Epidemiología Moderna: Conceptos y Modelos avanzados

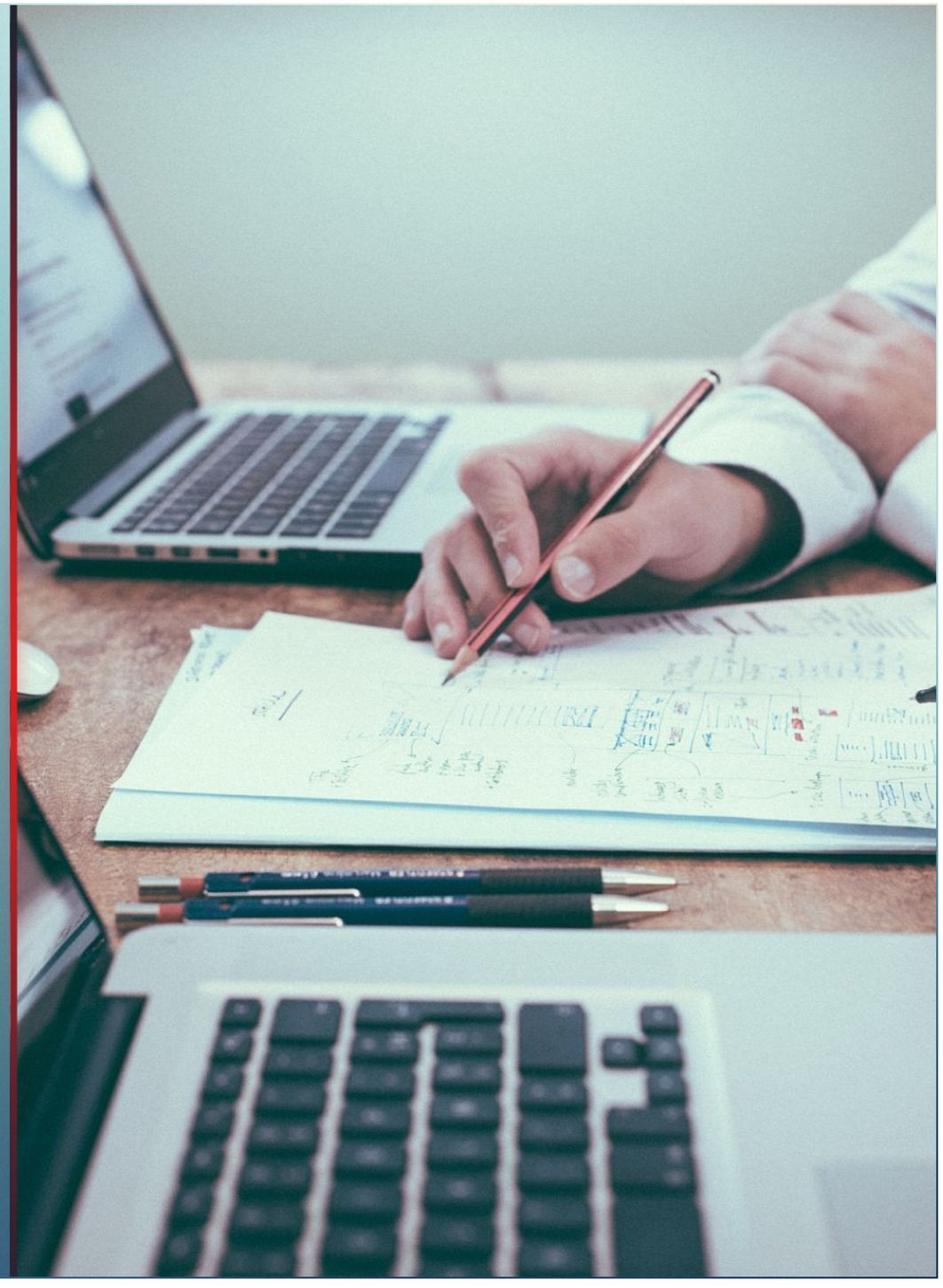
Introducción al modelado de ecuaciones estructurales - Conceptos básicos

Epidemiología para la Salud Pública

¿Recordemos?

Antes de comenzar la presentación de los conceptos de modelado de ecuaciones estructurales (MEE) es importante volver a los conceptos de regresión múltiple **(el concepto debe aparecer en un globo: el análisis de regresión es una metodología estadística multivariada que mide el cambio promedio que ocurre en una variable dependiente que está asociado con cambios en una o más variables independientes)**. En la regresión múltiple, la variable dependiente está asociada con más de una variable independiente.

El modelo más recurrente para modelar en estudios epidemiológicos es el uso de regresión múltiple (colocar un enlace con el contenido de la unidad 5 – tabla 2), a pesar de sus ventajas, los análisis que usan regresión múltiple tienen limitaciones. Una de las limitaciones en la regresión múltiple es que las variables se insertan en el modelo al mismo nivel. Para mitigar esta limitación, algunos académicos aplican modelos jerárquicos, que establecen niveles para las variables.

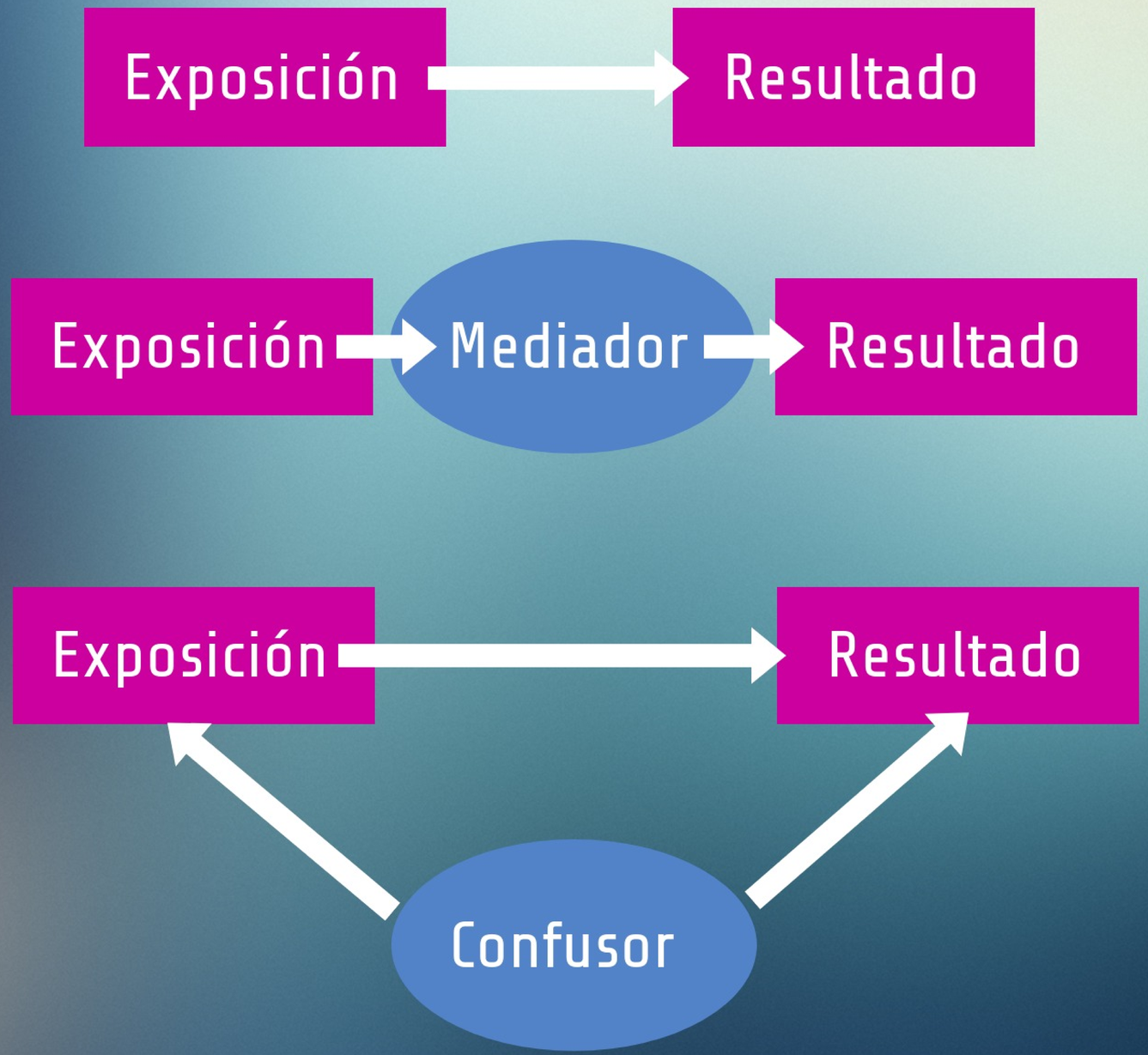


Modelos jerárquicos–
presuponen incluir simultáneamente las variables en el modelo y seleccionar el "mejor" modelo en base a conocimientos previos o criterios estadísticos. También debe incluir las variables basadas en un modelo conceptual de análisis de las interrelaciones. Se basa en un modelo teórico conceptual de asociaciones, que requiere un mayor conocimiento de los determinantes socioeconómicos, biológicos y conductuales para dilucidar las relaciones que se establecen entre ellos para determinar el efecto.

Figura 1: Esquema propuesto para el modelado jerárquico

Sin embargo, un nivel jerárquico es inapropiado para investigar relaciones simultáneas y efectos independientes entre variables y un resultado en situaciones donde una de estas variables es un modificador de efecto o factor de confusión.

Como ya hemos visto en la unidad 5, el método para corregir este fenómeno es procesar dos modelos: uno ajustado por el factor de confusión y otro sin el ajuste adecuado y, finalmente, comparar la estimación del efecto en los dos.
Para comprender la importancia de aplicar el análisis cuando hay modificadores de efectos y factores de confusión, recuerde:



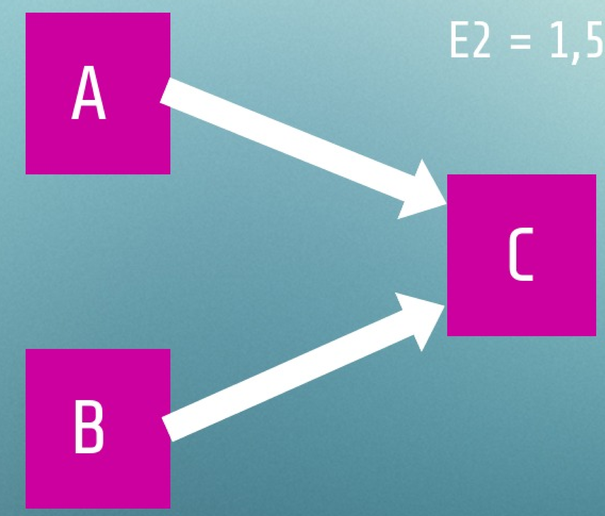
Efecto de A sobre C:

E1 = 2

Mediador – diferencia entre E1 y 2

Parte del efecto de A sobre C que está mediado (explicado por B)

E2 = 1,5



Confusor: diferencia entre E1 y E2

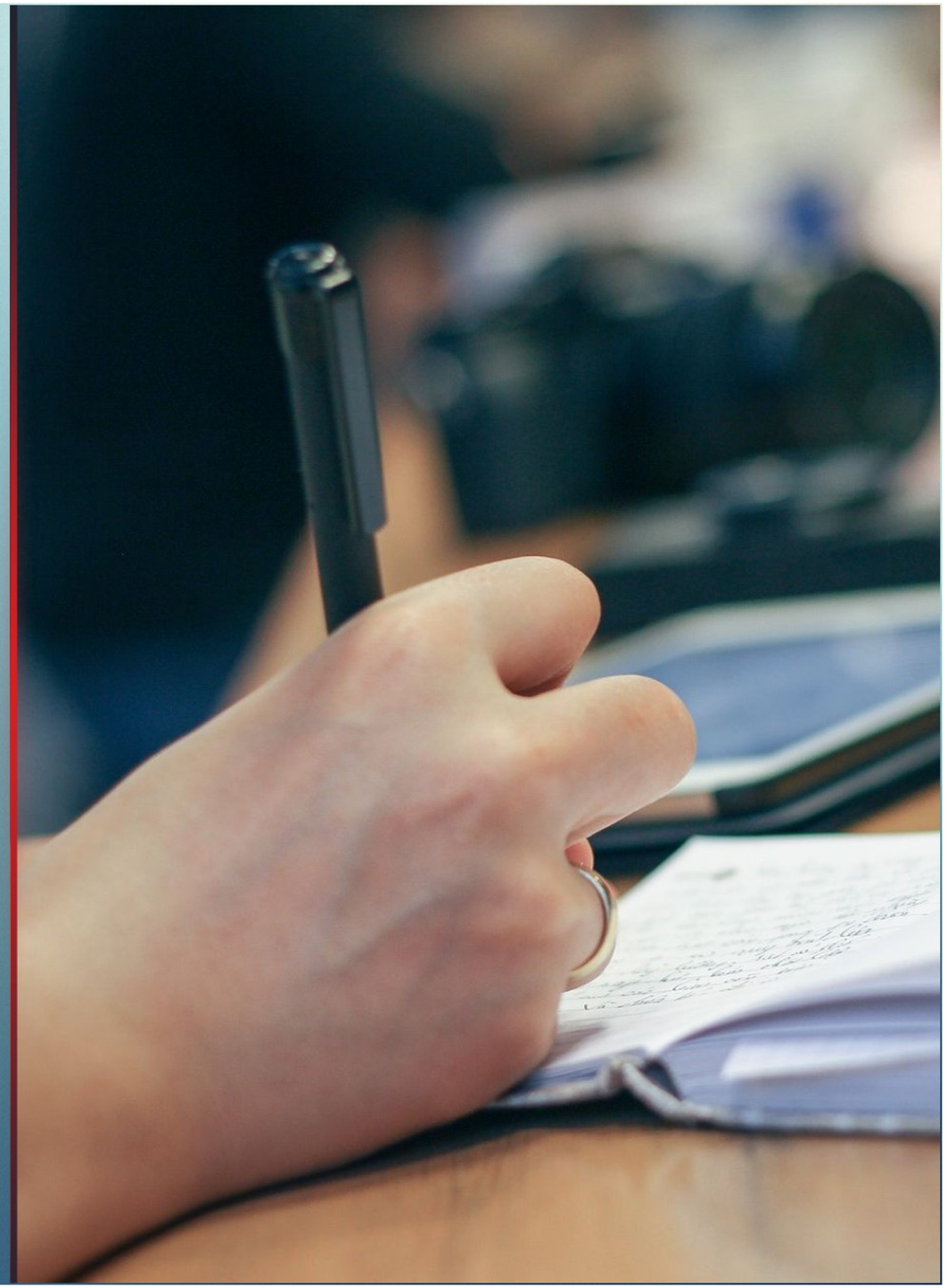
Diferencia en el efecto explicada por el efecto de confusión

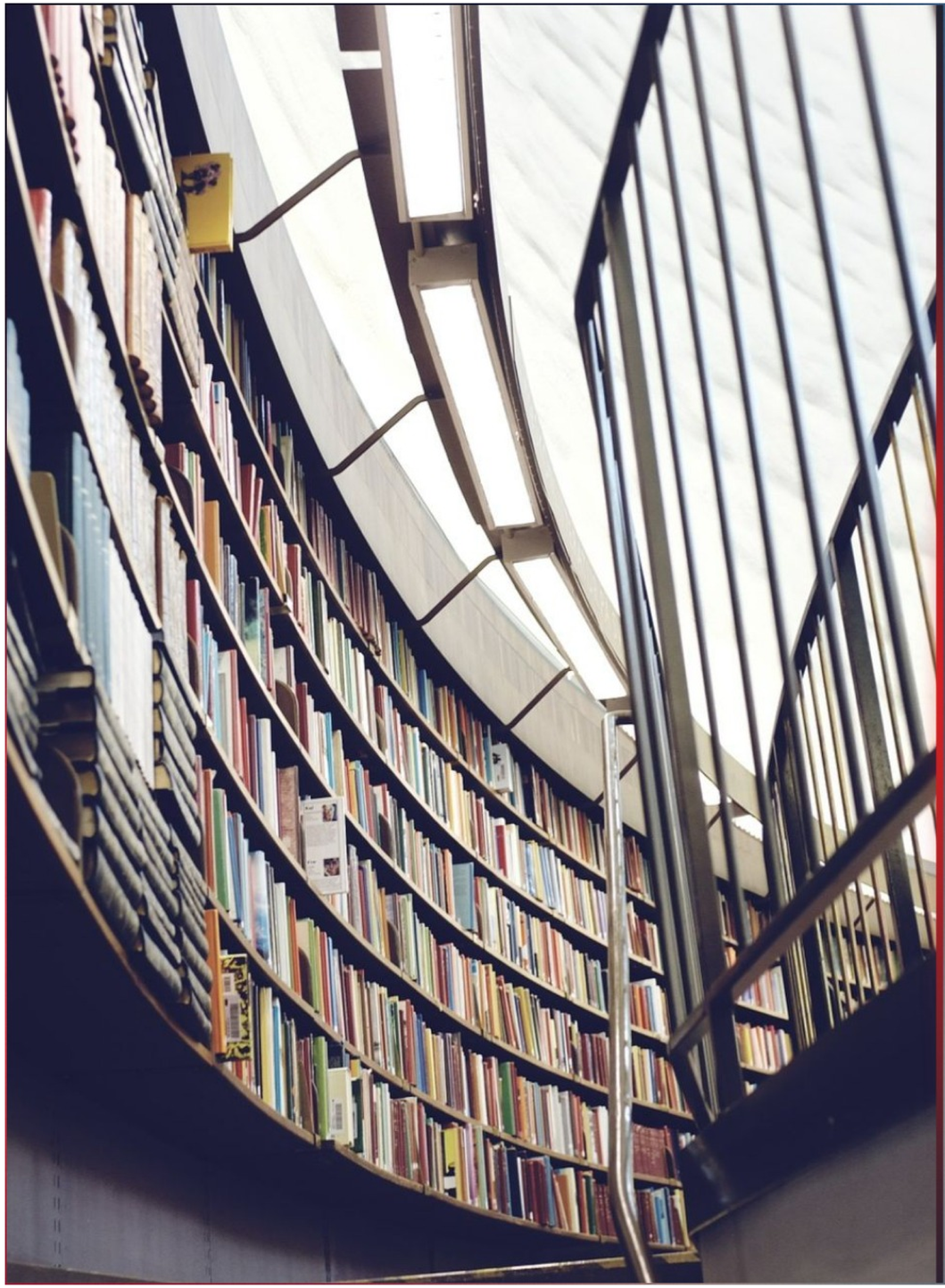
En este contexto, se destacan algunas limitaciones de la regresión múltiple, incluso con la aplicación de modelos jerárquicos: (contenido animado) .

- Estimar regresiones múltiples
- No hay estadísticas para evaluar la calidad del ajuste.
- Es difícil investigar hipótesis más específicas de mediación de efectos.
- Sin estimación de efectos indirectos.

Y así, el MEE, una técnica que involucra análisis de regresión múltiple + factor se aplica en estudios epidemiológicos para mejorar las estimaciones en los modelos de asociaciones causales. El modelado de ecuaciones estructurales (MEE) es una técnica general de modelado estadístico multivariante, que se usa ampliamente en las ciencias humanas y sociales, y se ha aplicado cada vez más en el área de la salud.

Aunque es una técnica generalizada en las humanidades, la salud aún se considera innovadora y con el potencial de estudiar asociaciones causales, principalmente para el análisis de caminos causales con la presencia de modificadores de efectos y factores de confusión. En salud es ampliamente utilizado en el área de la psicometría.





Puede verse como una combinación de análisis factorial

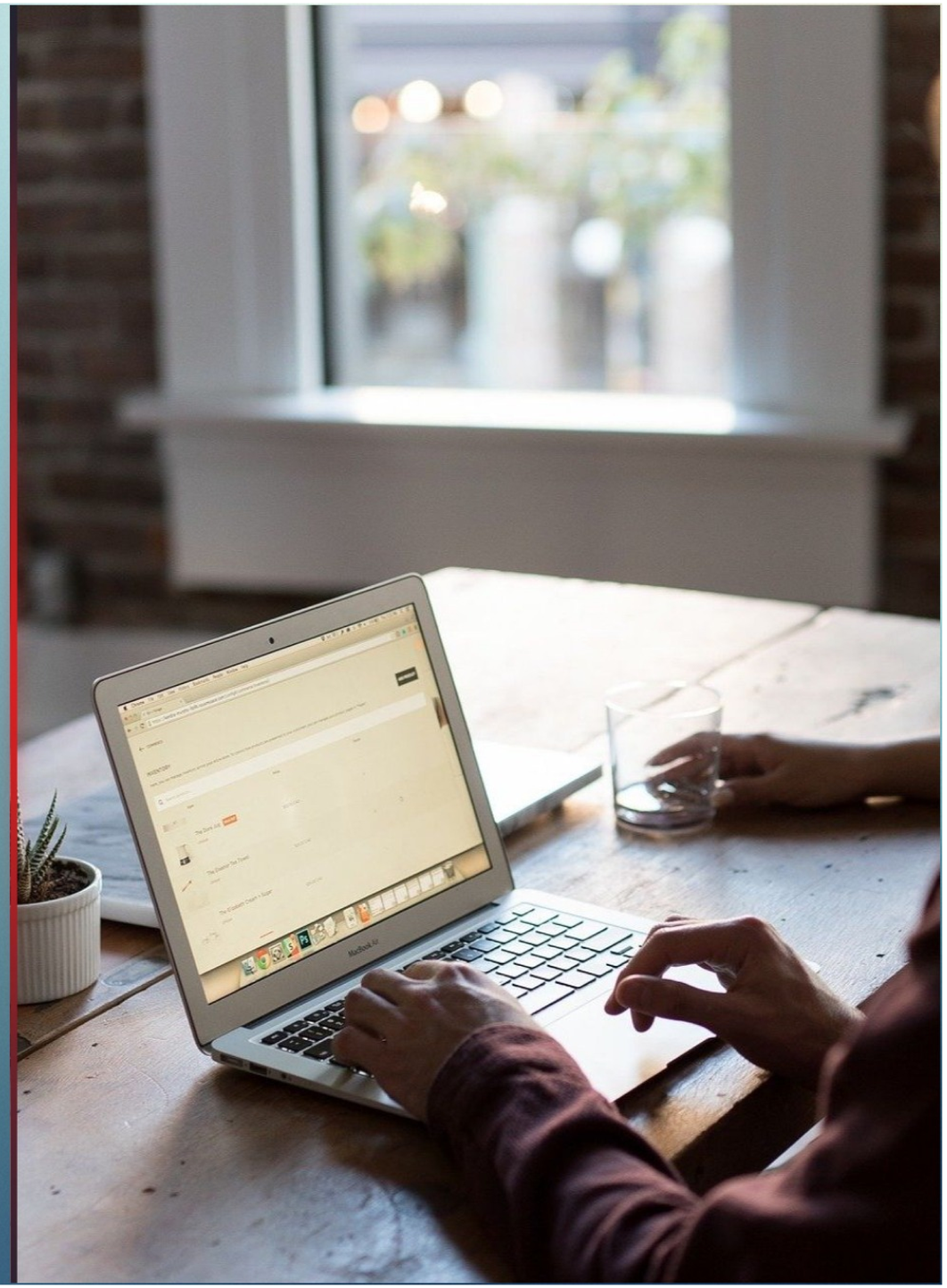
- ***Técnica estadística multivariada (no determinista) que permite la medición de variables latentes (construcciones no observadas directamente) a partir de un conjunto de variables manifiestas (observadas directamente).***

- ***Regresión o expandiéndose para el análisis de trayectorias o trayectos. (El análisis de trayectorias o trayectos es una extensión del análisis de regresión lineal de mínimos cuadrados. Esta extensión permite la descomposición de efectos estadísticos entre: efecto directo y efectos indirectos.)***

El interés de muchos investigadores y otros profesionales en MEE a menudo se deriva de las construcciones teóricas que se pueden desarrollar a partir de las construcciones latentes (cuando se usan otras variables para representar una variable que no era posible tener en la base de datos.

Las relaciones entre construcciones teóricas están representadas por coeficientes de regresión o coeficientes de trayectoria entre variables observadas y / o latentes. El modelo de ecuación estructural implica una estructura para la covarianza entre las variables observadas.

Los modelos de ecuaciones estructurales se visualizan, en la mayoría de los casos, mediante un diagrama de trayectoria.



Los modelos de ecuaciones estructurales son relevantes por las siguientes ventajas:

1. ***Permiten trabajar con estimación y medición simultáneamente;***

2. ***Permiten que se estimen los efectos directos e indirectos de las variables explicativas sobre las variables de respuesta;***

3. ***Son bastante robustas, debido a la relajación de los supuestos, en comparación, por ejemplo, con el modelo de regresión de mínimos cuadrados;***

4. ***Tienen una facilidad interpretativa resultante de sus interfaces gráficas.***

- *Análisis de ruta – modelo estructural con variables observadas*

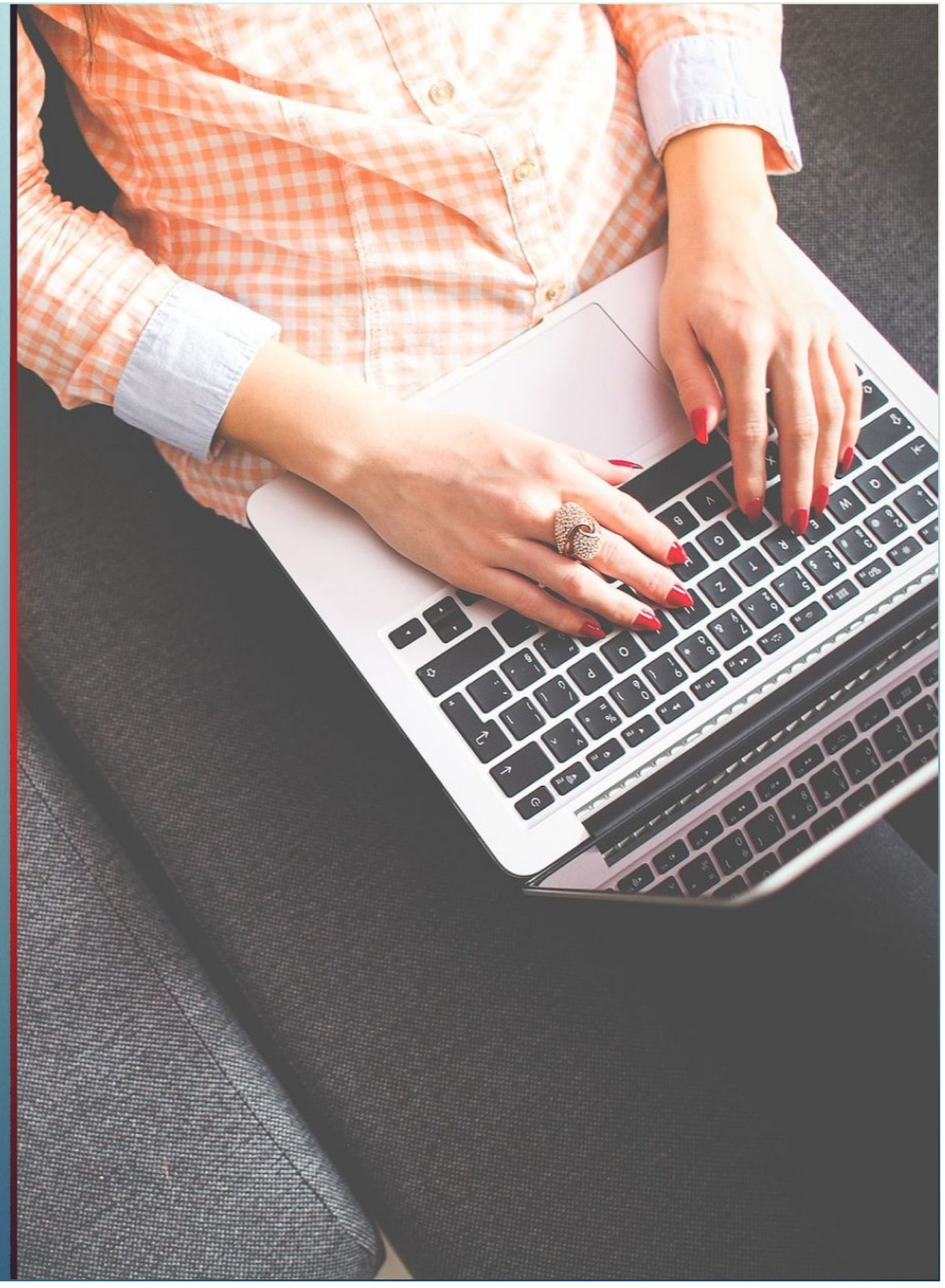
- *Análisis factorial confirmatorio – modelo de medición – con variable (s) latente (s)*

- *HÍBRIDO / MODELO COMPLETO – ANÁLISIS DE RUTA + ANÁLISIS FACTORIAL*
- *Notaciones en MEE*

- *En MEE los símbolos tienen representación universal:*

Variable observada, manifiesta o indicadora – medida por el observador

Variable latente – no observada, construcción hipotética, factor



Símbolos de relaciones entre variables.

1. ***X -> Y: Efecto directo - X afecta a Y unidireccionalmente.***

2. ***X-><-Y: Efecto recíproco - bidireccional.***

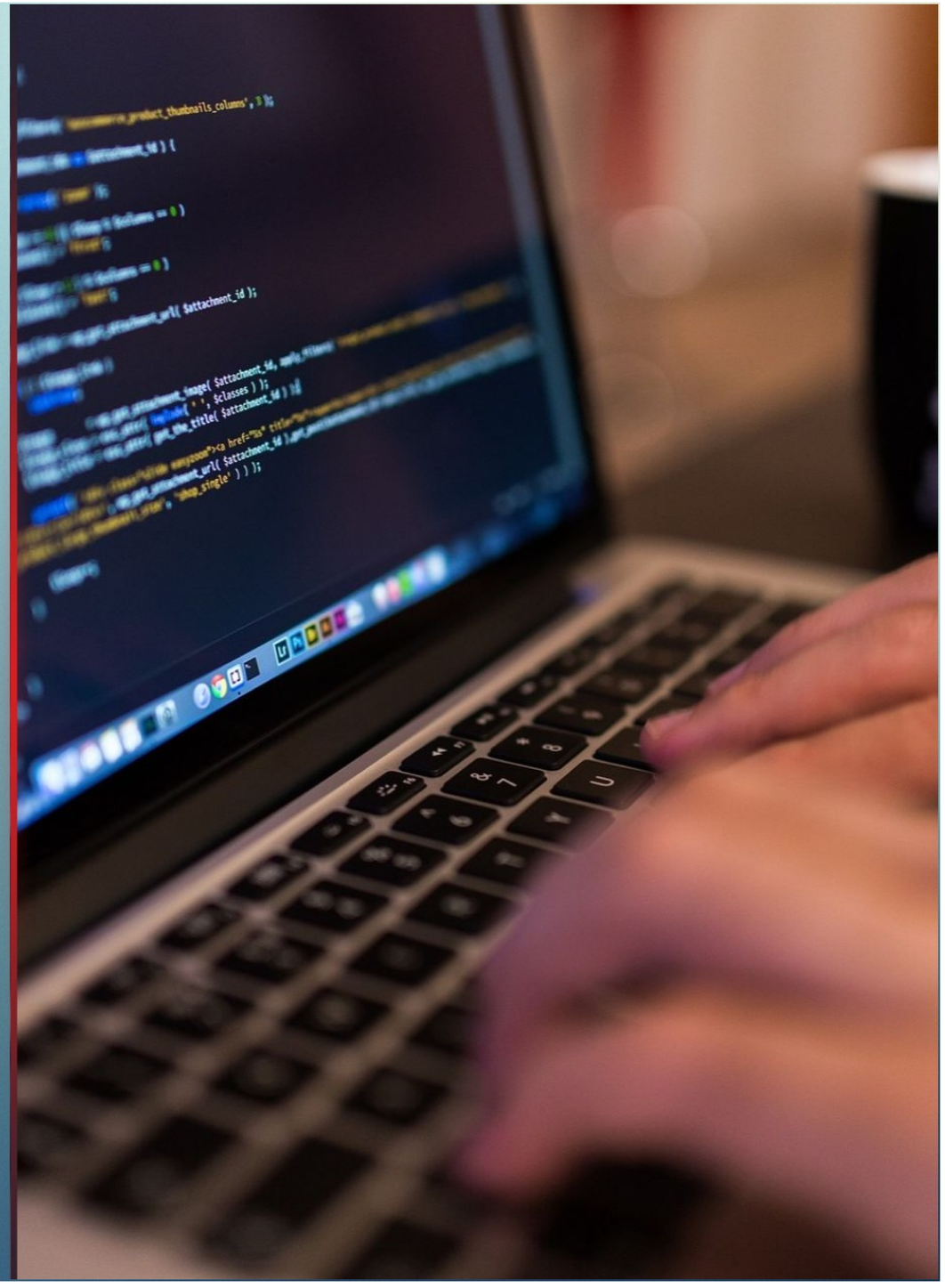
3. ***X<------>Y : Correlación o covarianza: asociación no analizada.***

En el modelo tenemos los siguientes tipos de variables:

Variable exógena: solo explicativa
Variable endógena: solo respuesta o variable explicativa

En cuanto a los efectos, tenemos lo directo o indirecto: es el efecto de una variable explicativa en la variable de respuesta por una variable mediadora o de confusión.

El desarrollo de computadoras y software proporcionó la generalización, integración y extensión de modelos del tipo de análisis de la varianza (ANOVA), análisis de regresión múltiple y análisis factorial confirmatorio, procesando simultáneamente el modelado de ecuaciones estructurales, lo que permite estimar simultáneamente las variables dependientes e independientes y las correlaciones entre ellas en un solo sistema de ecuaciones llamado estructural. El MEE se consolidó así como una contribución a la búsqueda de la mejora de los análisis causales, particularmente en las ciencias sociales y la psicología.



El epidemiólogo debe confiar en una buena teoría sustantiva y definiciones operativas rigurosas de sus categorías conceptuales, ya que es importante enfatizar que los análisis estadísticos correctos no tienen la capacidad de corregir errores teóricos o conceptuales, como la omisión de variables centrales o un concepto incorrecto guiando la medida de construcciones. Si, desde un punto de vista conceptual o teórico, en términos sustantivos, el análisis propuesto es correcto, la aplicación de modelos estadísticos multivariados, como el MEE, es de gran valor para la investigación científica o el análisis de datos técnicos.

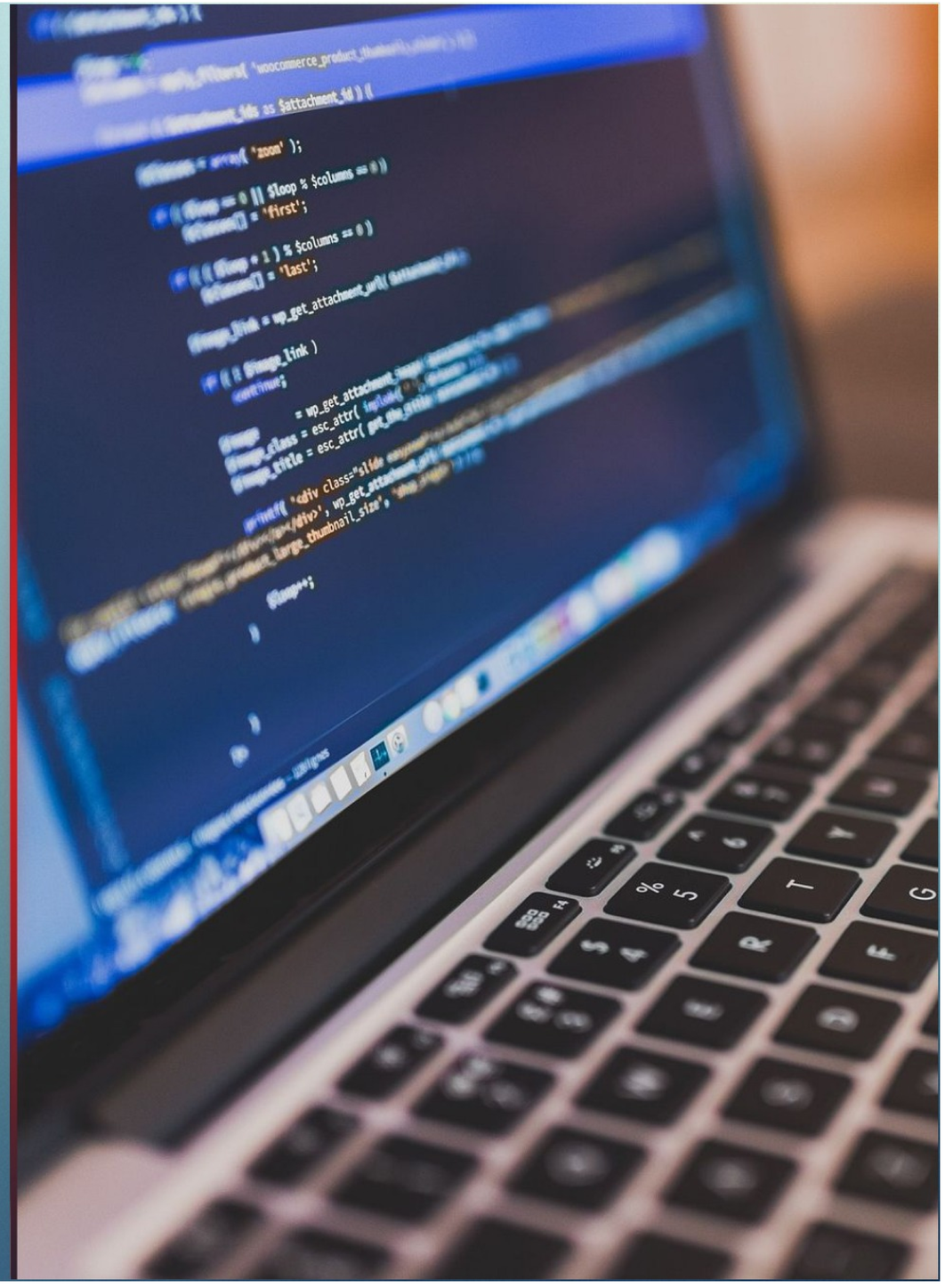
Por lo tanto, cuando alguien necesita desarrollar un MEE, lo mejor que puede hacer es comenzar con un modelo en el que haya una gran confianza teórica. Esta confianza deriva de fundamentos teóricos o empíricos sobre los vínculos entre las variables. Debido a su carácter confirmatorio, el MEE requiere que el análisis se inicie con un alto nivel de confianza en el modelo a estimar.

Por lo tanto, el MEE no es adecuado cuando hay poco conocimiento o confianza en el modelo inicial, independientemente de si el resultado final es teórica o empíricamente plausible.
Lo que se indica es que ya hay estudios exploratorios que dan pistas para la confirmación del modelo teórico y que se utilizan los conceptos de causalidad de la estadística clásica:

a) Una variable X debe considerarse una causa de una variable Y cuando existe una asociación estadística entre ellas (condición de covarianza).

b) Una variable X debe considerarse una causa de una variable Y, cuando X precede a Y en el tiempo (condición de precedencia temporal).

c) Una variable X debe considerarse una causa de una variable Y cuando otras posibles causas de Y se miden e incluyen en el análisis (condición de eliminar las causas en competencia).



¡Recordatorio! Claro que con el predominio del modelo teórico establecido a priori.

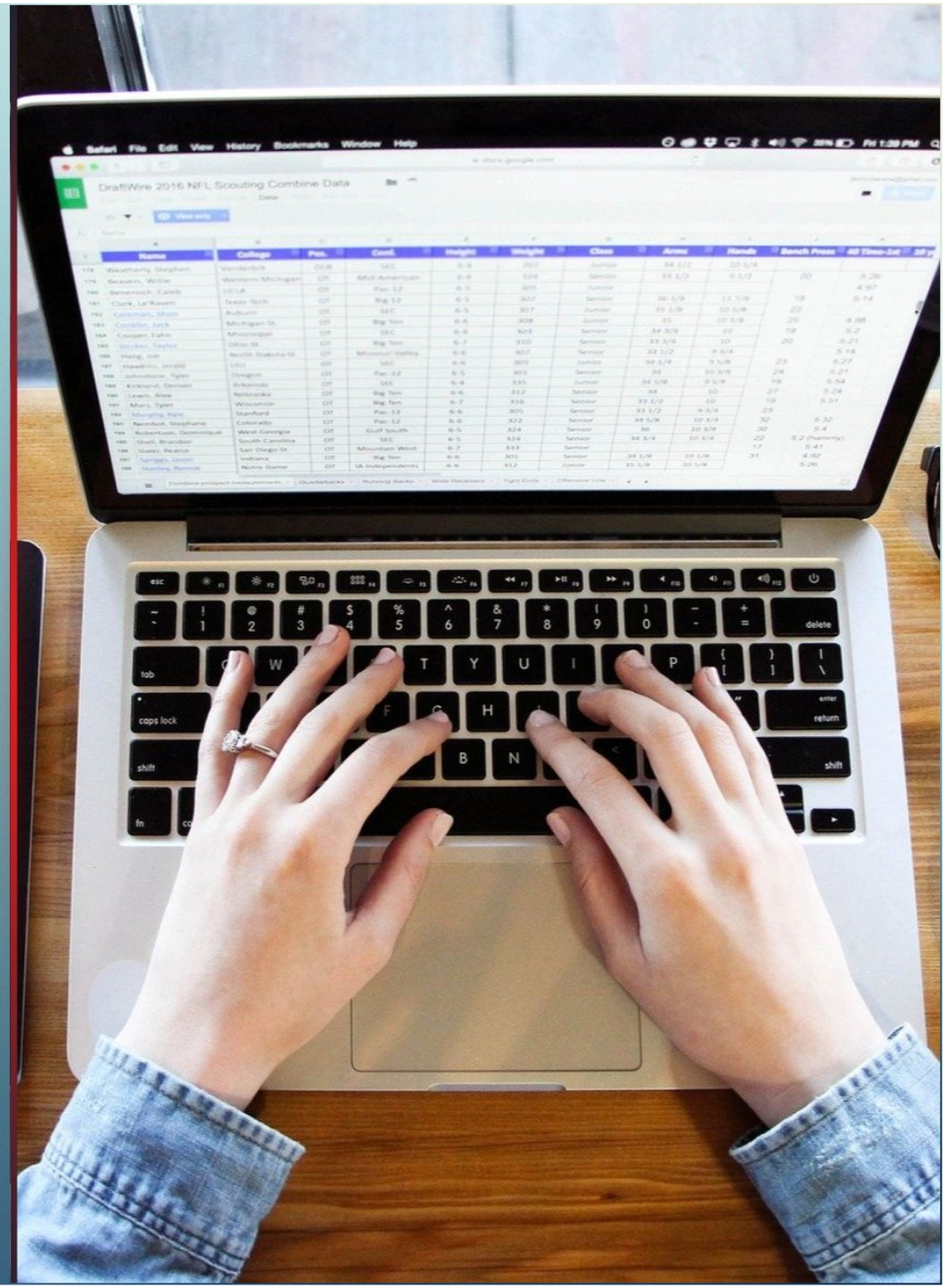
Los MEE recursivos (unidireccionales) cumplen la mayoría de los supuestos de los modelos de regresión de mínimos cuadrados, a saber:

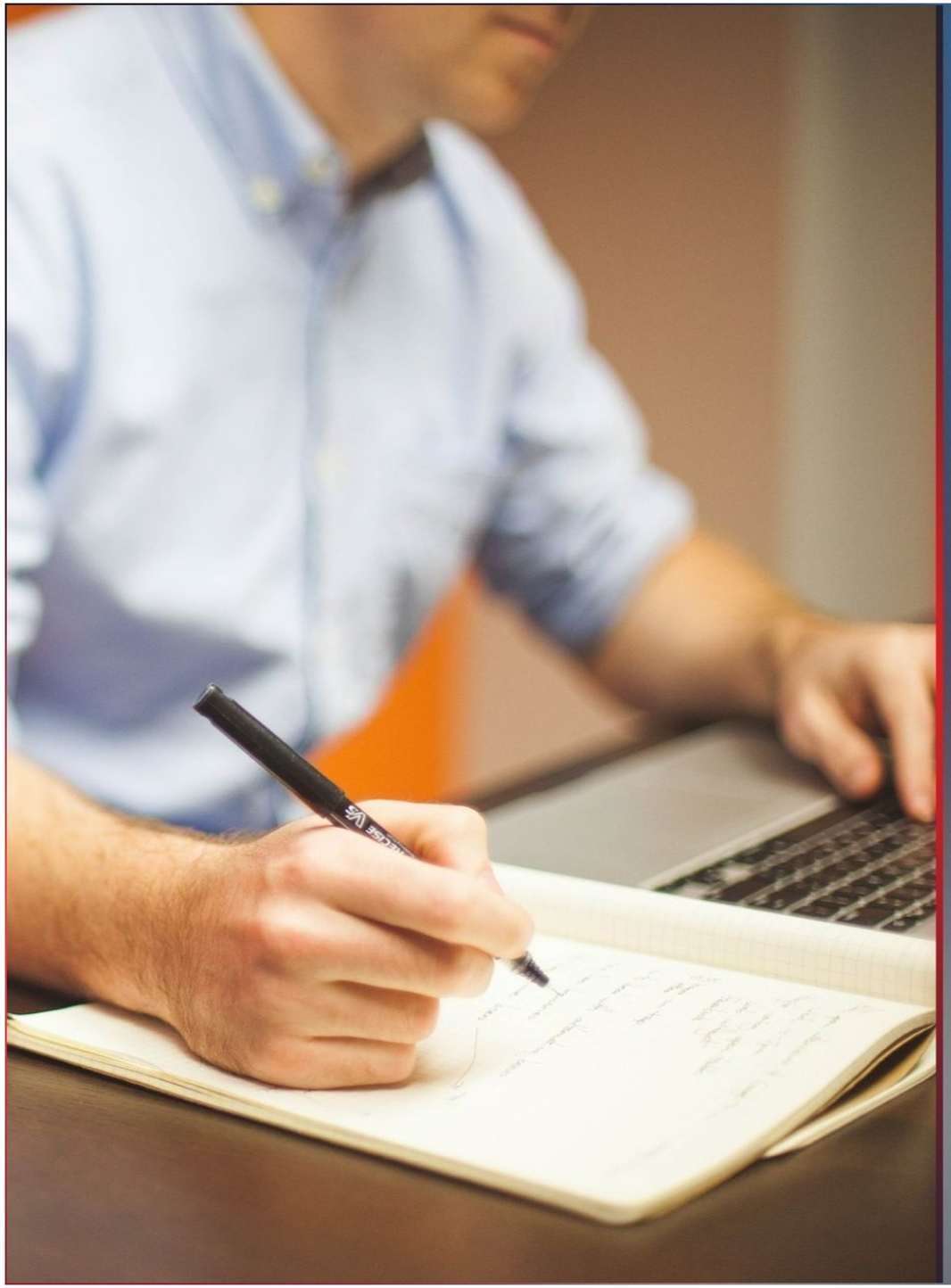
- Independencia de errores.
- Independencia de las observaciones.
- Linealidad.
- Normalidad de los errores.
- Homocedasticidad.
- Aleatoriedad.

Sin embargo, en general, los MEE permiten la estimación de modelos con altos niveles de colinealidad entre variables explicativas. En el caso de modelos no recursivos (bidireccionales), la suposición de independencia de error puede ser relajada.

El análisis factorial utiliza la matriz de covarianza para estimar el factor estructural resultante del análisis factorial desarrollado inicialmente para explicar la correlación entre las variables. El propósito de usar la matriz de covarianza (o correlación, en el caso de modelos con coeficientes estandarizados) es reducir los tipos de variables estandarizadas en el modelo.

Los modelos de factores se conocen como subidentificados, identificados o sobreidentificados. Si el modelo está subidentificado, no se puede estimar porque no tiene una solución única. Cuando se identifica el modelo, hay una solución exacta para la matriz de covarianza de los datos, lo que implica un ajuste perfecto. Si el modelo está sobreidentificado, él no se ajusta perfectamente.





El MEE completo, por lo tanto, se compone de la integración de modelos de regresión con análisis factoriales confirmatorios. Los modelos de regresión proporcionan la parte determinista del ESM (el modelo causal), mientras que los análisis factoriales confirmatorios proporcionan la parte relacionada con las mediciones.

Los conceptos presentados aquí son esenciales para la continuidad de nuestro aprendizaje en las próximas semanas. Para entender la aplicabilidad de lo poco que se presentó hoy, lea el artículo de Silva et al. (2010).

En las próximas semanas detallaremos el modelo de análisis confirmatorio, el modelo estructural y finalmente, analizaremos el modelo híbrido.

Unidad basada en el libro: Neves, J. A. B. (2018). Modelo de equações estruturais: Uma introdução aplicada. ENAP y en el curso de modelación con ecuaciones estructurales ofrecidas por la Universidad Federal de Maranhão (Brasil), por el profesor Antonio Augusto Moura.

Neves, J. A. B. (2018). Modelo de equações estruturais: Uma introdução aplicada. ENAP.

Silva, A. A. M. da, Vasconcelos, A. G. G., Bettiol, H., & Barbieri, M. A. (2010). Socioeconomic status, birth weight, maternal smoking during pregnancy and adiposity in early adult life: An analysis using structural equation modeling. Cadernos de Saúde Pública, 26(1), 15–29. https://doi.org/10.1590/S0102-311X2010000100003

PUCE Virtual

Quito
Av.12 de Octubre 1073 y Roca
Edificio de la Facultad de Comunicación,
Lingüística y Literatura. Primer Piso. Oficina 106.

Teléfono:
(593-2) 299 1592 / (593)09 8 851 2839

Correo:
soportevirtual@puce.edu.ec