Parte 1: Invarianza de Medida

Peter F. Halpin

## Descripción General del Taller

* Parte 1. Introducción + análisis factorial + IM
* Parte 2. TRI + FDI
* Parte 3. Escalamiento robusto + FDI + FDP
* Abreviaturas
  + FDI = funcionamiento diferencial del ítem
  + FDP = funcionamiento diferencial de la prueba
  + TRI = teoría de respuesta al ítem
  + IM = invarianza de medida

## Descripción General de la Parte 1

* Definiciones generales de IM y FDI
* Modelo factorial para datos categóricos
* "Niveles" de IM definidos para datos categóricos
* Pruebas de IM comparando modelos (pruebas de diferencia de chi-cuadrado)
* Ejemplo práctico

## Organización

* Sitio web: [peterhalpin.github.io/RDIF-workshop/](https://peterhalpin.github.io/RDIF-workshop/)
* Diapositivas: Estas diapositivas
* Notas: Estas diapositivas en formato DOCX (traducido, editable)
* Código: Código de ejemplos de estas diapositivas + análisis adicionales

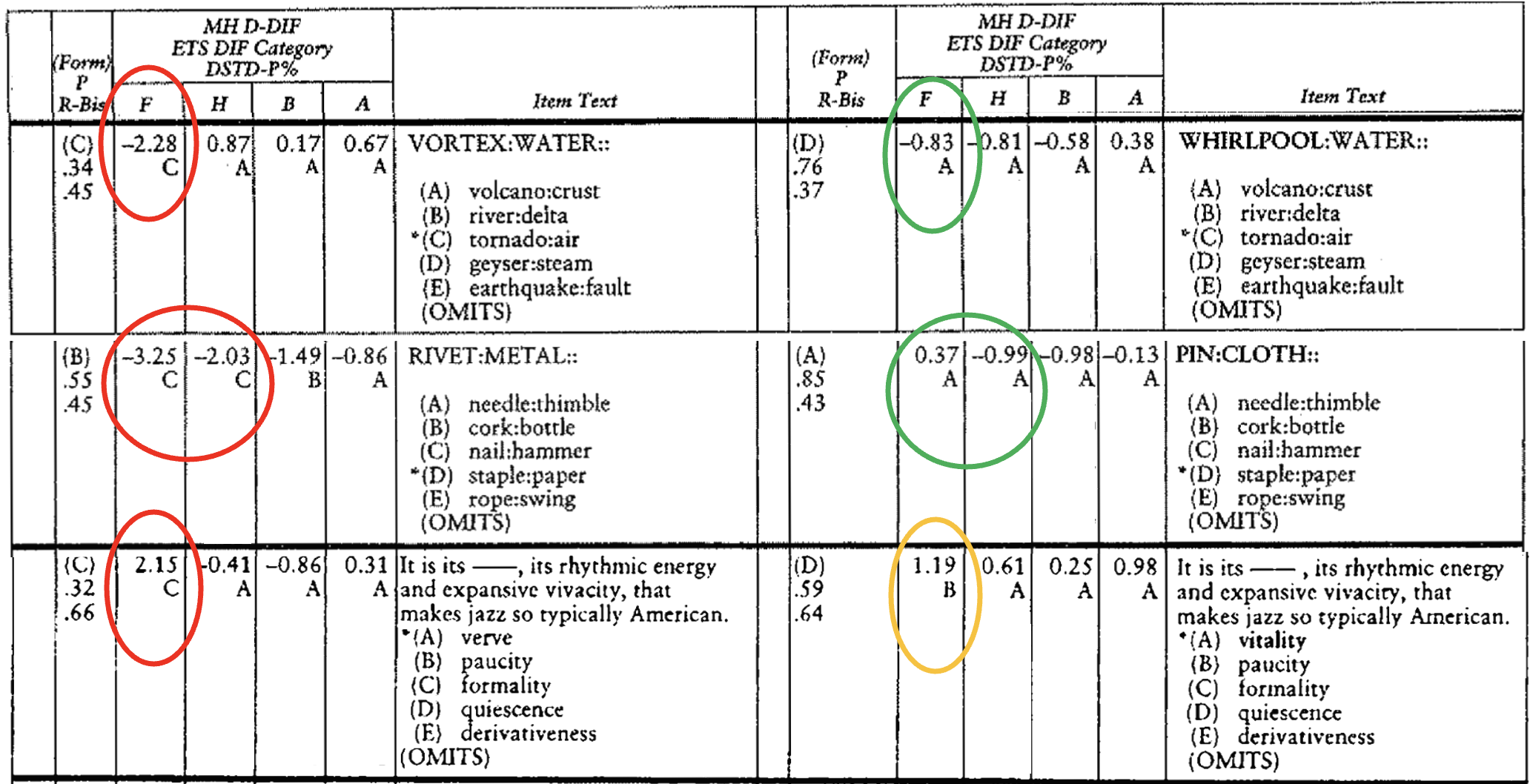
# IM y FDI en General

* Primer objetivo: Definir los problemas principales

## Definiciones intuitivas (Bauer, 2017)

* Invarianza de medida (IM): Una evaluación funciona igual a través de diferentes grupos de encuestados
* Funcionamiento diferencial del ítem (FDI): Un ítem funciona de manera diferente a través de diferentes grupos de encuestados
* Relación:
  + Si la IM se cumple, ningún ítem exhibe FDI
  + Si la IM no se cumple, al menos un ítem exhibe FDI

## Ejemplos (Curley & Schmitt, 1993)



## Un marco general

* Para discutir IM / FDI en general, represente modelos psicométricos usando la definición de distribución marginal
* : variables observadas (ítems de evaluación)
* : variable latente (rasgo, factor, constructo)
* : distribución de probabilidad (masa, densidad)

## Un marco general

* Diferentes suposiciones definen diferentes modelos (por ejemplo, Holland & Rosenbaum 1986)
* Análisis factorial: y están normalmente distribuidos con
* TRI: multinomial y es normal
* ….

## Un marco general

* Esta representación muestra que hay dos partes en un modelo psicométrico
* Útil para comprender IM / FD

## El modelo de medición

* La distribución condicional relaciona los datos observados con el rasgo latente
* Típicamente asumimos independencia condicional
* Las correlaciones entre los ítems se explican solo por el rasgo latente
  + "Los ítems miden el constructo"

## El modelo de población

* La distribución del rasgo latente describe cómo se distribuye el constructo objetivo
* Los modelos psicométricos requieren que establezcamos la escala del rasgo latente
  + p.ej., establecer y
* Esto será un aspecto complicado de IM / FDI
  + Diferentes niveles de IM permiten la estimación de diferentes parámetros de la población

## IM

* Sea cualquier otra variable
  + A menudo género, raza, pero podría ser cualquier cosa
* IM: para toda
* El modelo de medición no depende de
* “La medida no está sesgada con respecto a ”

## Implicaciones de IM

* El modelo marginal con IM:
* Si los grupos difieren en sus puntuaciones observadas, esto debe ser porque difieren en el rasgo latente

## FDI

* FDI: para el item
  + El modelo de medición **sí** depende de para algunos ítems
  + Esto es justo lo opuesto de IM
  + A veces llamado "sesgo de medición"

## Implicaciones de FDI

* Si los grupos difieren en sus puntuaciones observadas, esto podría ser porque:
  + 1. el modelo de población difiere entre grupos
    2. el modelo de medición difiere entre grupos
    3. ambos

## ¿Por qué el FDI es un problema?

* Por razones técnicas no podemos estimar este modelo si todos los ítems exhiben FDI
  + "La naturaleza circular del FDI"
  + Discutiremos más en la Parte 2
  + Por ahora, solo las implicaciones...

## ¿Por qué el FDI es un problema?

* Cuando calculamos y reportamos puntuaciones de pruebas usando x, estamos asumiendo implícitamente que los ítems no exhiben FDI (es decir, no están sesgados)
* Si esta suposición es errónea:
  + Las puntuaciones de los individuos pueden estar sesgadas
  + Las estimaciones de diferencias de grupo basadas en puntuaciones de pruebas observadas pueden estar sesgadas
  + Las estimaciones de impacto usando modelos de variables latentes pueden estar sesgadas
  + …

## ¿Qué pasa con la IM parcial?

* IM parcial significa que algunos, pero no todos, los ítems exhiben FDI
  + Mantener ítems sesgados puede estar bien en algunos entornos de investigación
* Pero el objetivo habitual del análisis de FDI es eliminar cualquier ítem con FDI
  + Es decir, el objetivo es IM completa, no IM parcial
  + Esta sigue siendo el enfoque estándar en el desarrollo de pruebas: eliminar ítems con FDI antes de informar los puntajes.

## ¿Qué pasa con los diferentes modelos?

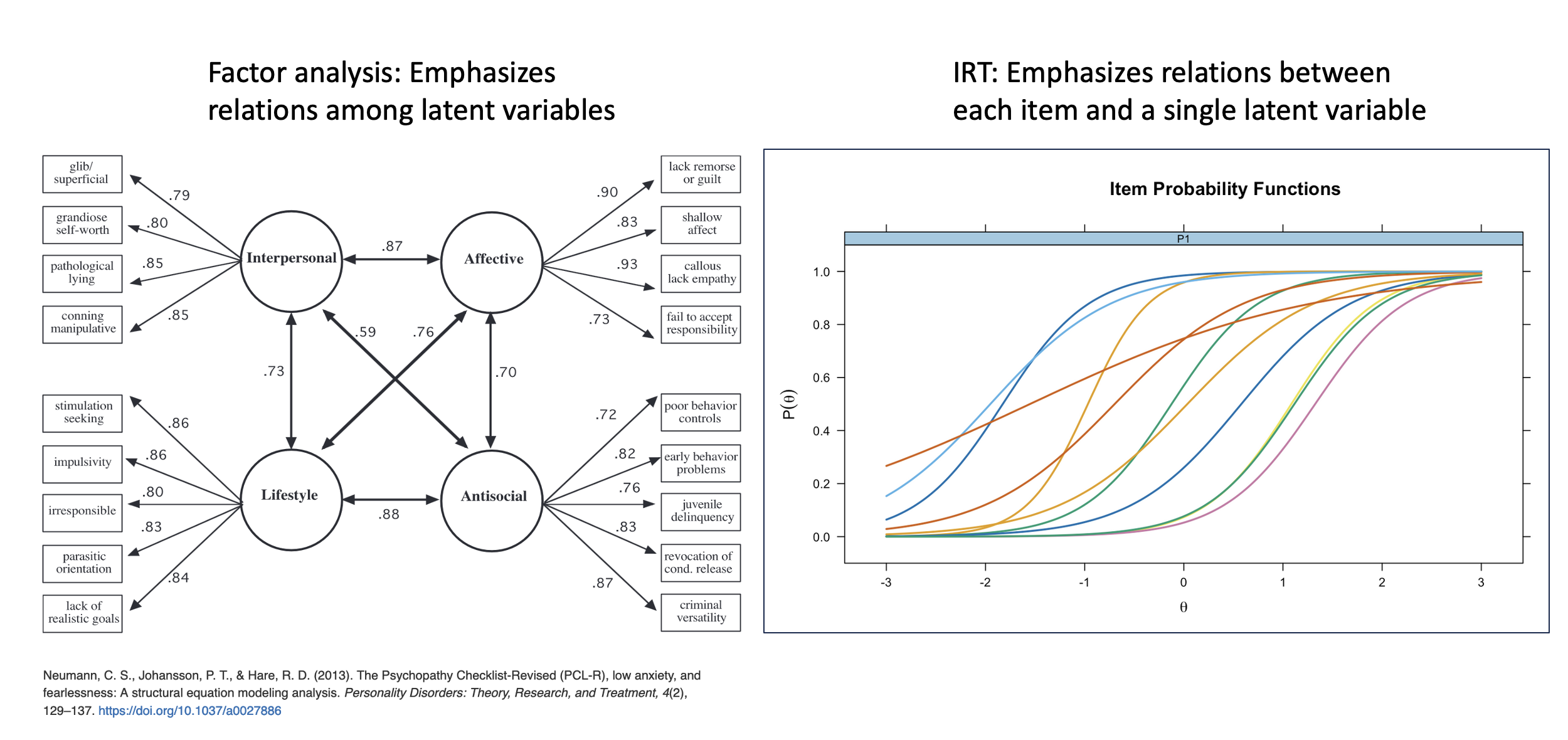
* Acabamos de ver cómo definir IM / FDI en general
* Sin embargo,
  + IM fue desarrollado en la literatura de análisis factorial
  + FDI fue desarrollado en la literatura de TRI
  + Ver Thissen (2023) para una revisión histórica

## ¿Qué pasa con los diferentes modelos?

* Los modelos, las pruebas, las convenciones y el software para IM y FDI difieren por razones históricas
* Podemos abordar tanto IM como FDI usando cualquiera de los modelos, pero actualmente es más fácil seguir con las distinciones tradicionales
  + ¡El software de análisis factorial facilita la prueba de IM!
  + ¡El software de TRI facilita la prueba de FDI!
  + Puedes cambiar esto, pero requiere (un poco) más de trabajo

## Comparación amplia entre modelos

| Característica | Análisis factorial | TRI |
| --- | --- | --- |
| Dimensionalidad del rasgo latente | **multidimensional** | Unidimensional (¡tradicionalmente!) |
| Tratamiento de datos categóricos | Variables de respuesta latentes | **Funciones de respuesta del ítem** |
| Estimación del modelo | Correlaciones policóricas (WLS) | Máxima verosimilitud |
| Parametrización del modelo | General, muchos "parámetros adicionales" | Específico, solo incluye parámetros utilizados en un modelo dado |
| Visualización principal | Diagrama de caminos | Funciones de respuesta del ítem |

**Comparación amplia entre modelos**

## Resumen

* IM / FDI tratan sobre el modelo de medición
  + Queremos asegurarnos de que la medición no dependa, por ejemplo, del género de una persona
* El impacto trata sobre el modelo de población
  + Puede o no haber diferencias de grupo en el constructo objetivo
* Sin IM (parcial), no podemos saber si las diferencias observadas se deben al sesgo de medición, diferencias "verdaderas" en el constructo objetivo, o ambos

# Análisis Factorial para Datos Categóricos

|  |
| --- |

* Enfoque en modelos unidimensionales

## Modelo factorial

* Para variables observadas continuas
* Suposiciones

. . .

* ¡No funciona cuando es categórica!

## Variables de respuesta latentes (VRL)

* El análisis factorial trabaja con datos categóricos introduciendo un nuevo tipo de variable latente
* Para cada variable observada categórica , asumimos que existe una variable de respuesta latente
  + ¡No es una variable de interés sustantivo, solo una comodidad matemática!
  + Ilustraciones en las siguientes diapositivas

## VRL: ¿Por qué?

* Beneficio: puede hacer análisis factorial "como de costumbre" con
  + Cuando la vida te da limones, haz limonada
  + Cuando la vida te da datos categóricos, haz datos continuos
* Costo: introduce nuevas variables (y sus parámetros) que no significan nada
  + Será un poco molesto más adelante

# VRL: Cómo funcionan

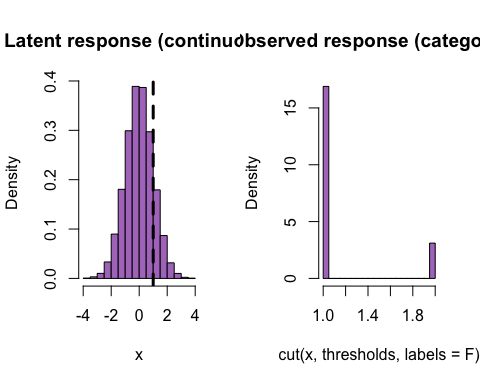
| *Figura: Wirth & Edwards, 2007* |
| --- |

## Dos ideas principales

* Primera idea: Definir el umbral de una variable de respuesta latente
  + Asume que las VRL están normalmente distribuidas
  + Nos permite manejar datos categóricos
* Segunda idea: correlaciones tetracóricas y policóricas
  + Asume que pares de VRL son bivariadamente normales
  + Nos permite estimar el modelo factorial

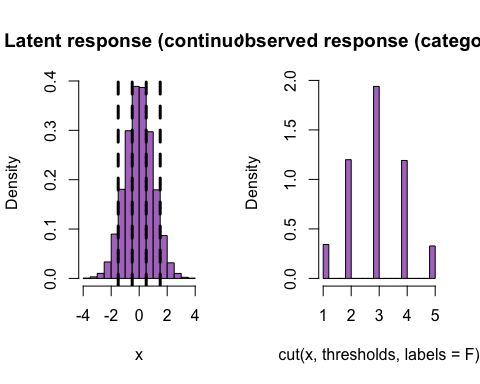
## Umbralización

x <- rnorm(20000)  
par(mfrow = c(1,2))  
hist(x, col = "#af7ac5", freq = F, main = "Latent response (continuous)")  
thresholds <- c(-10, 1, 10)  
abline(v = thresholds, lty = 2, lwd = 3, col = 1)  
hist(cut(x, thresholds, labels = F), col = "#af7ac5", freq = F, main = "Observed response (categorical)")



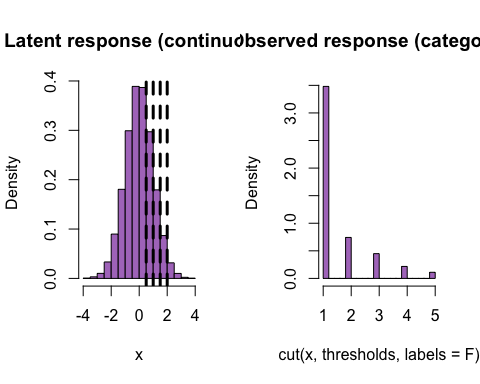
## Definir el umbral de una variable latente

par(mfrow = c(1,2))  
hist(x, col = "#af7ac5", freq = F, main = "Latent response (continuous)")  
thresholds <- c(-10, -1.5, -.5, .5, 1.5, 10)  
abline(v = thresholds, lty = 2, lwd = 3, col = 1)  
hist(cut(x, thresholds, labels = F), col = "#af7ac5", freq = F, main = "Observed response (categorical)")



## Definir el umbral de una variable latente

par(mfrow = c(1,2))  
hist(x, col = "#af7ac5", freq = F, main = "Latent response (continuous)")  
thresholds <- c(-10, .5, 1, 1.5, 2, 10)  
abline(v = thresholds, lty = 2, lwd = 3, col = 1)  
hist(cut(x, thresholds, labels = F), col = "#af7ac5",freq = F, main = "Observed response (categorical)")



## Correlación tetracórica

knitr::include\_graphics("tetrachoric.png")

|  |
| --- |

* Correlación de respuestas observadas: Coeficiente Phi
* Correlación de respuestas latentes: Correlación tetracórica

## Correlación policórica

knitr::include\_graphics("polychoric.png")

|  |
| --- |

* Correlación de ítems observados: Spearman, …
* Correlación de respuestas latentes: Correlación policórica

## Resumen

* Las VRL se utilizan para tratar con datos categóricos en el análisis factorial
* Las correlaciones entre las VRL se modelan en lugar de modelar directamente los datos categóricos
  + Estas se llaman correlaciones tetracóricas y policóricas
* ¡Todo esto se hace por conveniencia matemática!
  + Las VRL (usualmente) no representan conceptos sustantivos
  + No aparecen en TRI, que es una diferencia principal entre los modelos

# Volvemos al Modelo Factorial…

## Modelo factorial para datos categóricos

* Paso 1: Asumir que las variables categóricas con categorías surgen del umbral de una VRL
* Los parámetros *se llaman los umbrales del ítem*

## Modelo factorial para datos categóricos

* Paso 2: Modelo factorial para las VRLs
* Suposiciones
* Igual que el modelo continuo, pero ahora para las VRLs

## Identificación del Modelo

* La identificación del modelo para datos categóricos es complicada
* Las VRLs introducen muchos parámetros que no podemos estimar realmente
* Versión corta: En el caso de un solo grupo, los únicos parámetros que podemos estimar son
  + las cargas factoriales
  + los umbrales

## Identificación del Modelo

* Se complica más al probar para IM
* Podemos estimar algunos de los parámetros excluidos, pero diferentes autores usan diferentes enfoques
* Así que, para estar preparados para IM, ayuda revisar la versión larga de este problema para el caso de un solo grupo…

## Identificación del Modelo

* Estandarizar el rasgo latente como de costumbre:
  + por ejemplo, establecer y
  + Podemos fijar uno de los interceptos y las cargas factoriales en 1 en su lugar
* Para las VRLs, también debemos establecer su escala, y hay dos maneras de hacer esto
  + “Parametrización Delta” - recomendado para interpretación, por defecto en lavaan
  + “Parametrización Theta” - puede simplificar la estimación, no se discutirá mucho

## Parametrización Delta

* Estandarizar las VRLs como
  + es decir, establecer y
  + "Delta" se define como , por lo tanto, es equivalente a establecer
* Equivalente a estandarizar los datos continuos
  + Las cargas factoriales se pueden interpretar como correlaciones
  + Los umbrales se pueden interpretar como puntuaciones-z

## Implicaciones de la Parametrización Delta

* Establecer implica que los interceptos también son cero:
* Así que, $\_j = 0 $
* Implicación: los interceptos de las VRLs no pueden ser estimados (fijados a cero)

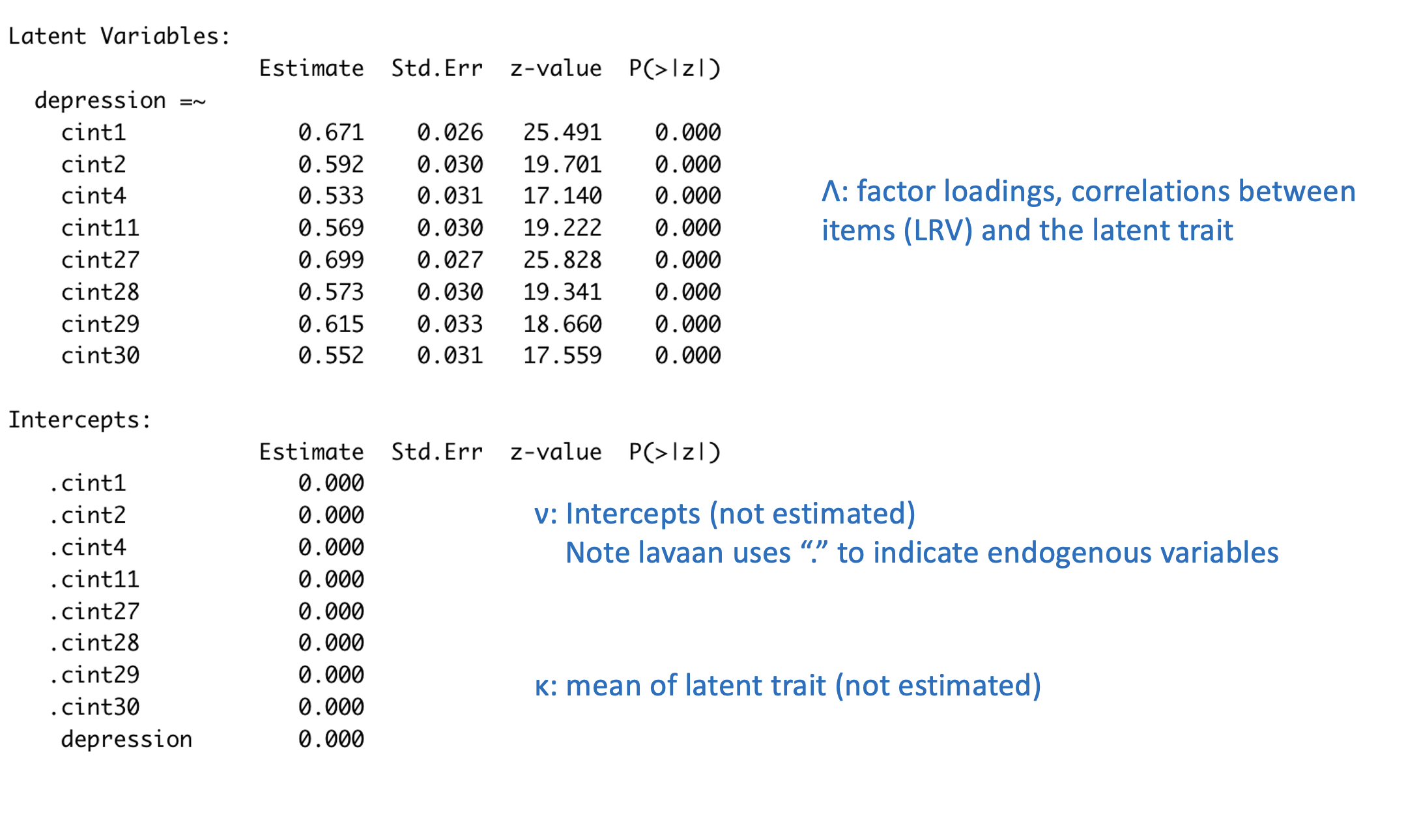
## Implicaciones de la Parametrización Delta

* Establecer implica el valor de las varianzas residuales
* Así que
* Implicación: la varianza residual del modelo factorial no puede ser estimada (fija a )

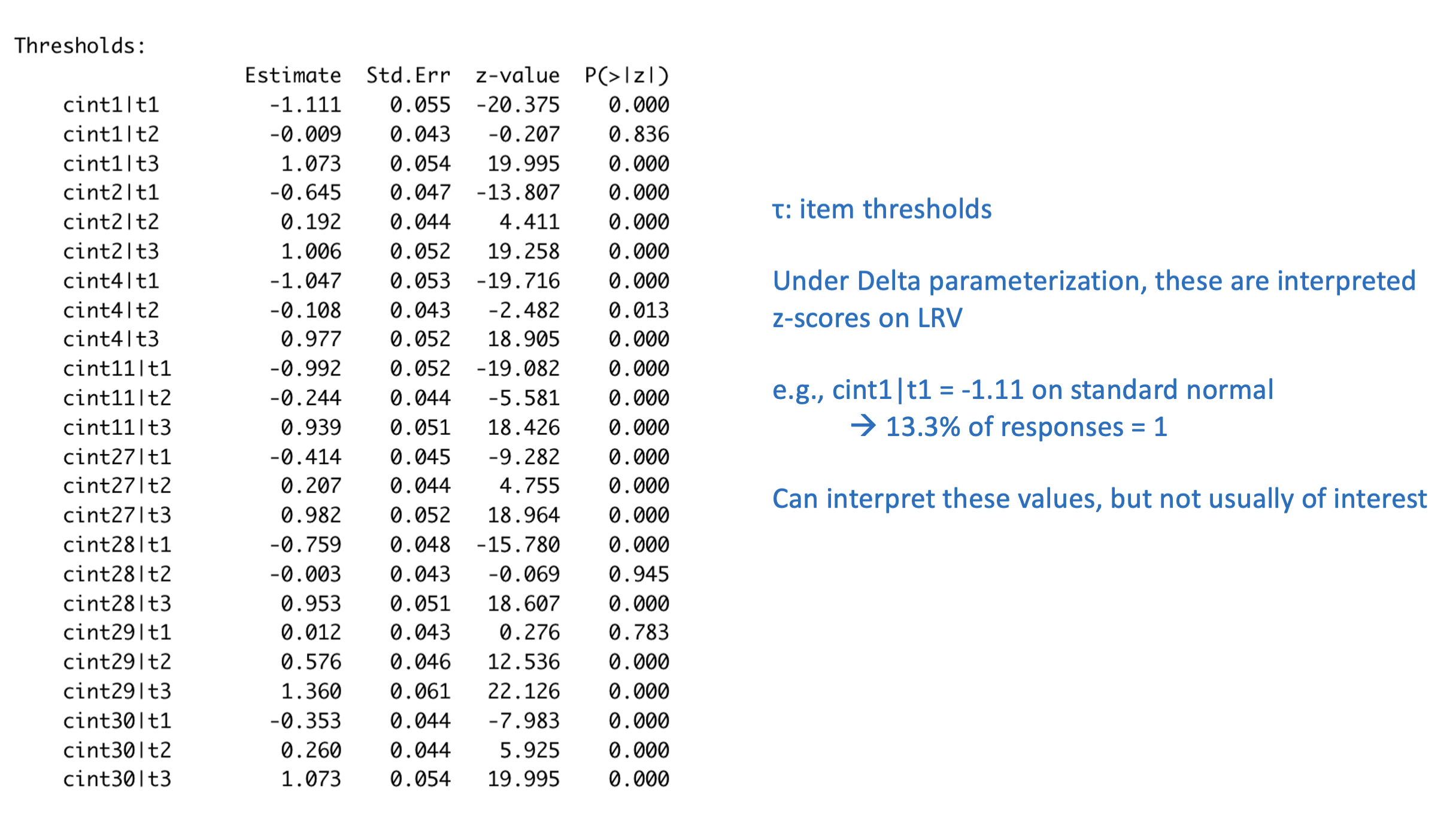
## Estimación en lavaan: Código

library(lavaan)  
dat <- read.csv("cint\_data.csv")  
  
# Model syntax  
mod1 <- 'depression =~ cint1 + cint2 + cint4 + cint11 +   
 cint27 + cint28 + cint29 + cint30'  
  
# Fit model  
fit.delta <- cfa(mod1,   
 data = dat,   
 std.lv = T, # standardize latent variable  
 ordered = T) # data are ordered  
   
# Print model summary  
summary(fit.delta)

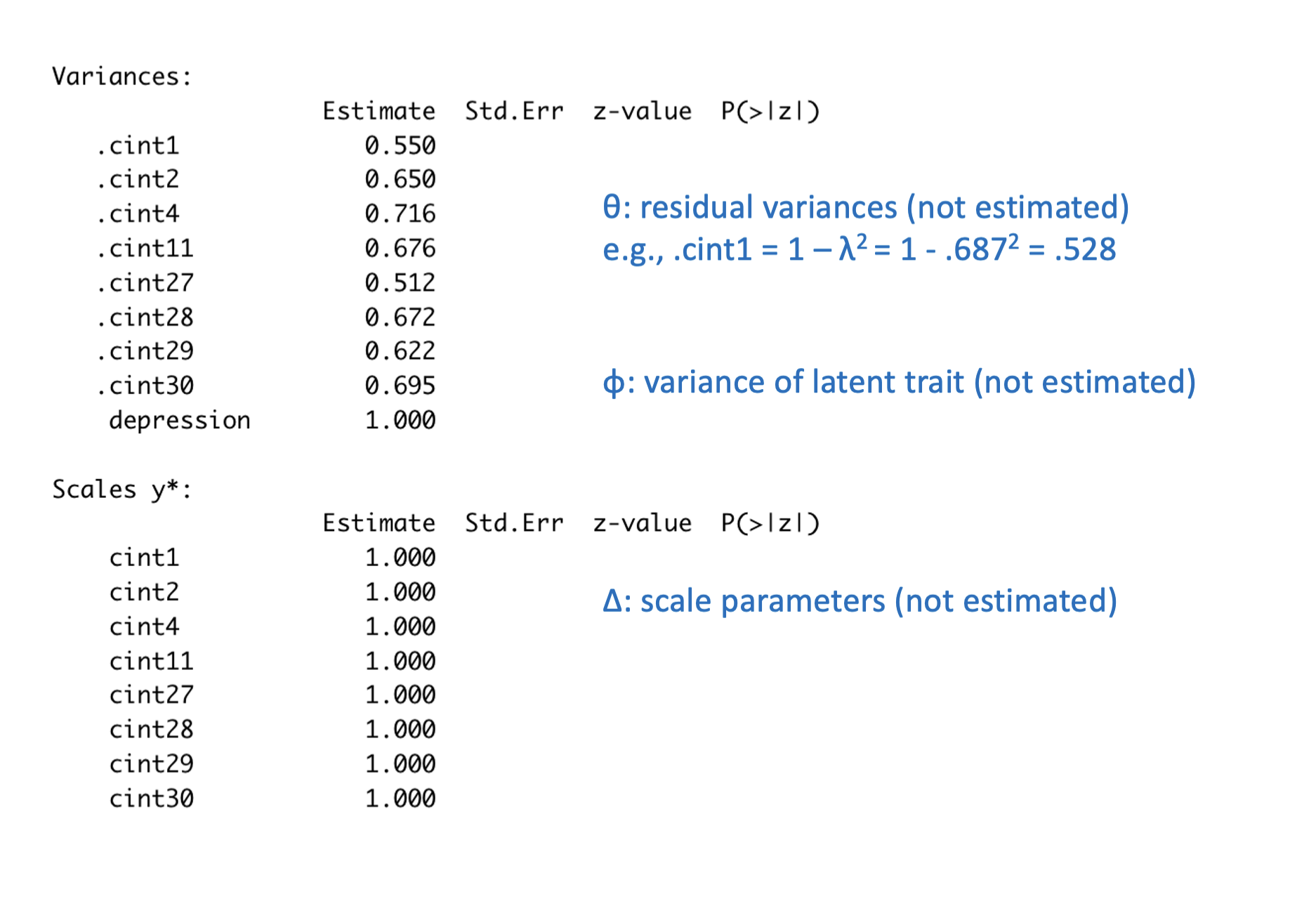
## Estimación en lavaan: Salida



## Estimación en lavaan: Salida



## Estimación en lavaan: Salida



## Resumen de la Parametrización Delta

* Los únicos parámetros que estimamos son las cargas factoriales y los umbrales
* El rasgo latente está estandarizado: ,
* Las VRLs están estandarizadas:
  + Lo mismo que establecer
* Los interceptos están fijados,
* Las varianzas residuales están fijadas,
  + ¡En algunos modelos de IM, se puede estimar , pero sigue fijo!

## Parametrización Theta

* En lugar de establecer , fijamos la varianza residua l
* Implica que la varianza de está fijada a
* Entonces el parámetro Delta está fijado a
* La interpretación del modelo es más complicada ya que
* Ver notas de codificación para ejemplos

## Resumen

* El modelo factorial para datos categóricos utiliza VRLs
  + Paso 1. Representar datos categóricos usando VRL
  + Paso 2. Analizar factorialmente las VRLs
* ¡Truco conveniente! Pero introduce muchos parámetros que no podemos estimar
* Los únicos parámetros que podemos estimar son las cargas factoriales y los umbrales
* ¡En Delta, estos son fáciles de interpretar!
* (En Theta, no son fáciles de interpretar)

# Invarianza de Medida

|  |
| --- |

## Recapitulación de IM

* *Queremos que nuestro modelo de medición sea el mismo a través de los grupos*
* Esto asegurará que cualquier diferencia de grupo en los datos observados $\bf x$ se deba únicamente a diferencias en el constructo objetivo (es decir, impacto)
* Importante para asegurar comparaciones imparciales entre grupos

## Parámetros del modelo de medición

* Para grupos
* Los pesos factoriales,
* Los umbrales de los ítems,
* Queremos probar si estos son iguales a través de los grupos:

## Parámetros del modelo de población

* En un solo grupo, tuvimos que estandarizar para estimar el modelo
* En múltiples grupos, este enfoque es problemático
* p.ej., si establecemos la media del factor en 0 en cada grupo:
* ¡Estamos afirmando que todos los grupos tienen la misma media en el rasgo latente – esto no es una restricción “arbitraria” en el modelo!

## Parámetros del modelo de población

* La IM nos permite estimar los parámetros del modelo de población (ver Muthen & Asparouhov 2002, Millsap & Yun-Tein 2004)
* De hecho, el objetivo de la IM puede interpretarse en términos de poner suficientes restricciones en el modelo para estimar el impacto
  + Más sobre esto pronto cuando hablemos de "niveles" de IM
* Incluso con IM, todavía necesitamos estandarizar en un grupo, llamado grupo de referencia

## Parámetros molestos/de estorbo

* ¿Qué pasa con los parámetros de las VRL?

  + Los interceptos,
  + Las varianzas residuales,
* Estos son técnicamente parte del modelo de medición
* Con IM podemos estimar ya sea (Delta) o la varianza residual (Theta)
  + La mayoría de softwares harán esto por defecto…

## Resumen

* Los parámetros de medición:
  + Los pesos factoriales,
  + Los umbrales de los ítems,
* Los parámetros de población:
* Los parámetros molestos/ de estorbo (VRL):
  + ; interceptos: ; varianzas residuales:

# Niveles de Invarianza de Medida

* configuracional, débil, métrica, escalar, fuerte, estricta, …

## Hay muchas versiones…

| *Tabla: Thissen, 2023* |
| --- |

## Resumen de niveles: Invarianza Configuracional

* **Modelo de medición**: Mismo patrón factorial sobre grupos (qué ítems van con qué factores)
* **Modelo de población**: No es suficiente para estimar el impacto en ningún parámetro
* Generalmente no se interpreta, pero es la base para probar otros modelos

## Resumen de niveles: Invarianza Débil / Métrica

* **Modelo de medición**: Todos los pesos factoriales son iguales sobre los grupos
* **Modelo de población**: Suficiente para estimar el impacto en las (co-) varianzas del factor
* Puede servir como base para el modelado de ecuaciones estructurales multigrupo (sin estructura media)

## Resumen de niveles: Invarianza Fuerte / Escalar

* **Modelo de medición**: Todos los pesos factoriales y umbrales son iguales a través de los grupos
* **Modelo de población**: Suficiente para estimar el impacto en las (co-) varianzas y medias del factor
* Considerado aceptable para comparar grupos en puntuaciones de pruebas observadas

## Resumen de niveles: Invarianza Estricta

* **Modelo de medición**: Todos los pesos factoriales, umbrales y varianzas residuales son iguales a través de los grupos
* **Modelo de población**: Suficiente para estimar el impacto en las (co-) varianzas y medias del factor
* Asegura que las puntuaciones de las pruebas son igualmente confiables en ambos grupos
* Nota: algunos problemas al distinguir la invarianza fuerte y estricta con datos categóricos (lo veremos pronto)

## El modelo configuracional: Recapitulación

* **Modelo de medición**: Mismo patrón factorial sobre los grupos (qué ítems van con qué factores)
* **Modelo de población**: No es suficiente para estimar el impacto en ningún parámetro
* Generalmente no se interpreta, pero es la base para probar otros modelos

## Código del modelo configuracional

# Model (same as above)  
mod1 <- ' depression =~ cint1 + cint2 + cint4 + cint11 +   
 cint27 + cint28 + cint29 + cint30'  
# Fit model  
fit.config <- cfa(mod1,   
 data = dat,   
 std.lv = T,   
 ordered = T,   
 group = "cfemale") # <--- new   
   
# Print model summary  
summary(fit.config)

**Salida del modelo configuracional**

|  |
| --- |

## Invarianza Débil / métrica: Recapitulación

* **Modelo de medición**: Todos los pesos factoriales son iguales a través de los grupos
* **Modelo de población**: Suficiente para estimar el impacto en las (co-) varianzas del factor
* Este modelo es más emocionante en configuraciones multidimensionales cuando estamos interesados en la matriz de covarianza de los factores, no solo en la varianza de un único factor

## Código de invarianza débil / métrica

# Fit model  
fit.weak <- cfa(mod1,   
 data = dat,   
 std.lv = T,   
 ordered = T,   
 group = "cfemale",  
 group.equal = "loadings") # <--- new   
   
# Print model summary  
summary(fit.weak)

## Salida de invarianza débil / métrica

|  |
| --- |

## Comparación de los modelos

* Los modelos CFA anidados se pueden comparar usando sus estadísticas chi-cuadrado (por ejemplo, Satorra & Bentler, 2001)
* Dos modelos están anidados si uno puede obtenerse del otro fijando algunos parámetros a valores fijos
* Deje que "A" denote el modelo más grande y "B" denote el modelo más pequeño
* Defina:
* Entonces tiene una distribución chi-cuadrado central con cuando el modelo restringido es verdadero

## Código para comparar los modelos

lavTestLRT(fit.config, fit.weak)

Scaled Chi-Squared Difference Test (method = "satorra.2000")  
  
lavaan NOTE:  
 The "Chisq" column contains standard test statistics, not the  
 robust test that should be reported per model. A robust difference  
 test is a function of two standard (not robust) statistics.  
   
 Df AIC BIC Chisq Chisq diff Df diff Pr(>Chisq)   
fit.config 40 37.534   
fit.weak 47 57.126 12.091 7 0.09761 .  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

. . .

* No rechazar la invarianza débil usando
* 8 pesos factoriales restringidos, se estimó 1 varianza para el rasgo latente, así que

## Resumen del ejemplo

* La invarianza débil con respecto al género fue satisfecha
  + Se pueden comparar la varianza del rasgo latente entre grupos
* En el ejemplo, la depresión fue ligeramente más variable para las mujeres
  + Hombres: Est = 1.000; SE = NA
  + Mujeres: Est = 1.263; SE = 0.120
* Para probar la homogeneidad de la varianza, ver ejemplos de codificación

## Invarianza Fuerte / escalar: Recapitulación

* **Modelo de medición**: Todos los pesos factoriales y umbrales son iguales a través de los grupos
* **Modelo de población**: Suficiente para estimar el impacto en las (co-) varianzas y medias del factor
* Considerado aceptable para comparar grupos en puntuaciones de pruebas observadas
* Con datos categóricos, también se puede estimar la varianza de las VRL con invarianza fuerte
  + La mayoría de softwares lo harán por defecto

## Código de invarianza fuerte / escalar

# Fit model  
fit.strong <- cfa(mod1,   
 data = dat,   
 std.lv = T,   
 ordered = T,   
 group = "cfemale",  
 group.equal = c("loadings", "thresholds")) # <--- new   
   
# Print model summary  
summary(fit.strong)

## Salida de invarianza fuerte / escalar

|  |
| --- |

## Comparación de modelos

lavTestLRT(fit.config, fit.weak, fit.strong)

Scaled Chi-Squared Difference Test (method = "satorra.2000")  
  
lavaan NOTE:  
 The "Chisq" column contains standard test statistics, not the  
 robust test that should be reported per model. A robust difference  
 test is a function of two standard (not robust) statistics.  
   
 Df AIC BIC Chisq Chisq diff Df diff Pr(>Chisq)   
fit.config 40 37.534   
fit.weak 47 57.126 12.091 7 0.09761 .   
fit.strong 62 112.075 63.910 15 5.3e-08 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

. . .

* Rechazar la invarianza fuerte usando
* umbrales por restringidos, pero se estimó la media del rasgo latente y 8 parámetros , así que

## Resumen del ejemplo

* La invarianza fuerte con respecto al género no fue satisfecha
* En el ejemplo, la depresión fue mucho más alta en promedio para las mujeres
  + Hombres: Est = 0.000; SE = NA
  + Mujeres: Est = .450; SE = 0.086
* Diferencia de 0.45 DE entre grupos (DE = 1 para hombres)
* Pero no sabemos si esto se debe al sesgo de medición, impacto o ambos (porque el modelo fue rechazado)

# Próximos pasos

## Resumen

* Hemos visto cómo probar la IM utilizando el análisis factorial para datos categóricos
* En nuestro ejemplo, encontramos que la evaluación CINT satisfizo la invarianza métrica pero no la invarianza escalar
  + Implicación – las comparaciones entre géneros pueden reflejar sesgo de medición, impacto o ambos
* A continuación, consideraremos cómo encontrar ítems que exhiben FDI
  + ¡Eliminar estos ítems de la evaluación asegurará que las comparaciones de medias en la CINT sean imparciales y justas con respecto al género!

## Lo que hemos hecho hasta ahora

* IM y FDI en general
* Análisis factorial con datos categóricos
* Pruebas de IM utilizando análisis factorial
  + Configuracional, débil, fuerte
  + Ver Apéndice y ejemplo de codificación para invarianza estricta
* Ilustró métodos utilizando un ejemplo

## Lo que haremos a continuación

* Cambiar perspectivas a TRI
* TRI con datos binarios y categóricos
* Pruebas de FDI utilizando TRI
* Ilustrar métodos utilizando un ejemplo

## Referencias

Bauer, D. J. (2017). A more general model for testing measurement invariance and differential item functioning. Psychological Methods, 22(3), 507–526.

Curley, W. E., & Schmitt, A. P. (1993). Revising Sat®-Verbal Items to Eliminate Differential Item Functioning. ETS Research Report Series, 1993(2), i–18.

Holland, P. W. & Rosenbaum. P. R. (1986). Conditional Association and Unidimensionality in Monotone Latent Variable Models. The Annals of Statistics, 14(4), 1523–1543.

Millsap, R. E., & Yun-Tein, J. (2004). Assessing Factorial Invariance in Ordered-Categorical Measures. Multivariate Behavioral Research, 39(3), 479–515.

Muthen, B., & Asparouhov, T. (2002). Latent Variable Analysis With Categorical Outcomes: Multiple-Group And Growth Modeling In Mplus.

Satorra, A., & Bentler, P. (2001). A scaled difference chi-square test statistic for moment structure analysis. Psychometrika, 66, 507–514.

Wu, H., & Estabrook, R. (2016). Identification of Confirmatory Factor Analysis Models of Different Levels of Invariance for Ordered Categorical Outcomes. Psychometrika, 81(4), 1014–1045.

# Apéndice

## Invarianza estricta: Recapitulación

* **Modelo de medición**: Todos los pesos factoriales, umbrales y varianzas residuales son iguales a través de los grupos
* **Modelo de población**: Suficiente para estimar el impacto en las (co-) varianzas y medias del factor
* Asegura que las puntuaciones de las pruebas sean igualmente confiables en ambos grupos

## Invarianza estricta vs invarianza fuerte

* En el análisis factorial para datos continuos, rara vez se prueba la invarianza estricta
  + No es necesario para estimar el impacto o comparar grupos en puntuaciones observadas
* Con datos categóricos: ¿Deberían tratarse las varianzas ("Deltas") de las VRL como parámetros reales?
  + No lo creo; ver material suplementario para otras opiniones
* Si queremos ignorar la varianza de las VRL, entonces deberíamos usar invarianza estricta en lugar de IM.
* Más fácil de hacer con la parametrización Theta, requiere algo de código nuevo en lavaan…
* En nuestro ejemplo, no hará diferencia ya que ya rechazamos la invarianza fuerte

## Código de invarianza estricta

# Model syntax to constrain Delta = 1 in both group  
mod.strict <-   
 'depression =~ cint1 + cint2 + cint4 + cint11 +   
 cint27 + cint28 + cint29 + cint30  
   
 cint1 ~\*~ c(1, 1)\*cint1  
 cint2 ~\*~ c(1, 1)\*cint2  
 cint4 ~\*~ c(1, 1)\*cint4  
 cint11 ~\*~ c(1, 1)\*cint11  
 cint27 ~\*~ c(1, 1)\*cint27  
 cint28 ~\*~ c(1, 1)\*cint28  
 cint29 ~\*~ c(1, 1)\*cint29  
 cint30 ~\*~ c(1, 1)\*cint30'  
  
fit.strict <- cfa(mod.strict, # <-- new  
 data = dat,   
 std.lv = T,   
 ordered = T,   
 group = "cfemale",  
 group.equal = c("loadings", "thresholds"))   
  
summary(fit.strict)

## Código de invarianza estricta

lavTestLRT(fit.config, fit.weak, fit.strict)

Scaled Chi-Squared Difference Test (method = "satorra.2000")  
  
lavaan NOTE:  
 The "Chisq" column contains standard test statistics, not the  
 robust test that should be reported per model. A robust difference  
 test is a function of two standard (not robust) statistics.  
   
 Df AIC BIC Chisq Chisq diff Df diff Pr(>Chisq)   
fit.config 40 37.534   
fit.weak 47 57.126 12.091 7 0.09761 .   
fit.strict 70 121.781 73.542 23 3.411e-07 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

. . .

* Nota: umbrales restringidos, pero se estimó la media del rasgo latente, así que
* Creo que este es el correcto para esta comparación