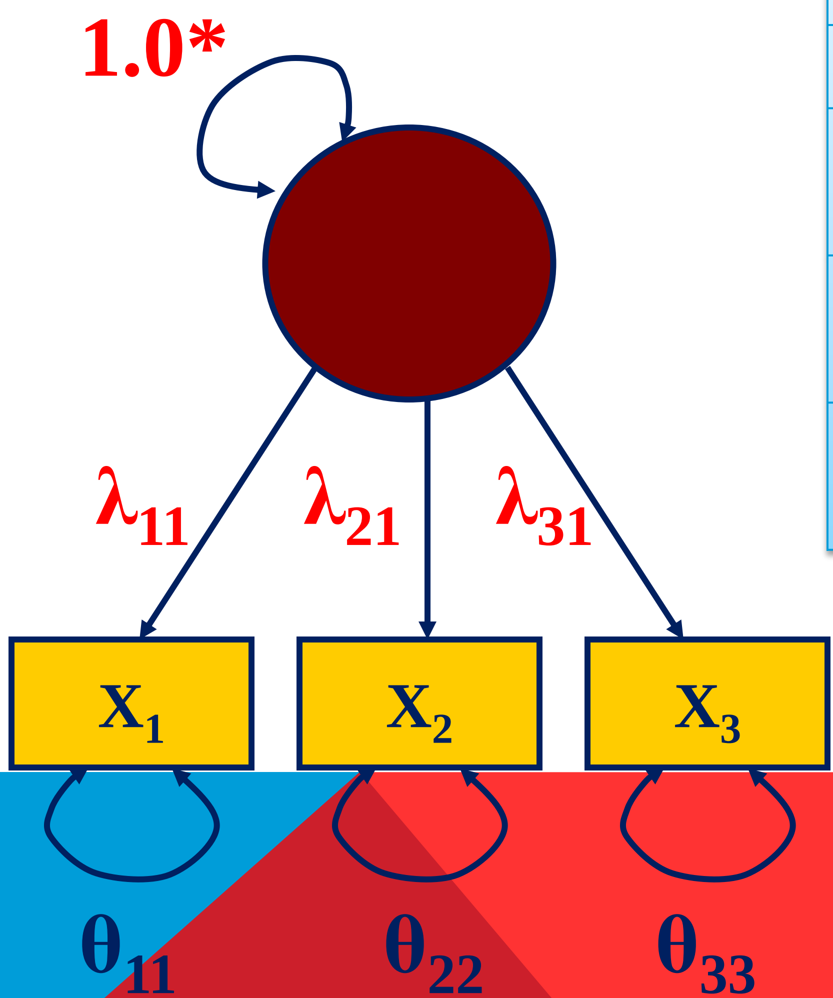
PARA TRES INDICADORES



<u>Implied variance/covariance matrix</u>

	X_1	X_2	X_3
X_1	$\lambda_{11}y_{11}\lambda_{11} + q_{11}$		
X_2	$\lambda_{11}y_{11}\lambda_{21}$	$\lambda_{21}y_{11}\lambda_{21} + q_{22}$	
X_3	$\lambda_{11}y_{11}\lambda_{31}$	$\lambda_{21}y_{11}\lambda_{31}$	$\lambda_{31}y_{11}\lambda_{31} + q_{33}$

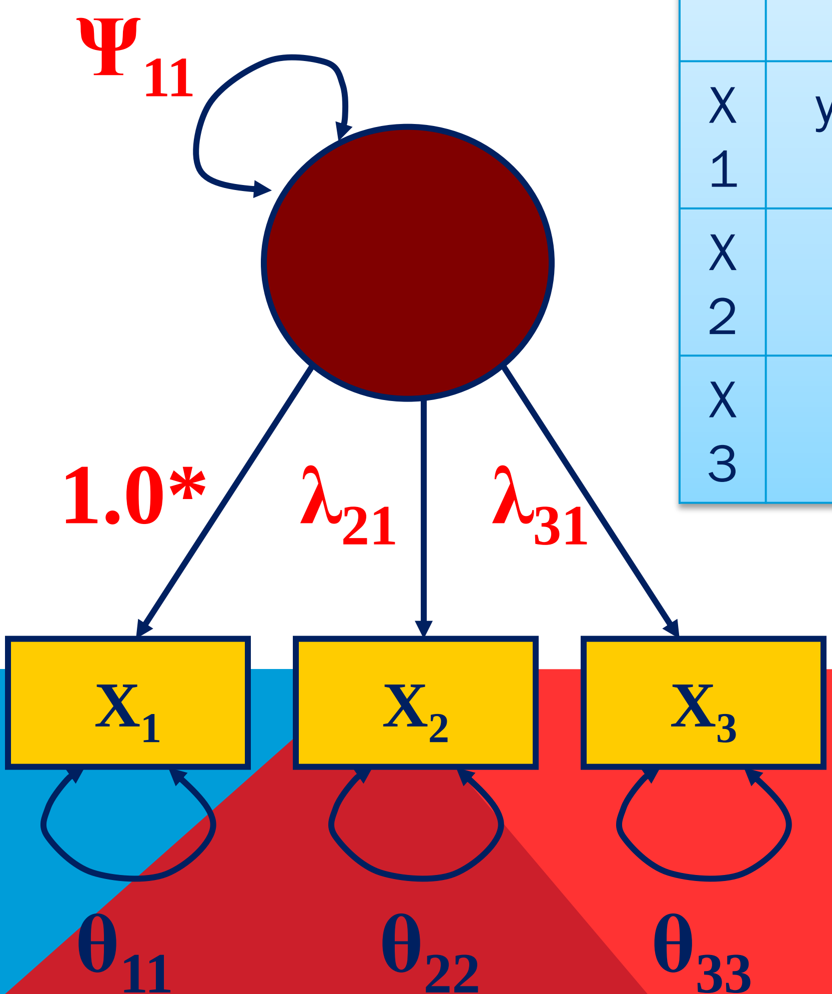
FIJAR LA ESCALA: FIX FACTOR



Implied variance/covariance matrix			
	X1	X2	X3
X1	λ112 + q11		
X2	λ11λ21	λ212 + q22	
X3	λ11λ31	λ21λ31	λ312 + q33

- Fijar la varianza latente: ψ_{11}= 1.0

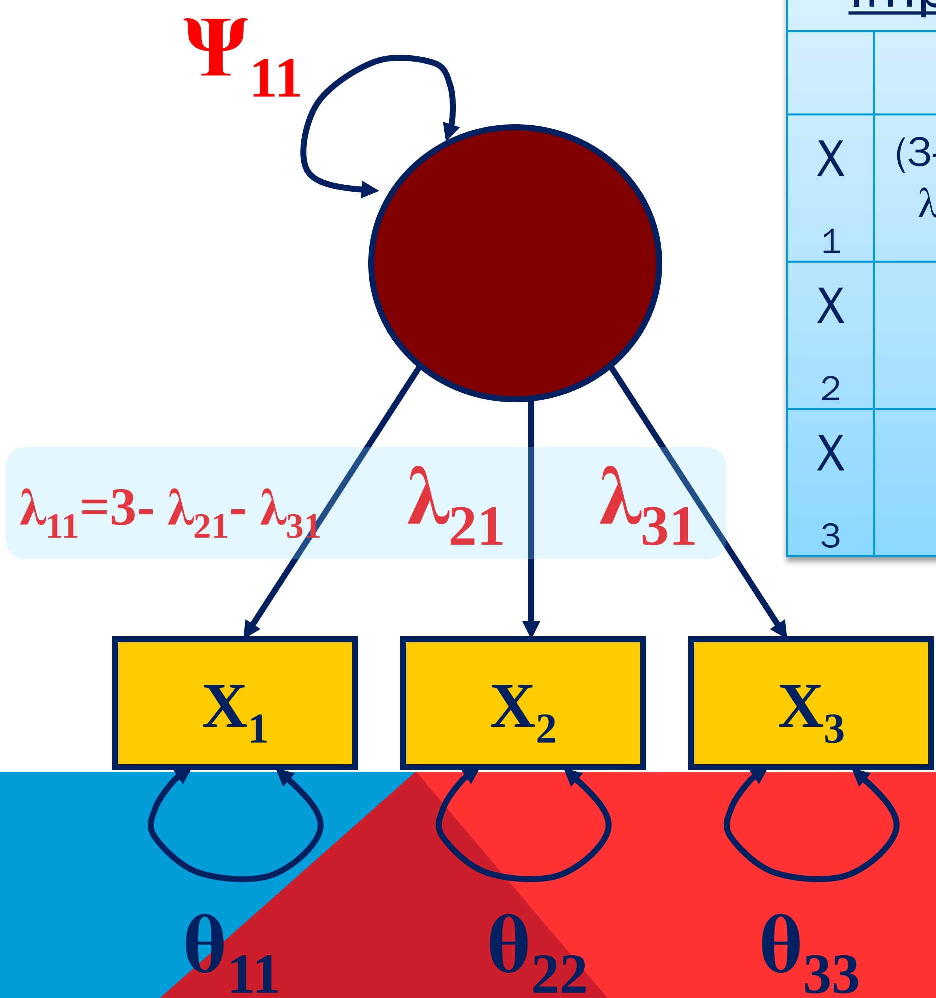
FIJAR LA ESCALA: MARKER VARIABLE



Implied variance/covariance matrix			
	X1	X2	X3
X1	y11 + q11		
X2	y11λ21	λ21y11λ21 + q22	
X3	y11λ31	λ21y11λ31	λ31y11λ31 + q33

- Fijar una carga factorial: $\lambda_{11} = 1.0$
- La seleccion de cual carga es arbitraria

FIJAR LA ESCALA: EFFECTS CODING

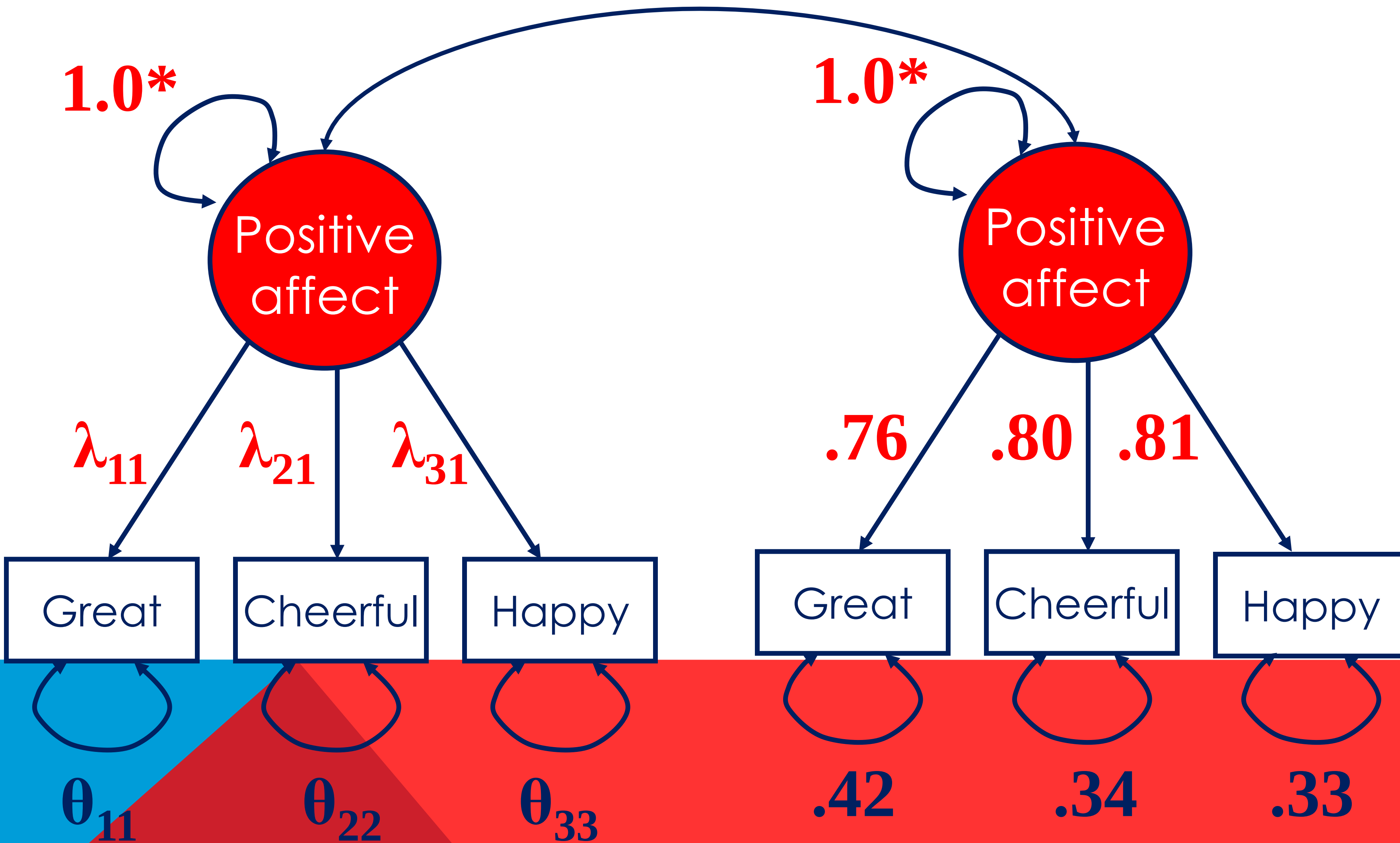


Implied variance/covariance matrix

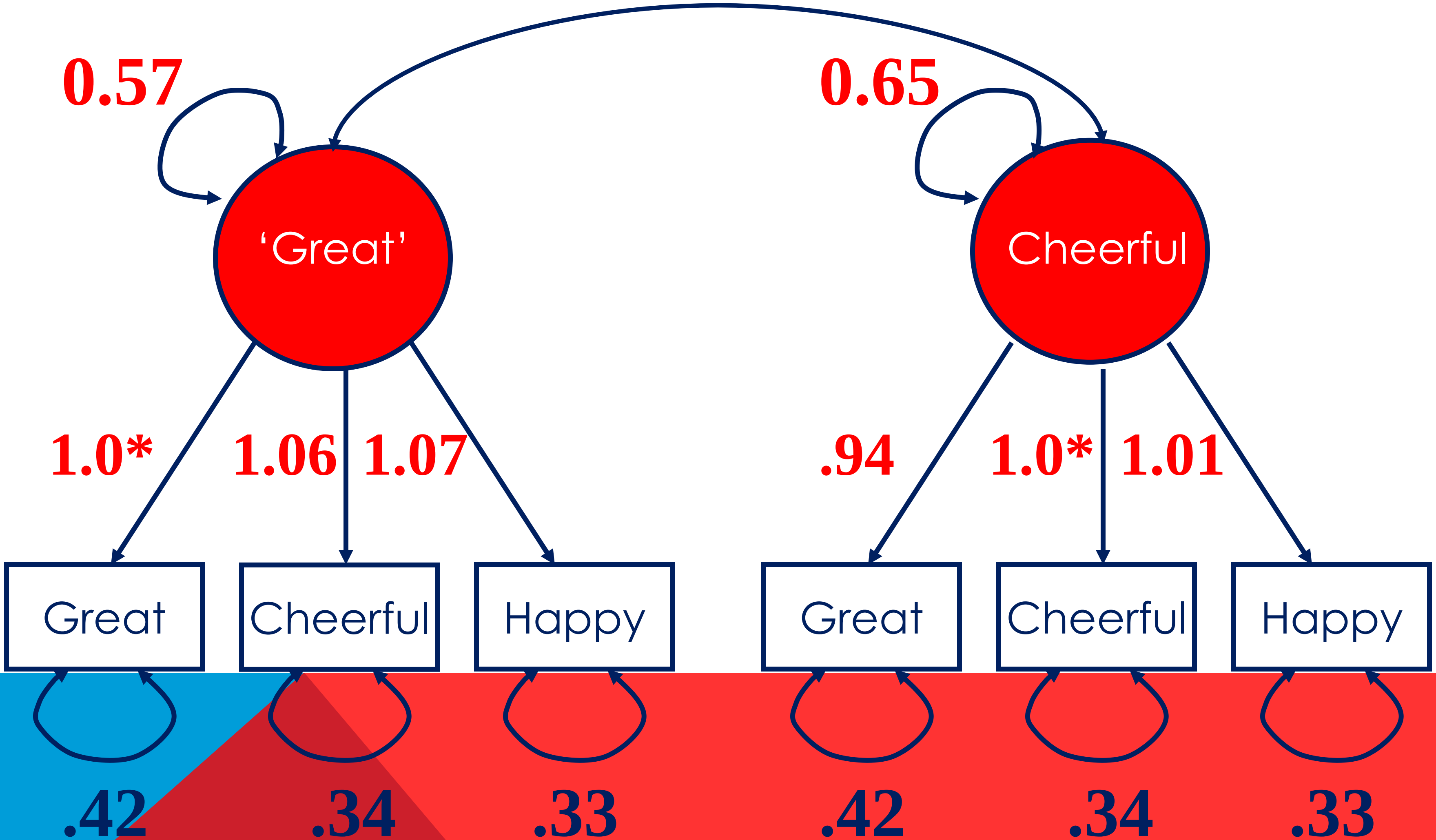
	X_1	X_2	X_3
X_1	$(3-\lambda_{21}- \lambda_{31})y_{11}(3-\lambda_{21}- \lambda_{31}) + q_{11}$		
X_2	$(3-\lambda_{21}-\lambda_{31})y_{11}\lambda_{21}$	$\lambda_{21}y_{11}\lambda_{21} + q_{22}$	
X_3	$(3-\lambda_{21}- \lambda_{31}) y_{11}\lambda_{31}$	$\lambda_{21}y_{11}\lambda_{31}$	$\lambda_{31}y_{11}\lambda_{31} + q_{33}$

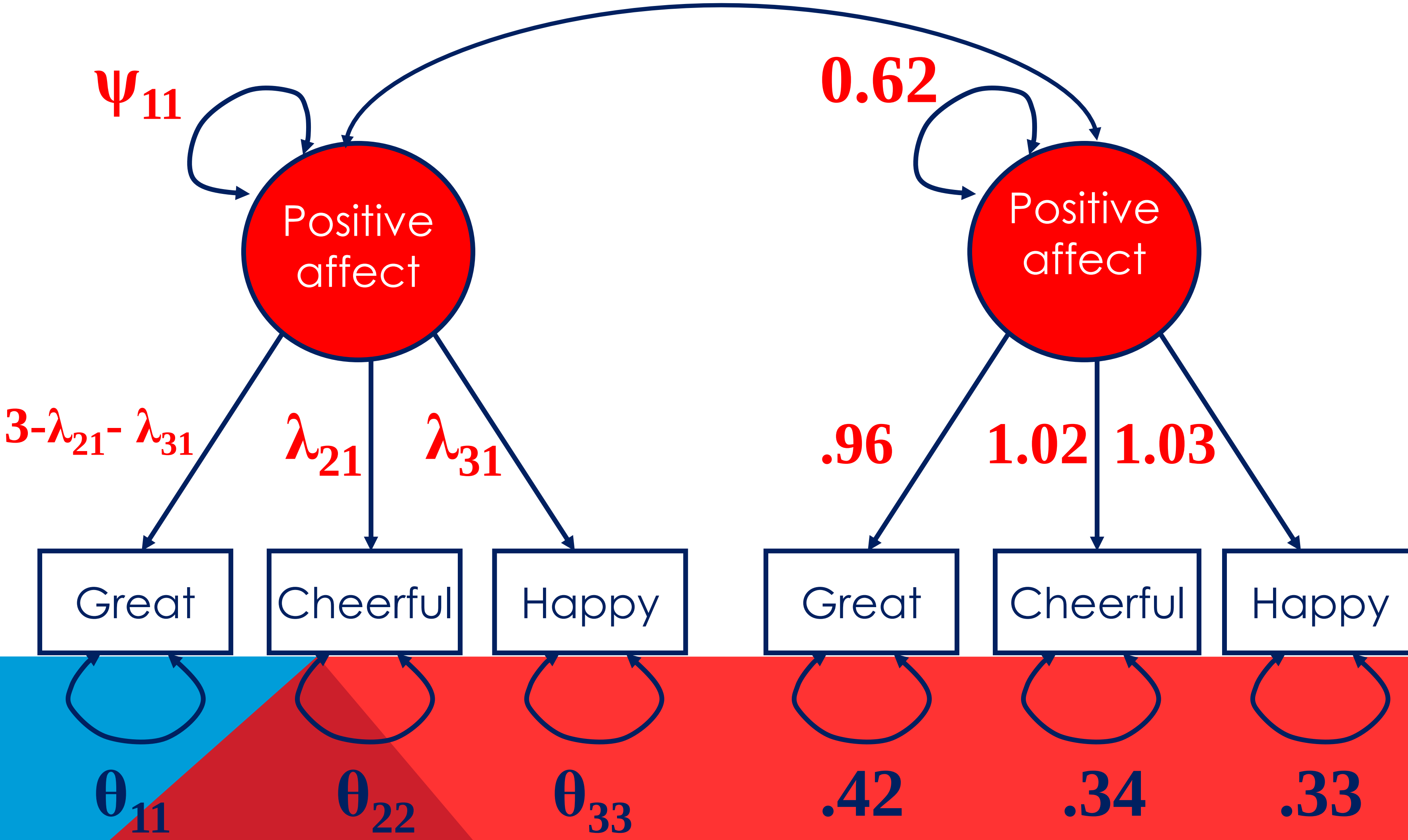
- Limitar las cargas a un promedio de 1.0
- La seleccion de cual carga es arbitraria

EJEMPLOS



1.0*
Positive affect
λ_{11}
λ_{21}
λ_{31}
Great
Cheerful
Happy
θ_{11}
θ_{22}
θ_{33}
1.0*
Positive affect
.76
.80
.81
Great
Cheerful
Happy
.42
.34
.33


0.57
0.65
'Great'
Cheerful
1.0*
1.06
1.07
.94
1.0*
1.01
Great
Cheerful
Happy
Great
Cheerful
Happy
.42
.34
.33
.42
.34
.33



ψ_{11}
0.62
Positive affect
Positive affect
$3-\lambda_{21}-\lambda_{31}$
λ_{21}
λ_{31}
.96
1.02
1.03
Great
Cheerful
Happy
Great
Cheerful
Happy
θ_{11}
θ_{22}
θ_{33}
.42
.34
.33

RECORDAR

El método de fijar la escala no afecta el ajuste del modelo

Afecta los estimados de los parámetros y su interpretación

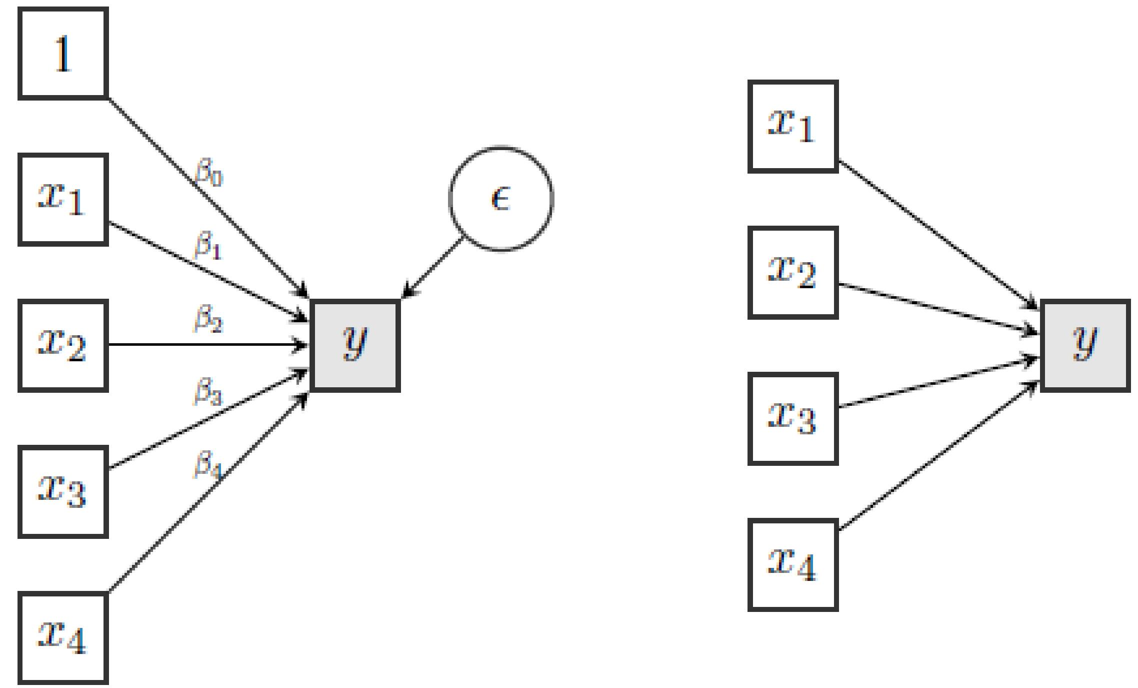
Usar:

- Effects coding: si la escala original es significativa
- Fixed factor: para obtener estimados estandarizados, libre de unidad de medida

INTRODUCCIÓN AL ANÁLISIS FACTORIAL CONFIRMATORIO (CFA)

DE REGRESIÓN A SEM

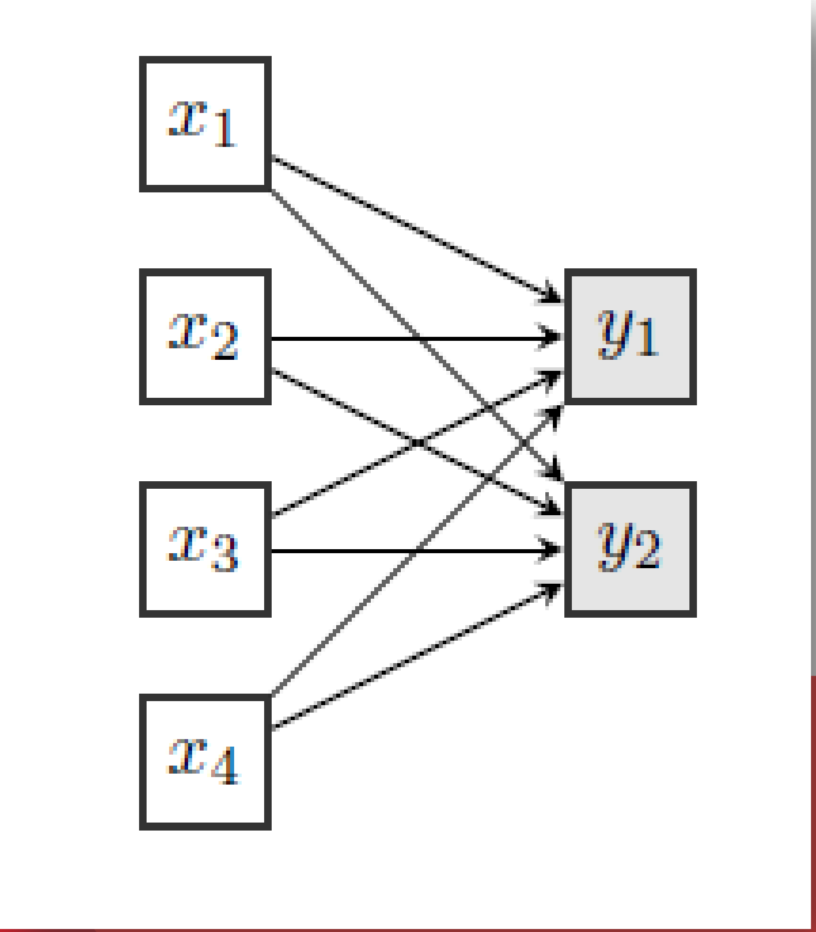
Regresión lineal univariada



$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \epsilon_i \quad (i = 1, 2, \ldots, n)$$

DE REGRESIÓN A SEM

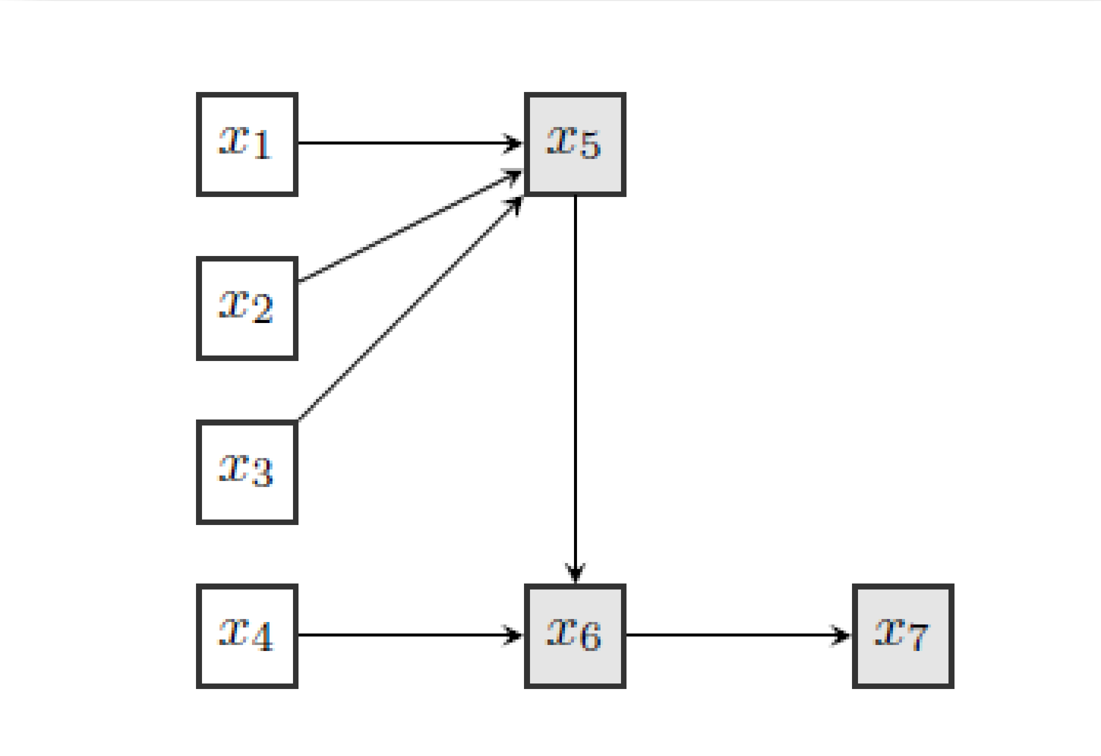
Regresión multivariada



PATH ANÁLISIS

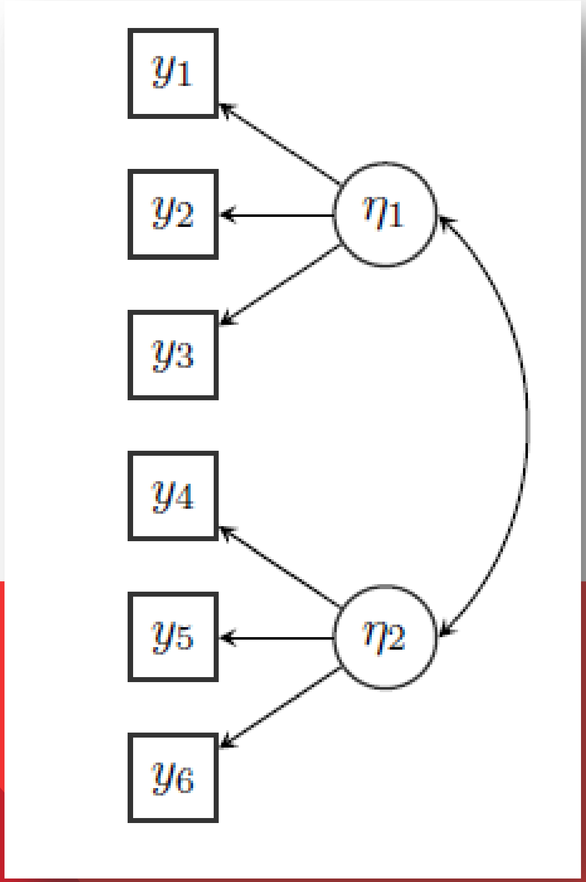
Probar modelos causales entre variables manifiestas

Sistema de ecuaciones de regresión



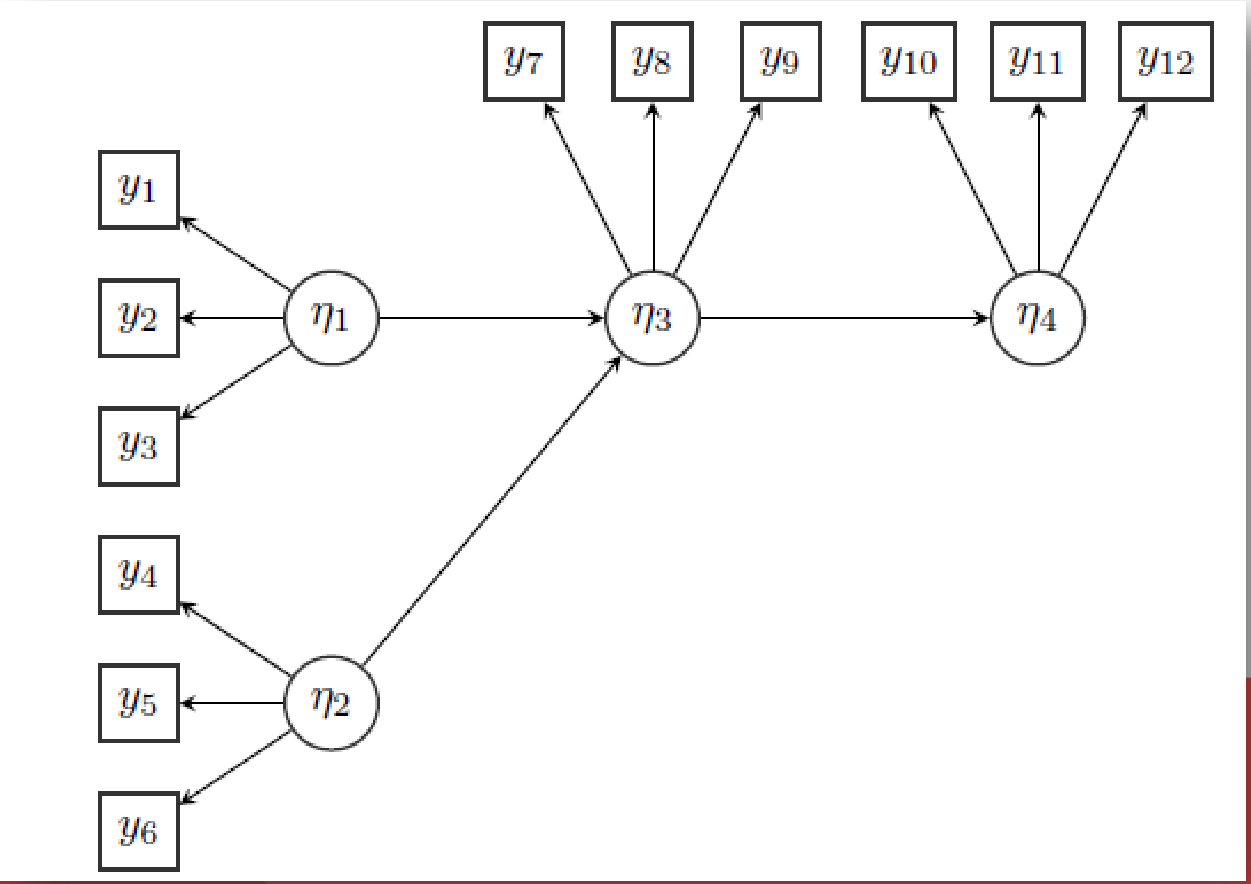
ANÁLISIS FACTORIAL CONFIRMATORIO (CFA)

Análisis factorial: representación de la relación entre una o más variables latentes y sus indicadores



MODELADO DE ECUACIONES ESTRUCTURALES (SEM)

Path análisis con variables latentes



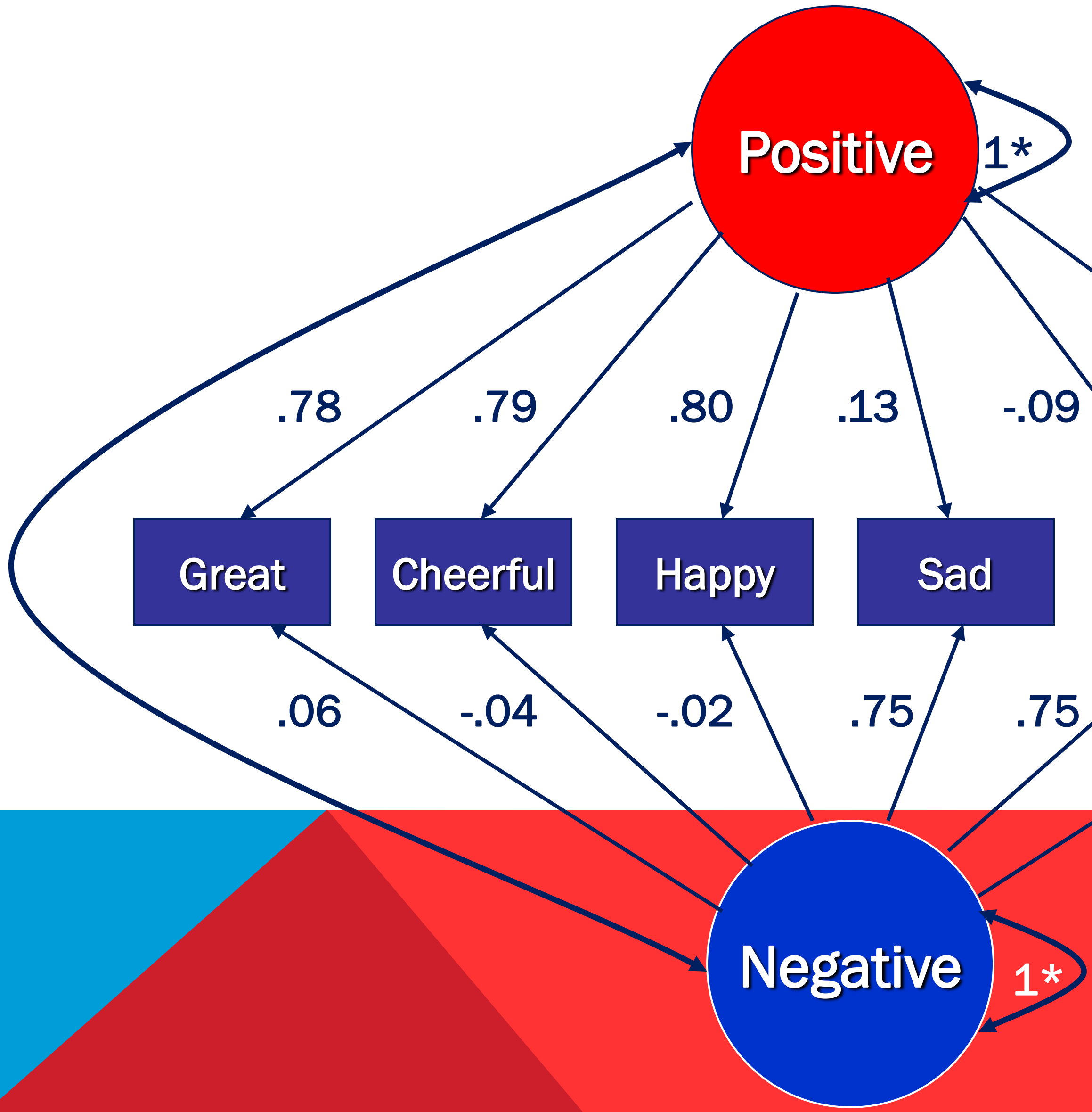
EFA VS CFA

EFA es exploratorio, ninguna expectativa de como se deben comportar los indicadores, empírico, sin ninguna teoría de base

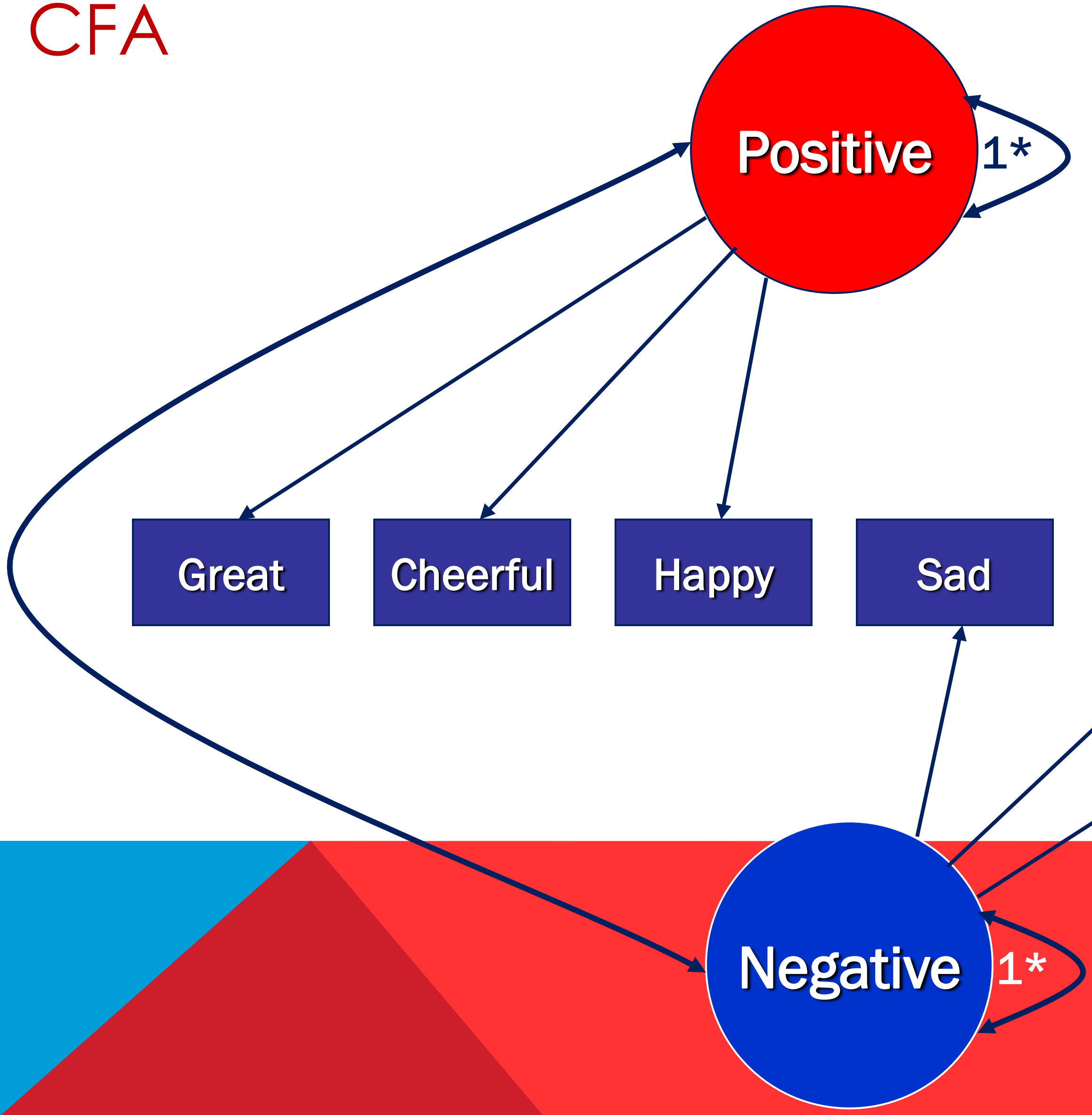
- Cuando de verdad uno no tiene ni la menor idea de como se van a comportar los indicadores? Cuando uno no tiene ninguna teoría en que basarse?

CFA se evalúa la estructura teórica que se espera

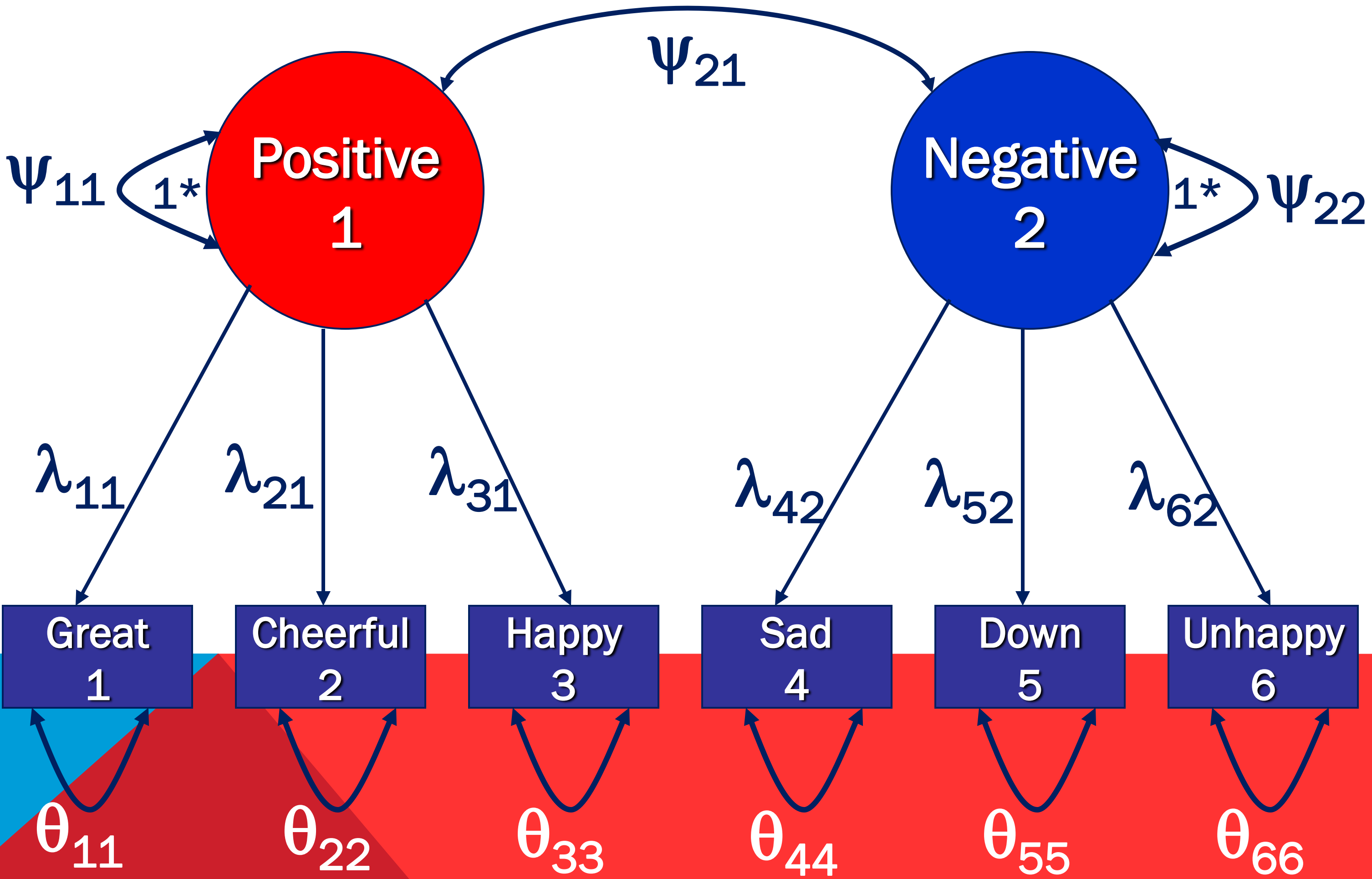
EFA EN NOTACIÓN DE SEM



CFA


Positive
1*
Great
Cheerful
Happy
Sad
Down
Unhappy
Negative
1*

CONSTRUCTOS OBLICUOS (CORRELACIONADOS)



ESTIMACIÓN DE LOS PARÁMETROS

Proceso de generación de parámetros óptimos basados en los datos observados

Métodos

- ***Maximum Likelihood (ML)***
  - ***MLM***
  - ***MLR***
- ***Generalized Least Squares (GLS)***
- ***Weighted Least Squares (WLS)***
- Ordinary Least Squares (OLS)
- Bayesiano

ÍNDICES DE AJUSTE

No hay un modelo verdadero

Modelos son aproximaciones parsimoniosas del fenómeno del mundo real

“A finding of good fit does not imply that a model is correct or true, but only plausible”

MacCallum& Austin, 2000

ÍNDICES DE AJUSTE

Tipos de índices de ajuste:

- Ajuste estadístico (χ^2)
- Ajuste absoluto (RMSEA)
- Ajuste relativo (CFI, NNFI)

Son las relaciones predichas por nuestro modelo consistentes con los datos observados?

ÍNDICES DE AJUSTE

χ^2 es un índice estadístico

- Prueba de ajuste exacto
- Representa la posibilidad de la hipótesis nula

Índices de ajuste practico

- Ajuste absoluto: sin referencia a un modelo nulo
- Ajuste relativo: relativo a un modelo nulo

CONTINUUM DE MODELOS

Ajuste absoluto: juzga la distancia desde el ajuste perfecto

Ajuste relativo: juzga la distancia del peor modelo posible

CHI-SQUARE: χ^2

$$\chi^2 = (N - 1)\, F_{ML}$$

…con $df = v\,(v + 1)/2 - p$

- v = # de indicadores; p = # de parámetros

Base de las otras estadísticas de ajuste

Influenciable por el tamaño de la muestra

H_0: Matrices observada y de modelo son iguales

H_1: Matrices observada y de modelo son diferentes

Meta: no rechazar la H_0

Recordar: todos los modelos son incorrectos a cierto grado

ÍNDICES DE AJUSTE ABSOLUTO

Root Mean Square Error of Approximation (RMSEA), <.08

- Compara ajuste con el modelo saturado
- Desajuste por df

Standardized Root Mean Residual (SRMR), <.08

- Media de los residuos estandarizados (cuadrados)

ÍNDICES DE AJUSTE RELATIVO

Compara el modelo con el modelo Null

Que tanto mejor es que el modelo Null

Modelo Null:

- Solo varianzas y medias de las variables observadas

Usados comúnmente:

- Tucker-Lewis Index (TLI)** recomendado
  → AKA Non-Normed Fit Index (NNFI)
- Comparative Fit Index (CFI)** recomendado
- Relative Noncentrality Index (RNI)
- Normed Fit Index (NFI)
- Incremental Fit Index (IFI)
- Relative Fit Index (RFI)

EVALUACIÓN DE LOS ÍNDICES DE AJUSTE

Tomar en cuenta muchos factores, no todos los índices tienen que ajustar

Cada índice tiene "problemas"

- TLI: tiene problemas con muestras pequeñas
- RMSEA: tiene problemas modelos pequeños, especialmente con muestras muy grandes
- SRMR: es bueno para modelos de un grupo de un solo tiempo de medida

EVALUACIÓN DE LOS ÍNDICES DE AJUSTE

Si χ^2 es no significativo, éxito

De otra manera examine los otros índices de ajuste

Revisar posibles cambios

Revisar índices de modificación

ÍNDICES DE MODIFICACIÓN

Cambio esperado en χ^2 si se cambia un parámetro

No son independientes, mas de uno puede arreglar el problema

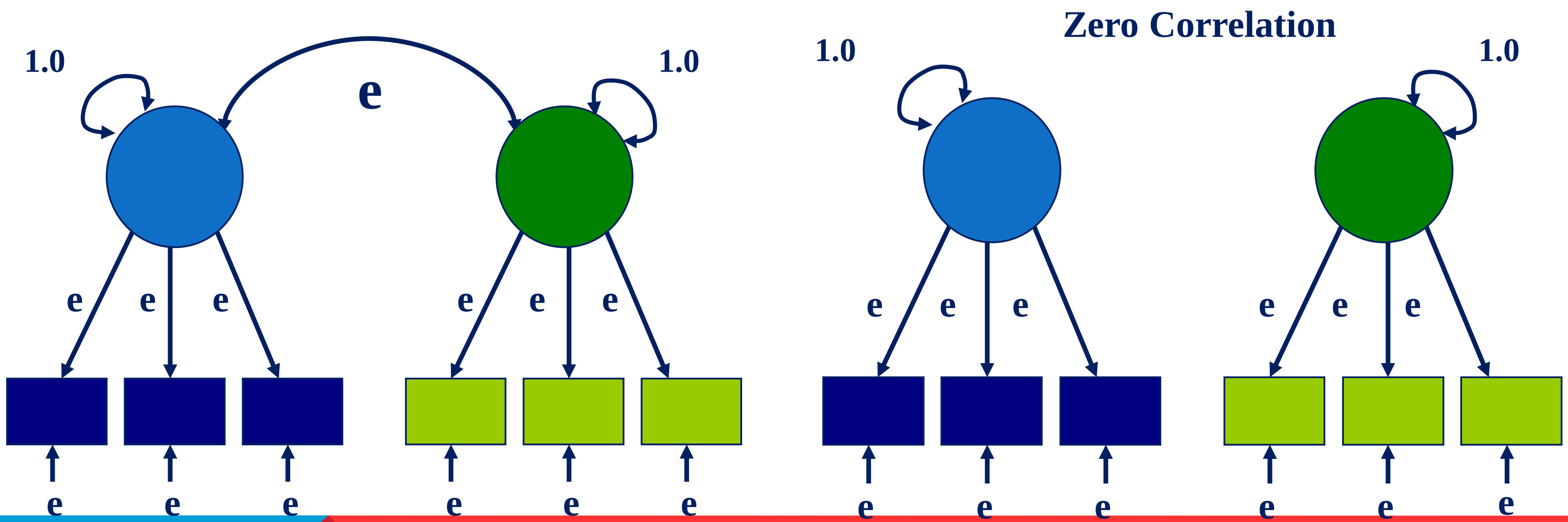
No son aditivos, haga uno a la vez

Solo aproxima del cambio en χ^2

COMPARANDO MODELOS ANIDADOS

Modelos anidados: cuando un modelo (A) es igual al otro (B), excepto por uno o dos parámetros en (A)

Modelo A esta anidado en el modelo B



COMPARANDO MODELOS ANIDADOS

Evaluar $\Delta\chi^2 = \chi^2(A) - \chi^2(B)$, con $\Delta df = df(A) - df(B)$

Si $\Delta\chi^2$ es no significativo: ambos modelos son igualmente buenos para explicar los datos, se escoge el modelo mas parsimonioso (menos parámetros, mas df)

Si $\Delta\chi^2$ es significativo: el modelo con mas parámetros explica mejor los datos

INTRODUCCIÓN A LAVAAN

QUE ES LAVAAN?

Un paquete de R para latent variable analysis:

- CFA, función cfa()
- SEM, función sem()
- Growth curve modeling, función growth()
- Modelado general, función lavaan()

Desarrollado para proveer una opción libre, de igual calidad que los programas comerciales

A largo plazo lavaan implementara todas las capacidades de los programas comerciales

'LAVAAN MODEL SYNTAX'

Tipos de formula:

formula type	operator	mnemonic
latent variable	`=~`	is manifested by
regression	`~`	is regressed on
(residual) (co)variance	`~~`	is correlated with
intercept	`~ 1`	intercept
defined parameter	`:=`	is defined as
equality constraint	`==`	is equal to
inequality constraint	`<`	is smaller than
inequality constraint	`>`	is larger than

MODELO COMPLETO

Escriba el modelo entre ''

```
mymodel<- '# regresiones
y1 ~ f1 + f2 + x1 + x2
f1 ~ f2 + f3
f2 ~ f3 + x1 + x2
# definición de variables latentes
f1 =~ y1 + y2 + y3
f2 =~ y4 + y5 + y6
f3 =~ y7 + y8 + y9 +y10
#varianzas y covarianzas
y1 ~~ y1
y1 ~~ y2
f1 ~~ f2
#intercepts
y1 ~1
f1 ~1'
```

EJEMPLO 1

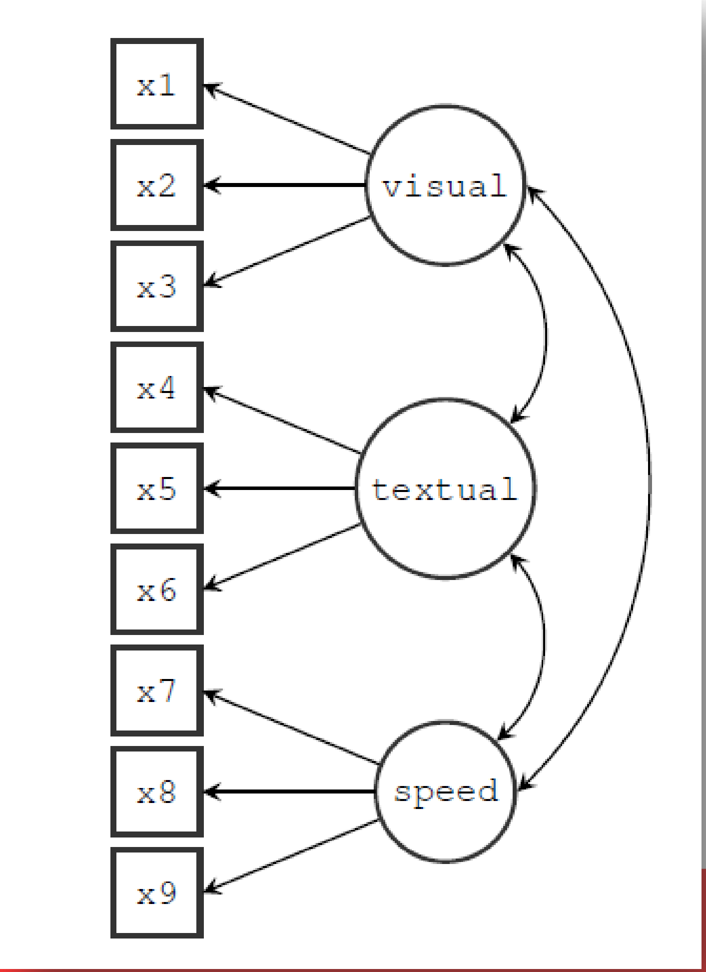
'visual =~ x1 + x2 + x3

textual =~ x4 + x5 + x6

speed =~ x7 + x8 + x9'

Estimar un modelo consiste en 3 pasos:

- Especificar el modelo
- Estimar el modelo (funciones: cfa, sem, growth, lavaan)
- Ver los resultados



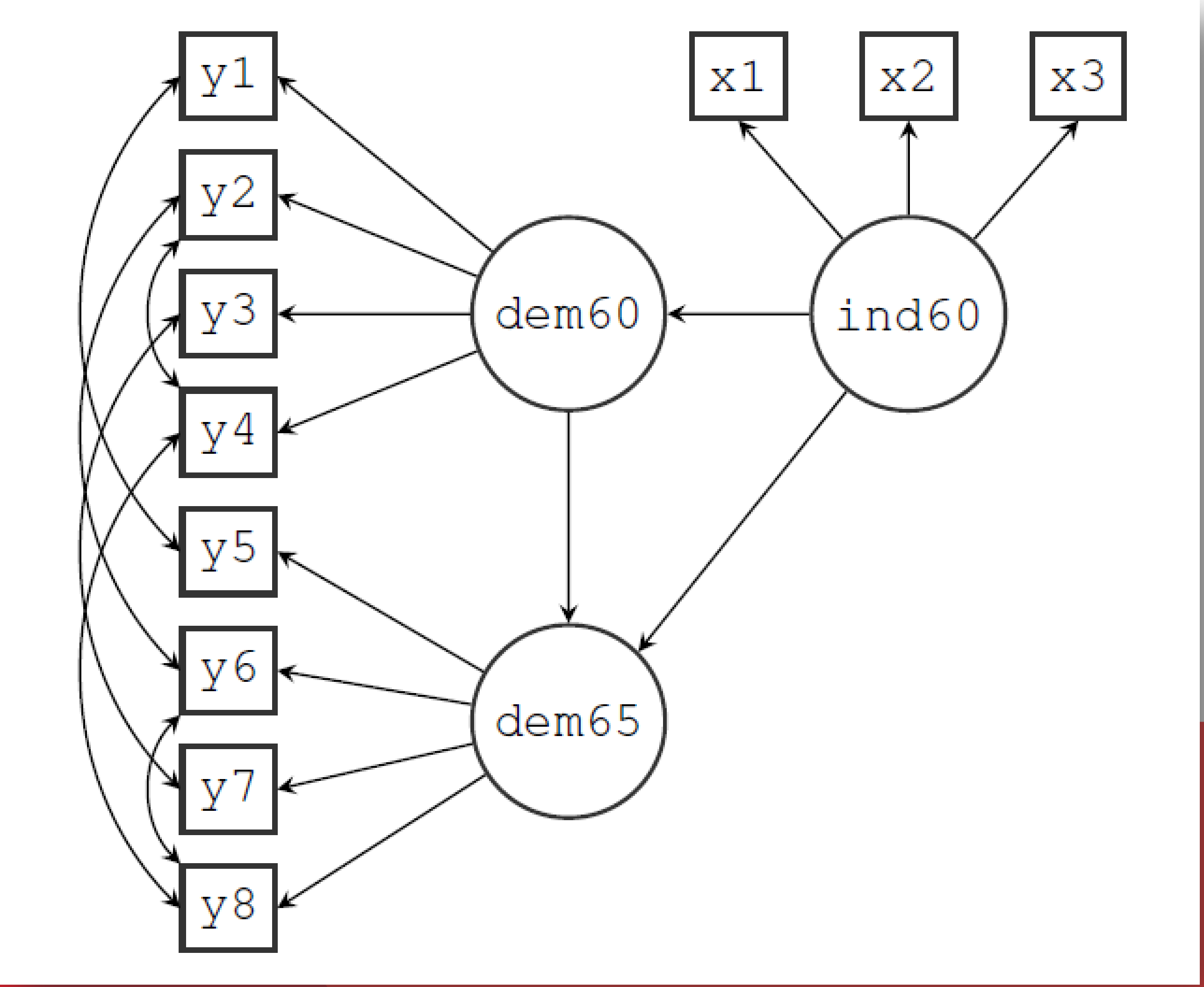
EJEMPLO 1

```
# 1 especificar el modelo
hs.model <- 'visual =~ x1 + x2 + x3
        textual =~ x4 + x5 +x6
        speed =~ x7 + x8 + x9'

# 2 estimar el modelo
fit <- cfa (hs.model, std.lv=T, data=HolzingerSwineford1939)

# 3 mostrar los resultados
summary(fit, fit.measures=T, standardized=T)
```

VEAMOS LAS FUNCIONES

- ? lavaan
- ?cfa
- ?sem
- ?growth
- ?fitMeasures
- ?parameterEstimates
- ?coef
- ?inspect
- ?modindices
- ?partable
- ?fitted

EJEMPLO 2



EJEMPLO 2

```
mod2 <- '
ind60 =~ x1 + x2 + x3
dem60 =~ y1 + y2 + y3 + y4
dem65 =~ y5 + y6 + y7 + y8

dem60 ~ ind60
dem65 ~ ind60 + dem60

y1 ~~ y5
y2 ~~ y4 + y6
y3 ~~ y7
y4 ~~ y8
y6 ~~ y8'

fit2 <- sem(mod2, std.lv=T, data=PoliticalDemocracy, meanstructure=T)
summary(fit2, fit.measures=T, standardized=T)
```

INSTALAR

simsem

- https://github.com/simsem/simsem/wiki

semTools

- https://github.com/simsem/semTools/wiki

Gracias