



INTRODUCCIÓN AL ANÁLISIS DE SENDEROS

Gabriel Sotomayor
Mayo 2023

CONTENIDOS

Calendarización

Retroalimentación Prueba 1

Análisis de Senderos

1. Introducción
2. Conceptos Básicos
3. Descomposición de los efectos path
4. Supuestos del Path Analysis

CALENDARIZACIÓN

Fechas	Contenidos	Evaluaciones
22 de mayo	Análisis de sendero	Tarea de AFC
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

ÍTEM 1 — REGRESIÓN LINEAL MÚLTIPLE

- a) Tipo de variable dependiente en RLM: Intervalo o de razón. Numérica es parcialmente correcta.
- b) Mínimos cuadrados ordinarios: Método de estimación del a Regresión. Busca minimizar la suma del cuadrado de los residuos.
- c) Ejemplos: Poca problematización sociológica. Exceso de variables demográficas. No se hace referencia específica a la interpretación de coeficientes ni estadísticos de ajuste.
- d) Problemas con la interpretación de coeficientes:

Escolaridad: El coeficiente para la variable escolaridad (87210.88) indica que, manteniendo constantes las otras variables del modelo, por cada año adicional de escolaridad, se espera un aumento promedio en los ingresos autónomos de \$87210.88.

ÍTEM 2 — REGRESIÓN LOGÍSTICA

- a) Diferencia entre RLM y RLG: Nivel de medida de la variable dependiente. Transformación logarítmica de los odds. Se modelan las chances (logaritmo).
- b) Ejemplo: Buena selección de variables. En general hay buen manejo de la intuición tras la técnica, pero problemas en las tecnicidades específicas. No es necesario ajustar dos modelos.
- c) Bien en la comparación de modelos, problemas en la interpretación de coeficientes. Ej:

La variable “Mujer” (ref.hombre) tiene un coeficiente de 0.39 (0.03)***, lo que indica que los hogares con jefatura femenina tienen mayores probabilidades de encontrarse en situación de pobreza que los con jefatura masculina. En concreto, las odds (razón de probabilidades) de pobreza para una hogar con jefatura femenina son $\exp(0.39) = 1.48$ veces mayores que las de uno con jefatura masculina, manteniendo constantes el resto de variables.

La variable “Edad” tiene un coeficiente de -0.03 (0.00)***, lo que indica que a medida que aumenta la edad del jefe de hogar, disminuyen las probabilidades de que el hogar se encuentre en situación de pobreza. En concreto, las odds de pobreza disminuyen en un factor de $\exp(-0.03) = 0.97$, es decir en un 3%, por cada año de aumento en la edad del jefe de hogar, manteniendo constantes las demás variables.

ÍTEM 3 — ANÁLISIS FACTORIAL EXPLORATORIO

a) Supuestos

- Nivel de medición: En principio se espera que sean continuas. Se requiere al menos ordinalidad.
- Tamaño muestral: Cómo mínimo se esperan unos 200 o 300 casos en total
- Normalidad (univariada y multivariada): Todas las variables observadas y sus combinaciones lineales han de estar distribuidas normalmente.
- Multicolinealidad: Para la extracción de factores comunes debe existir varianza común entre las variables, de lo contrario es poco probable encontrar estructuras latentes relevantes.

b) Ejemplo: Poco prolijo en la especificación de las variables y los factores comunes que se podrían obtener.

c) Análisis: Poca consideración del ajuste general. Falta de claridad respecto a los términos (autovalores, varianza acumulada, estructura simple).



1. ANÁLISIS DE SENDEROS (PATH ANALYSIS)

INTRODUCCIÓN AL ANÁLISIS DE SENDEROS

Análisis de Senderos: método que evalúa el ajuste de modelos teóricos con relaciones de dependencia entre variables.

No prueba la causalidad sino que ayuda a seleccionar o inferir hipótesis causales.

Extensión del modelo de regresión múltiple, examina la contribución directa e indirecta de variables.

Orígenes en estudios filogenéticos, introducido en ciencias sociales en el siglo XX.

Uso creciente en sociología, psicología, economía, ciencias políticas, ecología y otras disciplinas.

EVOLUCIÓN Y APLICACIÓN DEL ANÁLISIS DE SENDEROS

- Surgimiento de programas informáticos para Modelos de Ecuaciones Estructurales (SEM) en 1980.
- Diferencia entre PA y SEM: SEM mide variables latentes usando múltiples medidas y modela el error de medición.
- Aunque las variables observables pueden medirse directamente, no son reflejos exactos de la variable, hay factores aleatorios e imprevisibles.
- Ventajas del SEM: estima el efecto adicional del error de medición, permite establecer la validez de constructo de las variables latentes.
- A pesar de las ventajas del SEM, el PA sigue siendo útil y muy utilizado en la investigación psicológica.

FUNCIONAMIENTO Y PROPÓSITOS DEL ANÁLISIS DE SENDEROS

- El investigador realiza regresiones para analizar las relaciones entre variables, las cuales pueden operar como variables independientes de otras variables en el modelo.
- El PA evalúa el ajuste del modelo: el grado en que el modelo propuesto representa las relaciones entre las variables bajo estudio.
- El PA permite detectar modelos poco ajustados a la realidad y provee estimaciones de la magnitud y la significación de las relaciones hipotetizadas entre un conjunto de variables.
- Se puede representar el modelo mediante la creación de un diagrama con flechas que conectan las variables en estudio, estimando coeficientes path que son análogos a los coeficientes beta del análisis de regresión múltiple.



2. CONCEPTOS BÁSICOS

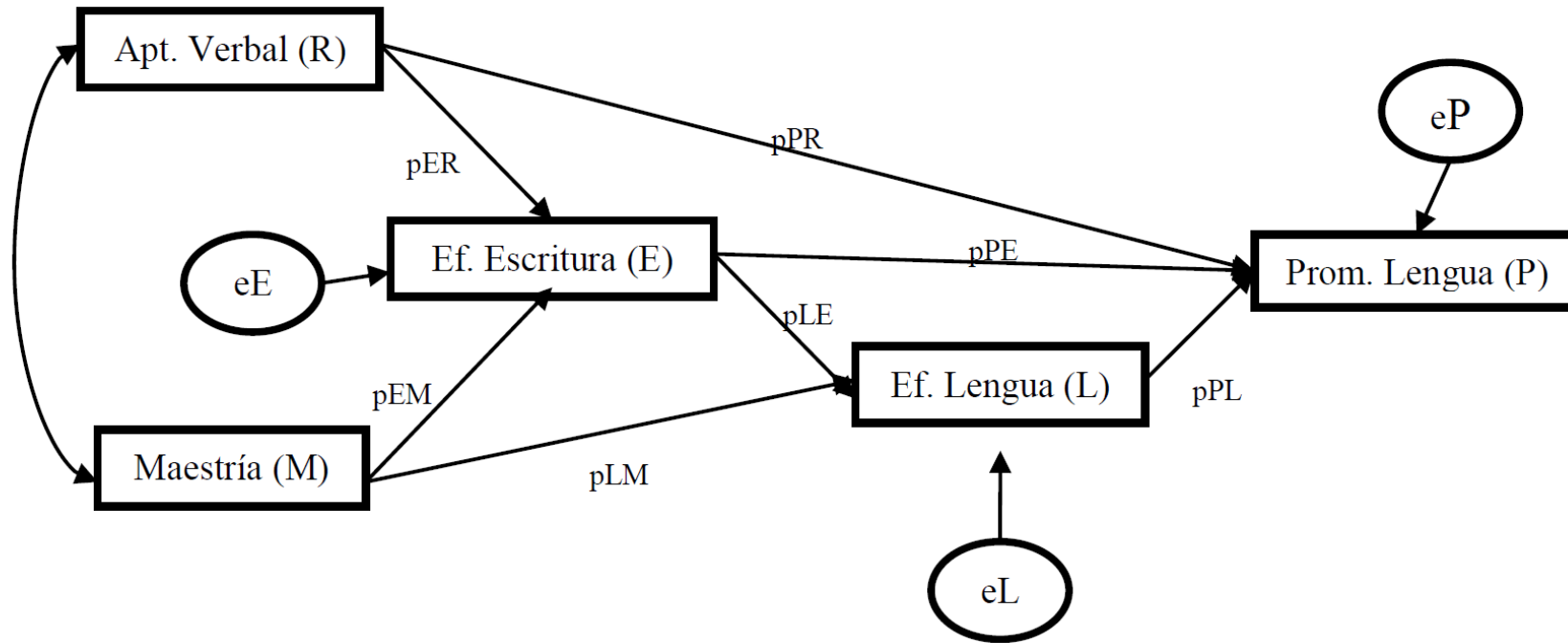


Figura 1. Un modelo social-cognitivo del rendimiento académico en Lengua. Nota: Tot. Verb=Aptitud Verbal; Efwriti=autoeficacia para la escritura; Maestría=Estructura de maestría de las metas de clase; Eflengu=Autoeficacia para rendimiento en Lengua; Promed=Promedio de calificaciones en Lengua; e=error de medición.

CONVENCIONES EN EL ANÁLISIS DE SENDEROS

- Diagramas son comunes para representar modelos hipotéticos en PA.
- Convenciones a tener en cuenta:
 - Flechas indican la relación entre variables; su sentido indica la dirección de la relación.
 - Flechas bidireccionales indican la covariación entre variables sin dirección especificada.
 - Cada flecha tiene un coeficiente path que indica la magnitud del efecto de la relación entre las variables.
 - Variables que reciben influencia de otras se denominan endógenas, las que no reciben influencia son exógenas.
 - Variables observables se enmarcan en cuadrados, variables latentes en círculos.
 - Los efectos directos operan directamente de una variable a otra.
 - Los efectos indirectos ocurren cuando la relación entre dos variables es mediada por una o más variables.

VARIABLES Y COEFICIENTES EN EL ANÁLISIS DE SENDEROS

- 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.
- Términos de error o residuales: variables exógenas no medidas directamente, reflejan causas inespecíficas de variabilidad en la variable dependiente o varianza no explicada más cualquier error debido a la medición.
- Coeficientes path: indican la magnitud y el signo del efecto de una variable sobre otra variable endógena, representan el efecto de una variable sobre otra, controlando el resto de las variables.

APLICACIÓN PRÁCTICA DEL ANÁLISIS DE SENDEROS

- Modelo del rendimiento académico en Lengua, formulado por Pérez, Medrano y Ayllón (2010).
- Variables consideradas: aptitud cognitiva verbal, creencias de autoeficacia para la escritura y rendimiento en Lengua, estructura de metas de aula de maestría.
- El modelo propone relaciones directas, indirectas y correlaciones entre estas variables.
- El diagrama de este modelo representa las relaciones causales hipotetizadas.

APLICACIÓN PRÁCTICA DEL ANÁLISIS DE SENDEROS

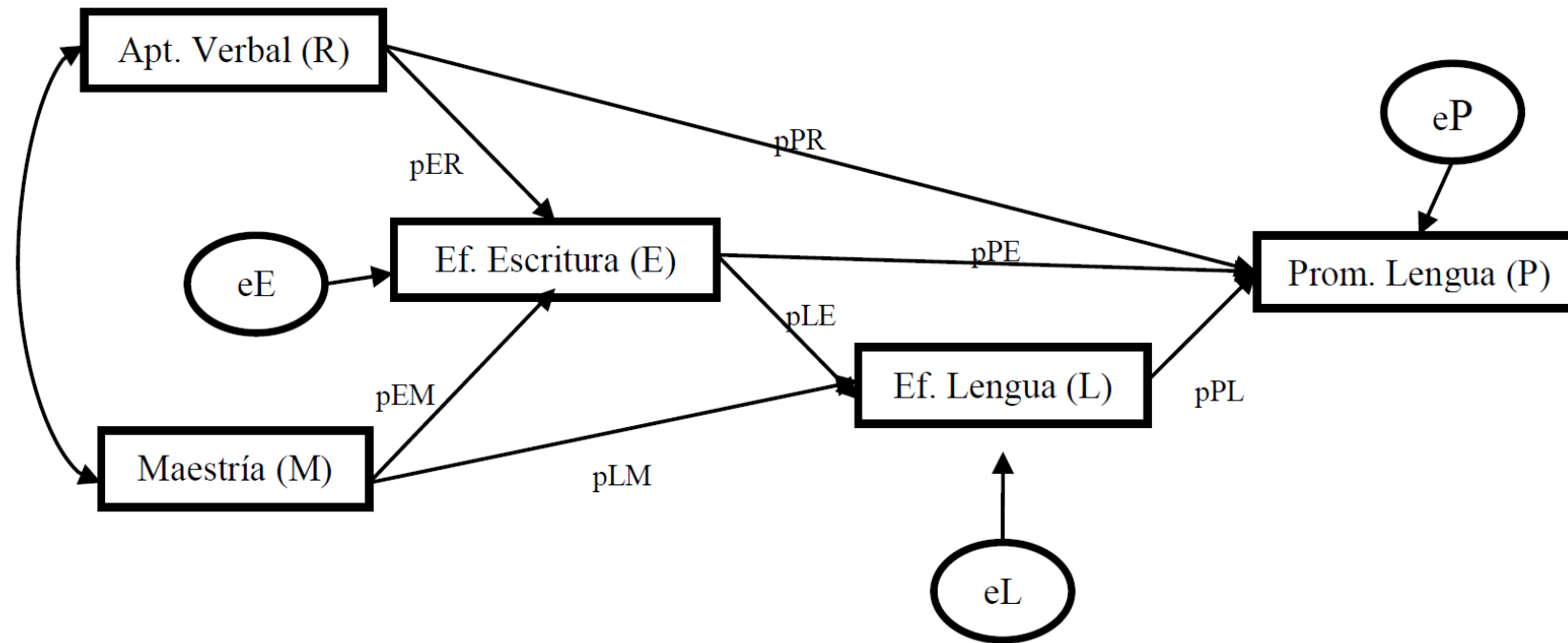


Figura 1. Un modelo social-cognitivo del rendimiento académico en Lengua. Nota: Tot. Verb=Aptitud Verbal; Efwriti=autoeficacia para la escritura; Maestría=Estructura de maestría de las metas de clase; Eflengu=Autoeficacia para rendimiento en Lengua; Promed=Promedio de calificaciones en Lengua; e=error de medición.

APLICACIÓN DE ECUACIONES EN EL ANÁLISIS DE SENDEROS

- PA es una extensión del análisis de regresión múltiple y sigue sus supuestos: todas las relaciones entre las variables son lineales, aditivas y causales.
- El modelo puede especificarse mediante un conjunto de ecuaciones estructurales que describen las relaciones directas entre las variables.
- Ejemplo:
 - $\text{Prom. Lengua} = p_{PR} + p_{PE} + p_{PL} + e_P$
 - $\text{Ef. Lengua} = p_{LE} + p_{LM} + e_L$
 - $\text{Ef. Escritura} = p_{ER} + p_{EM} + e_E$
- Cada variable endógena tiene un término de error o path residual que representa la variación no explicada por las variables predictoras.



3. DESCOMPOSICIÓN DE LOS EFECTOS PATH



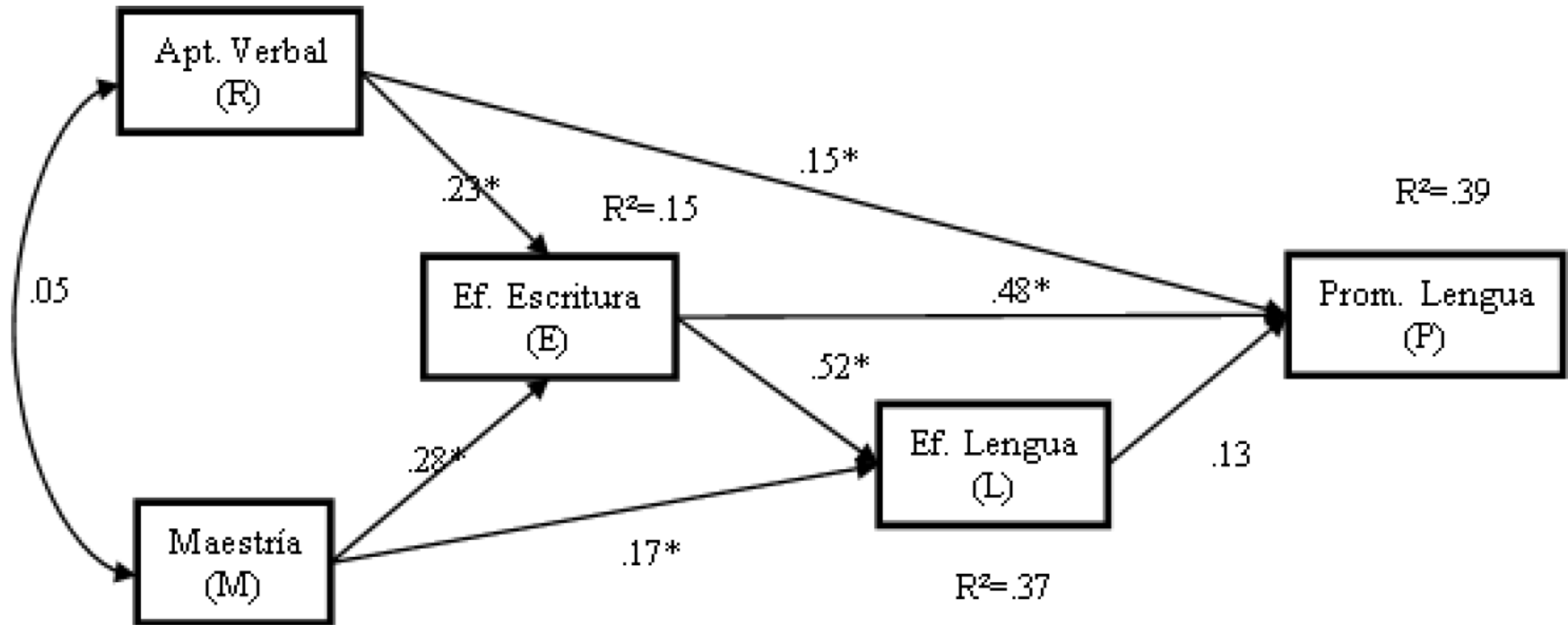
DESCOMPOSICIÓN DE EFECTOS PATH EN EL ANÁLISIS DE SENDEROS

- Una contribución específica del PA es su capacidad para descomponer asociaciones entre variables en efectos directos, indirectos y espurios.
- Efectos directos: influencia inmediata de una variable sobre otra.
- Efectos indirectos: influencia mediada por una o más variables intermedias.
Ejemplo: Ef. Escritura afecta al Prom. Lengua indirectamente a través de Ef. Lengua.
- Efectos espurios: relación entre dos variables endógenas es influenciada por una tercera variable no contemplada en el modelo. Ejemplo: relación entre Ef. Escritura y Ef. Lengua influenciada por Maestría.

ESTIMACIÓN Y SIGNIFICACIÓN DE LOS EFECTOS EN EL ANÁLISIS DE SENDEROS

- Los efectos de una variable sobre otra se estiman mediante coeficientes path estandarizados.
- La magnitud de los efectos indirectos se estima al multiplicar los coeficientes path a lo largo de la línea causal entre dos variables.
- Ejemplo: Efecto indirecto de Ef. Escritura sobre Prom. Lengua = $(p_{LE} \times p_{PL}) = (.52 \times .13) = .07$.
- Efecto total = efectos directos + efectos indirectos indirectos.
- Significación estadística de los efectos se calcula dividiendo los coeficientes no estandarizados por el error estándar, obteniendo un valor z.
- Valores z superiores a $\pm 1,96$ indican un efecto significativo a un nivel $p < 0,05$ (Test de Radio Crítico).
- Estos datos suelen ser proporcionados por programas estadísticos como R (paquete Lavaan).

ESTIMACIÓN Y SIGNIFICACIÓN DE LOS EFECTOS EN EL ANÁLISIS DE SEMEDOS





4. SUPUESTOS DEL ANÁLISIS DE SENDEROS

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 ± 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.