



# ANÁLISIS DE SENDEROS (II)

Gabriel Sotomayor  
Mayo 2023

# CONTENIDOS

- Recordatorio de la sesión anterior: Conceptos básicos y supuestos del PA
- Pasos de aplicación del PA
  - Especificación
  - Identificación
  - Estimación de parámetros
  - Evaluación del ajuste
  - Re-especificación del modelo
  - Interpretación de resultados
- Aplicación de PA en R

# CALENDARIZACIÓN

Fechas	Contenidos	Evaluaciones
29 de mayo	Análisis de sendero	
5 de junio	Modelos de ecuaciones estructurales	
12 de junio	Modelos de ecuaciones estructurales	Tarea de Senderos
19 de junio	Modelos de ecuaciones estructurales	
26 de junio	Feriado	
3 de julio	Repaso	Trabajo Final



# 1. REPASO DE LA SESIÓN ANTERIOR

# CONCEPTOS CENTRALES

**Análisis de senderos:** es un método que permite evaluar el ajuste de modelos teóricos en los que se proponen un conjunto de relaciones de dependencia entre variables. Extensión de RLM.

- **Variables exógenas:** sus causas son externas al modelo, su función es explicar las otras variables internas del modelo.
- **Variables endógenas:** tienen sus causas en una o más variables del modelo, incluyen variables dependientes e intervinientes.
- **Efectos directos:** influencia inmediata de una variable sobre otra.
- **Efectos indirectos:** influencia mediada por una o más variables intermedias.
- **Efectos espurios:** relación entre dos variables endógenas es influenciada por una tercera variable no contemplada en el modelo.

# SUPUESTOS DEL PATH ANALYSIS (I)

- Path Analysis (PA) es una extensión del análisis de regresión múltiple y requiere el cumplimiento de sus supuestos junto con otros adicionales.
- **Exploración de datos:** Detectar valores extremos (outliers) y valores perdidos (missing) para evitar distorsiones en el análisis. Para los outliers, se pueden usar puntajes Z (rango  $\pm 3$ ) y la distancia de Mahalanobis ( $D^2$ ).
- **Manejo de outliers:** Recomendado removerlos o recodificarlos al puntaje extremo más próximo.
- **Valores perdidos:** Su impacto depende de la cantidad y el patrón.

# SUPUESTOS DEL PATH ANALYSIS (II)

- **Tamaño de la muestra:** Se recomienda entre 10 y 20 casos por parámetro y al menos 200 observaciones.
- **Independencia de errores:** El término de error de cada variable endógena no debe correlacionarse con otras variables.
- **Normalidad:** Los datos deben seguir una distribución normal. Se puede verificar la normalidad univariada y multivariada examinando los índices de asimetría y curtosis, y el índice multivariado de Mardia.

# SUPUESTOS DEL PATH ANALYSIS (III)

- **Linealidad y Multicolinealidad:** Los datos deben tener una relación lineal y las correlaciones bivariadas entre variables no deben ser demasiado altas (más de 0.85 indica posible multicolinealidad).
- **Recursividad:** Las influencias causales deben ser unidireccionales y sin efectos retroactivos.
- **Nivel de medición intervalar:** Se asume para la mayoría de las variables, aunque a veces se pueden usar variables nominales u ordinales.
- **Confiabilidad:** Los instrumentos de medición utilizados deben tener propiedades de confiabilidad al menos moderadas.





# PASOS DE APLICACIÓN DEL ANÁLISIS DE SENDEROS



# PASOS DEL PATH ANALYSIS

**Especificación:** Define las variables y relaciones en el modelo explicativo.

**Identificación:** Verifica si el modelo está correctamente identificado.

**Estimación de parámetros:** Estima los parámetros con base en las varianzas y covarianzas muestrales.

**Evaluación del ajuste:** Revisa si las relaciones del modelo reflejan adecuadamente las observadas en los datos.

**Re-especificación del modelo:** Mejora el ajuste del modelo si es necesario.

**Interpretación de resultados:** Extrae conclusiones significativas de los datos.

# ESPECIFICACIÓN

Determina las variables y su relación basada en el conocimiento teórico del fenómeno.

Evita errores de especificación interna y externa para desarrollar un modelo de alto valor explicativo y relevancia teórica.

Errores de especificación interna: por omisión de parámetros relevantes o inclusión de parámetros irrelevantes.

Errores de especificación externa: variables omitidas.

Ejemplo: Un modelo de rendimiento académico en Lengua debería incluir la variable de autoeficacia para la escritura.

# IDENTIFICACIÓN

- Antes de recoger los datos, asegura que el modelo esté correctamente identificado: determinar si se cuenta con la cantidad suficiente de información para contrastarse el modelo.
- Calcula los grados de libertad del modelo para determinar si se dispone de suficiente información.
- $gl = \frac{1}{2} \times (N^{\circ} \text{ de variables observadas} \times (N^{\circ} \text{ de variables observadas} + 1)) - N^{\circ} \text{ parámetros a estimar}$
- Diferentes estados del modelo: identificado ( $gl=0$ , ajuste perfecto), por lo que no resulta de interés, solo reproduce la matriz de correlaciones original, sub-identificado ( $gl<0$ , necesita más información), **sobre-identificado ( $gl>0$ , puede ser estimado y contrastado).**

# ESTIMACIÓN DE PARÁMETROS

Estima los valores de los parámetros que proporcionen un ajuste óptimo entre la matriz reproducida y la observada.

Si la matriz residual es próxima a cero (matriz observada - matriz reproducida = matriz residual) el ajuste es bueno.

Método más común: Máxima Verosimilitud (ML, Maximum Likelihood), que requiere supuestos como muestra de tamaño adecuado, medidas al menos de nivel intervalar, y distribución normal multivariada.

Este método es robusto a ligeras desviaciones de la distribución normal (valores de hasta 70 en el coeficiente de Mardia).

Otros métodos pueden ser utilizados en casos de ausencia de normalidad o variables categóricas: WLS, DWLS.

# EVALUACIÓN DEL AJUSTE

- Revisa si las relaciones en el modelo reflejan adecuadamente las relaciones observadas en los datos.
- Evaluar en términos de: (a) magnitud y significación de los parámetros estimados, (b) varianza explicada por las variables, y (c) ajuste del modelo a los datos.
- Uso de estadísticos de bondad de ajuste: ajuste absoluto (chi cuadrado, RMSEA), ajuste relativo (CFI, TLI), ajuste parsimonioso (NFI).

# EVALUACIÓN DEL AJUSTE

**Tabla 1.**

Estadísticos de bondad de ajuste y criterios de referencia.

Estadístico	Abreviatura	Criterio
<i>Ajuste absoluto</i>		
Chi-cuadrado	$\chi^2$	Significación > .05
Razón Chi-cuadrado / grados de libertad	$\chi^2/gl$	< 3
Índice de bondad de ajuste	GFI	$\geq .95$
Índice de bondad de ajuste corregido	AGFI	$\geq .95$
Raíz del residuo cuadrático medio	RMR	Próximo a 0
Raíz cuadrada media del error de aproximación	RMSEA	< 0,05
<i>Ajuste comparativo</i>		
Índice de ajuste comparativo	CFI	$\geq .95$
Índice de Tucker-Lewis	TLI	$\geq .95$
Índice de ajuste normalizado	NFI	$\geq .95$
<i>Ajuste parsimonioso</i>		
Corregido por parsimonia	PNFI	Próximo a 1

# RE-ESPECIFICACIÓN DEL MODELO

Si el ajuste no es óptimo, el modelo puede ser re-especificado.

Las decisiones de añadir o eliminar parámetros deben estar en línea con la teoría subyacente al modelo propuesto.

Índices de modificación y análisis de residuos pueden ser utilizados para orientar las modificaciones del modelo.

El valor del índice de modificación corresponde aproximadamente a la reducción en el  $X^2$  que se produciría si el coeficiente fuera estimado.

Un valor  $> 3.84$  sugiere que se obtiene una reducción estadísticamente significativa en el  $X^2$  cuando se estima el coeficiente.

La existencia de residuos elevados entre parejas de variables ( $> 2.58$ ) señalaría la necesidad de introducir parámetros adicionales susceptibles de explicar la relación entre las variables en cuestión.



# INTERPRETACIÓN DE RESULTADOS

- Extrae conclusiones significativas basadas en los datos y el ajuste del modelo.
- Proporciona una visión general del fenómeno bajo estudio basado en los resultados del análisis de senderos.



**APLICACIÓN EN R**

