



ANÁLISIS DE SENDEROS (III)

Gabriel Sotomayor
Junio 2023

CONTENIDOS

Efectos indirectos (medicación) en sociología.

Construcción de diagramas de senderos en lavaan

Interpretación de coeficientes.

Estadísticos de ajustes.

Significación estadística en regresiones y análisis de senderos.

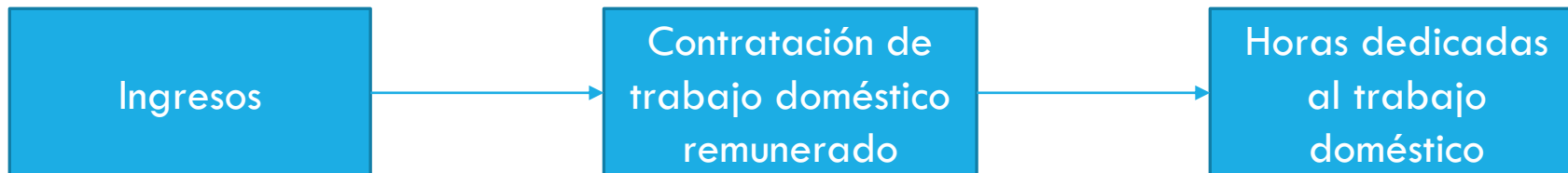
EFFECTOS INDIRECTOS

El análisis de senderos es útil para la investigación sociológica porque nos permite analizar relaciones más complejas entre variables que las que permite una regresión lineal múltiple.

- **Efectos directos:** influencia inmediata de una variable sobre otra.



- **Efectos indirectos:** influencia mediada por una o más variables intermedias.



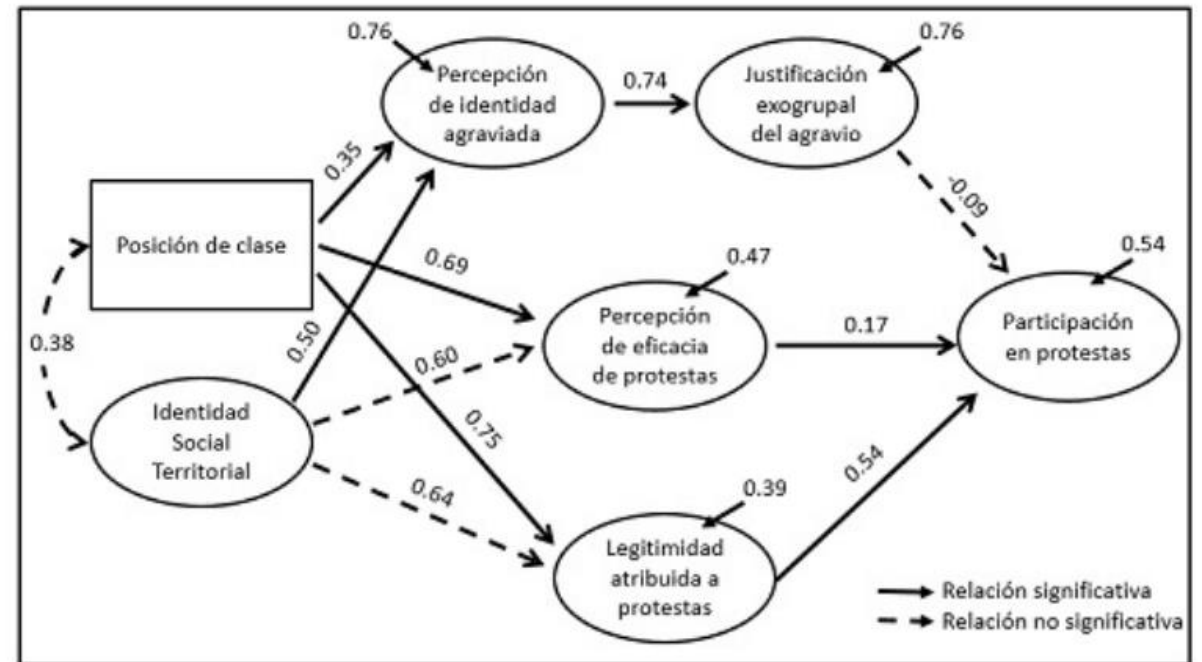
ACTIVIDAD 1: TRANSFORMEN LOS SIGUIENTES PLANTEAMIENTOS EN DIAGRAMAS

1. La esperanza de vida es afectada por el consumo de drogas. La posibilidad de consumir drogas también está influenciada por el ingreso, el cual a su vez también afecta la esperanza de vida.
2. La calidad del ambiente en el hogar afecta el desarrollo cognitivo infantil, el cual a su vez afecta el rendimiento escolar. La calidad del ambiente del hogar también afecta el desarrollo emocional infantil, el cual a su vez afecta la autoeficacia, lo cual afecta a su vez del rendimiento escolar. El ingreso afecta el rendimiento escolar, así como también la calidad del ambiente en el hogar.
3. La clase social afecta la participación en protestas de manera indirecta a partir de 3 vías. Por un lado, afecta la “Percepción identidad agraviada”, que a su vez afecta la “Justificación exogrupal del agravio” (que este es causado por otro grupo externo), que a su vez afecta la participación en protestas. Por otro lado la posición de clase afecta la percepción de eficacia de las protestas que afecta la participación en estas. Por último la posición de clase afecta la legitimidad atribuida a las protestas, que a su vez afecta la participación en protestas.

RECUPERANDO LA CLASE SOCIAL EN EL ESTUDIO DE LA ACCIÓN COLECTIVA. EL CASO DE UN MOVIMIENTO SOCIOAMBIENTAL CHILENO

Respecto a la posición de clase, esta sí tiene un efecto indirecto total significativo ($p < 0.001$) y fuerte ($r^2 = 0.29$) sobre la participación en protestas

Figura 4.
Resultados del modelo.



ACTIVIDAD 2:EFECTOS INDIRECTOS

En grupo de 2 o 3 personas, piensen en al menos 3 efectos indirectos que podríamos medir en sociología. Redacten las hipótesis propuestas (una por relación).

Para esto deben al menos contar con una variable independiente, una interviniente y una dependiente.



También puede haber otras formas:



CONSTRUCCIÓN DE DIAGRAMAS DE SENDEROS EN LAVAAN

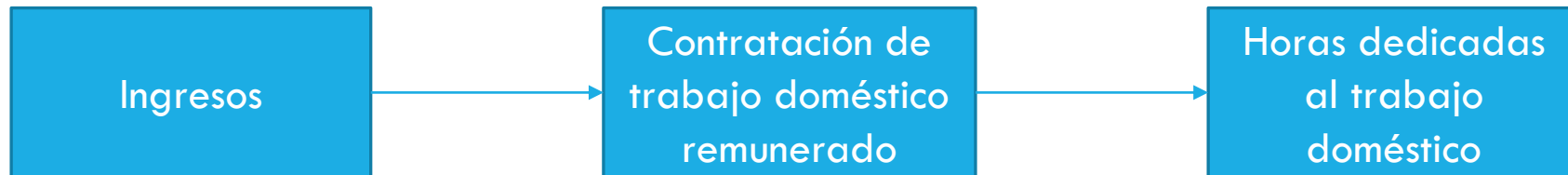
Para evaluar los modelos de análisis factorial confirmatorio y de senderos (y ecuaciones estructurales) en R, necesitamos expresarlos mediante el lenguaje de fórmulas del paquete lavaan.

Sintaxis	Comando	Ejemplo
~	Regresar en	Regresar B sobre A: $B \sim A$
~~	(Co)varianza	Varianza de A: $A \sim\sim A$
=~	Definir variable latente	Definir Factor 1 por A-D: $F1 =\sim A + B + C + D$
:=	Definir parámetro fuera del modelo	Definir parámetro u2 como doble del cuadrado de u: $u2 := 2*(u^2)$
*	Etiquetar parámetros (etiqueta antes de símbolo)	Etiquetar la regresión de Z sobre X como b: $Z \sim b*X$

EJEMPLO

En este caso, siendo ingresos “ing”, contratación de trabajo doméstico “ctd” y horas dedicadas al trabajo domestico “htd”, la formula quedaría definida de la siguiente manera:

```
mod_sendero <- 'ctd ~ ing  
                htd ~ ctd'
```



Ahora practiquemos con algunos ejemplos.

Como criterio general, cada variable endógena (con flechas apuntándole), será una línea de la especificación del modelo.

EJERCICIO 3

- Escribamos los diagramas confeccionados en el ejercicio 1 y ejercicio 2 a partir de la sintaxis de formulas de lavaan.

Sintaxis	Comando	Ejemplo
~	Regresar en	Regresar B sobre A: $B \sim A$
~~	(Co)varianza	Varianza de A: $A \sim\sim A$
=~	Definir variable latente	Definir Factor 1 por A-D: $F1 =\sim A + B + C + D$
:=	Definir parámetro fuera del modelo	Definir parámetro u2 como doble del cuadrado de u: $u2 := 2*(u^2)$
*	Etiquetar parámetros (etiqueta antes de símbolo)	Etiquetar la regresión de Z sobre X como b: $Z \sim b*X$

INTERPRETACIÓN DE COEFICIENTES PATH

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. Corresponden a coeficientes estandarizados.

castigo_media ~ rwa_media (0.284, $p < 0.001$):

Interpretación para una variable intervalar: por cada desviación estándar que aumenta en el "autoritarismo de derechas" (rwa_media), se espera que el "castigo severo" (castigo_media) aumente en promedio en 0.284 desviaciones estándar, controlando por las demás variables del modelo.

rwa_media ~ izquierda (-0.35, $p < 0.001$):

Interpretación para una variable nominal: En promedio, ser de izquierdas está asociado con una disminución de 0.35 desviaciones estándar en el "autoritarismo de derechas" (rwa_media), comparado con ser independiente, controlando por las demás variables del modelo.

INTERPRETACIÓN DE COEFICIENTES PATH

Componentes de la interpretación:

- Tamaño.
- Dirección.
- Control estadístico.
- Efecto promedio (o predicho por el modelo).
- **Significación estadística.**

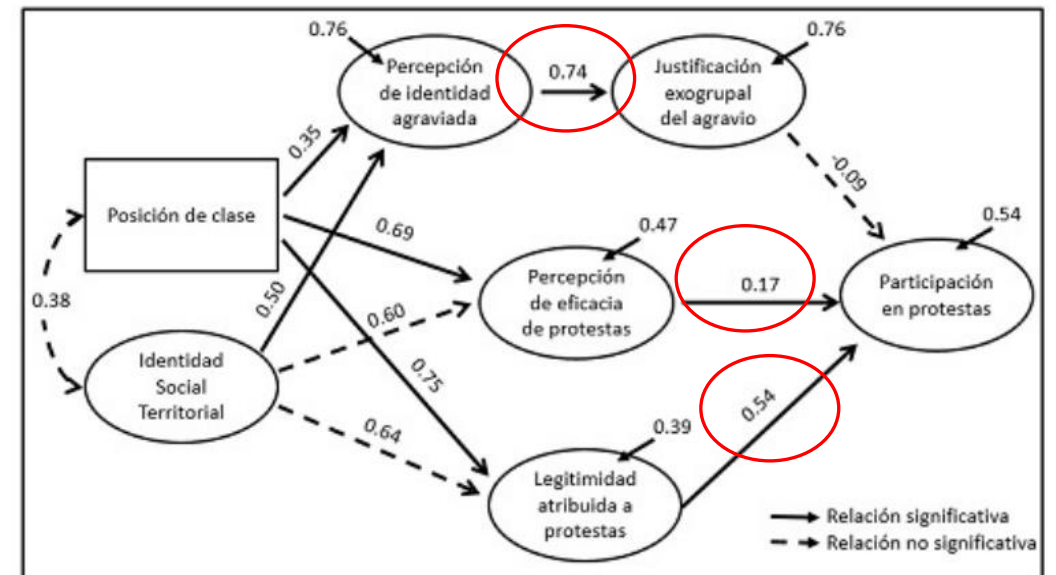
EJERCICIO 4

Escriban (escriban!!) la interpretación de los siguientes coeficientes path, considerando todos los elementos señalados.

- Tamaño.
- Dirección.
- Control estadístico.
- Efecto promedio (o predicho por el modelo).

Interpretación para una variable intervalar: por cada desviación estándar que aumenta en el "autoritarismo de derechas" (rwa_media), se espera que el "castigo severo" (castigo_media) aumente en promedio en 0.284 desviaciones estándar, controlando por las demás variables del modelo.

Figura 4.
Resultados del modelo.



¿POR QUÉ USAMOS MUESTRAS?

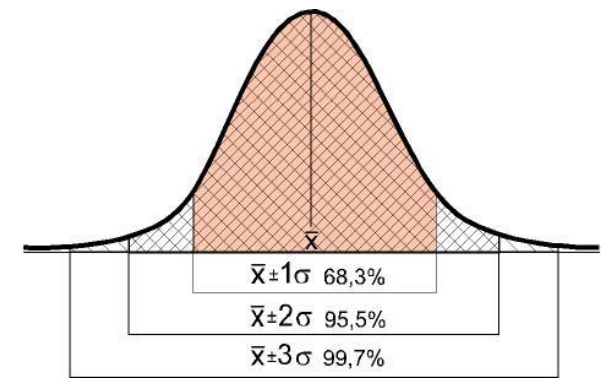
Usamos muestras porque nos permiten obtener datos **representativos** de una población a partir de una selección aleatoria de casos, una muestra. Este proceso se llama **estimación**: la estimación de un parámetro poblacional a partir de un estadístico muestral.

¿Porqué es esto posible?

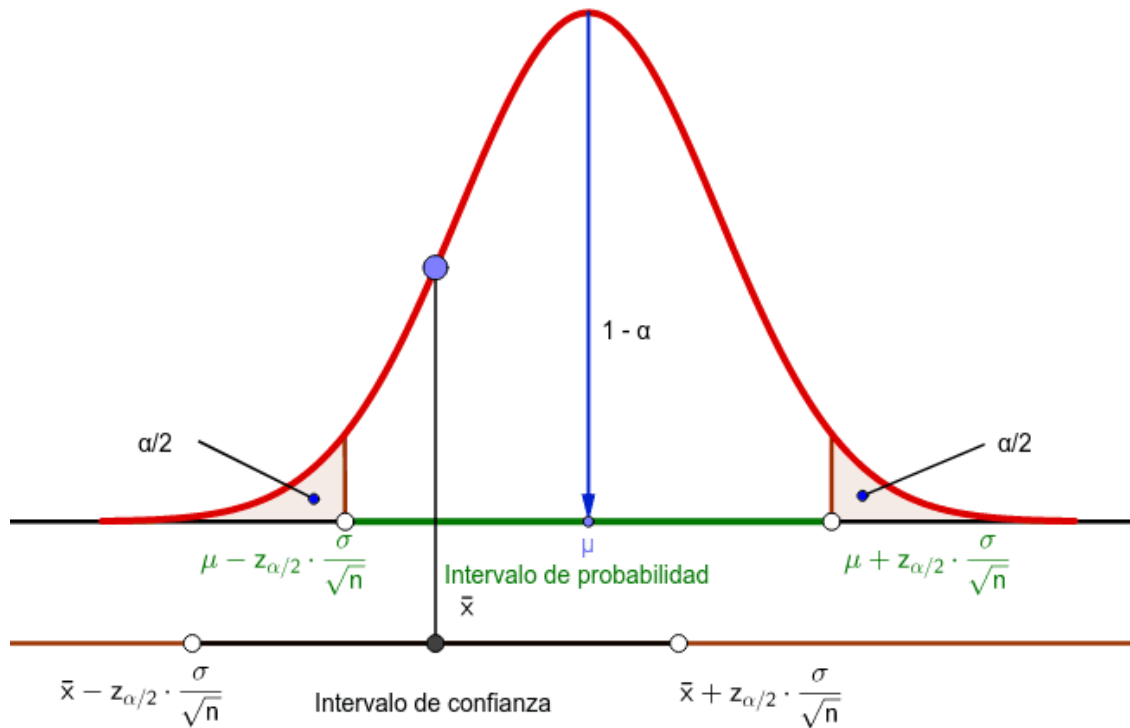
Sabemos que la representatividad es posible, gracias a dos leyes estadísticas esenciales:

- **El teorema del límite central** -> La distribución de medias muestrales extraídas de forma aleatoria de una población, se aproxima a la distribución normal a medida que aumenta el tamaño de la muestra.
- **La ley de los grandes números** -> la diferencia entre el estadístico muestral y el parámetro poblacional tiende a 0 cuando el tamaño de la muestra tiende al infinito.

DISTRIBUCIONES MUESTRALES



A partir del Teorema del Límite central, vimos que la media de las distintas muestras que podemos extraer de una población se distribuye de forma normal, con muestras suficientemente grandes (>50).



$$N(\mu; \frac{\sigma}{\sqrt{n}})$$

INTERVALOS DE CONFIANZA

Estimación de punto: Puede ser útil cuando solo tenemos la media (o proporción) muestral y no contamos con más información, sin embargo no sabemos la precisión (el error) de nuestra medición.

Estimación de Intervalo: Podemos calcularlo a partir de las propiedades de la distribución muestral de las medias, usando los estadísticos muestrales como estimadores de los parámetros poblacionales para calcular el error estándar.

Medias

$$IC = \bar{x} \pm Z_{1-\alpha/2} * \frac{s}{\sqrt{n}}$$

Proporciones

$$IC = \hat{p} \pm Z_{1-\alpha/2} * \sqrt{\frac{p(1-p)}{n}}$$

Error estándar

INFERENCIA EN ANÁLISIS DE SENDEROS

Al interpretar nuestros resultados queremos saber si contamos con evidencia suficiente para señalar las relaciones que observamos en nuestra muestra son estadísticamente significativas, es decir, si es probable que estas existan en la población.

Para esto debemos comparar el coeficiente path con su error estándar.

```
## Regressions:
##           Estimate Std.Err  z-value  P(>|z|)  Std.lv  Std.all
##  castigo_media ~
##    rwa_media      0.284   0.018   15.941   0.000   0.284   0.283
##  rwa_media ~
##    derecha        0.182   0.047    3.856   0.000   0.182   0.074
##    izquierda     -0.404   0.042   -9.679   0.000  -0.404  -0.187
##    centro        -0.094   0.033   -2.833   0.005  -0.094  -0.056
```


EVALUACIÓN DE AJUSTE

Tabla 1.

Estadísticos de bondad de ajuste y criterios de referencia.

Estadístico	Abreviatura	Criterio
<i>Ajuste absoluto</i>		
Chi-cuadrado	χ^2	Significación > .05
Razón Chi-cuadrado / grados de libertad	χ^2/df	< 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