Epidemiología Moderna: Conceptos y Modelos Avanzados

Análisis multivariante en epidemiología: estratificación y ajuste

Epidemiología para la Salud Pública

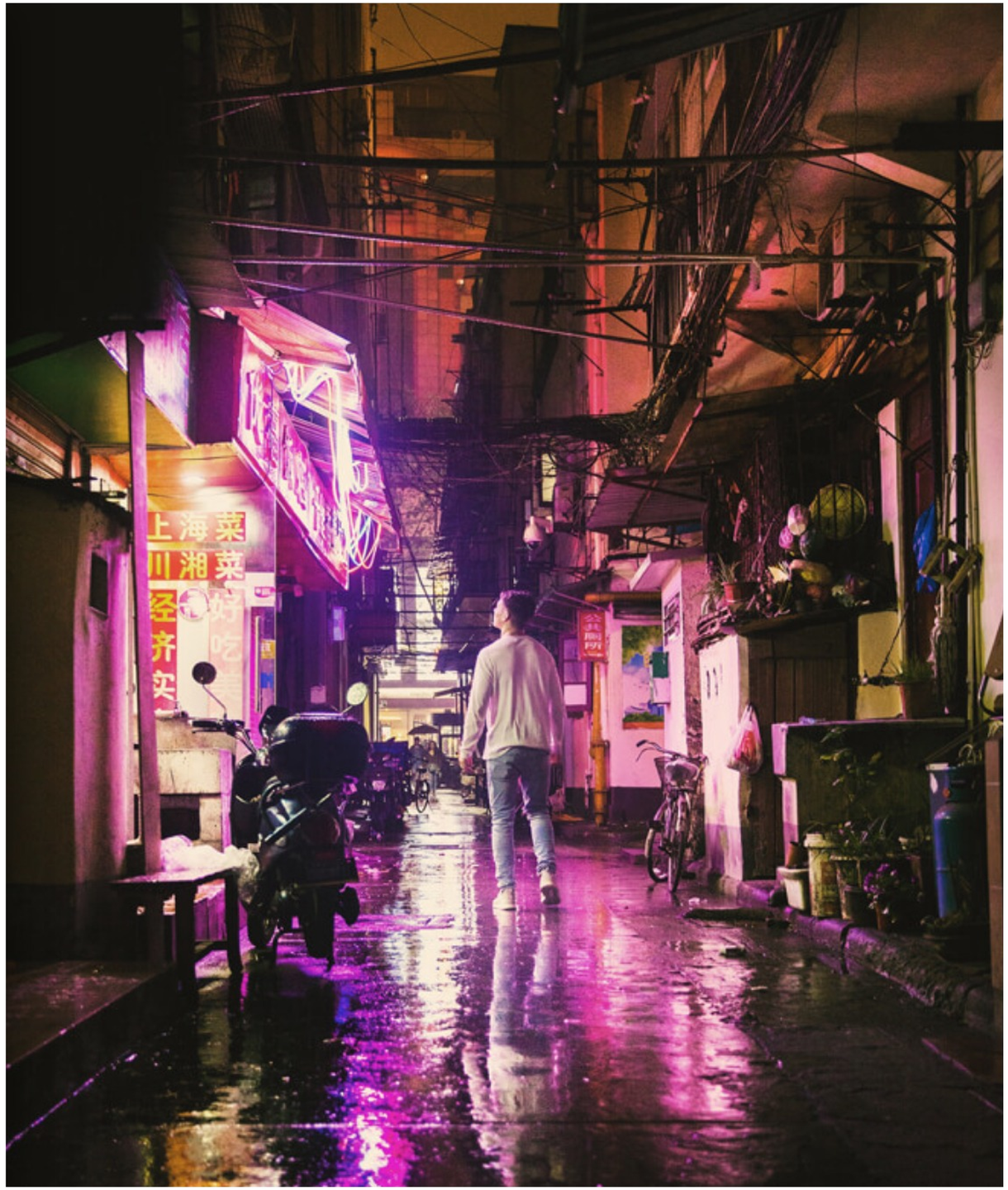
Tema 5.1. Métodos de ajuste basados en estratificación

El método más simple para analizar la posible presencia de confusión es la estratificación. Permite un enfoque directo y simultáneo de la posible presencia de modificación del efecto y confusión. Los resultados estratificados suelen ser útiles para elegir la técnica estadística adecuada para el ajuste.

1- Técnicas de estratificación y ajuste para el manejo de la confusión.

Para comprender cómo la estratificación aclara el fenómeno de la confusión, analicemos el ejemplo en la siguiente tabla (Tabla 1) , que resume los resultados de un estudio de casos y controles sobre el uso de anticonceptivos orales (AC) como un posible factor de riesgo de infarto de miocardio en mujeres en edad fértil.

Como se muestra en la parte superior de la Tabla 1, se estimó que la probabilidad de enfermedad entre las mujeres que usaban anticonceptivos orales era aproximadamente 70% mayor que la de las que no usaban anticonceptivos orales. Sin embargo, los autores de este estudio consideraron la posibilidad de confusión por edad.

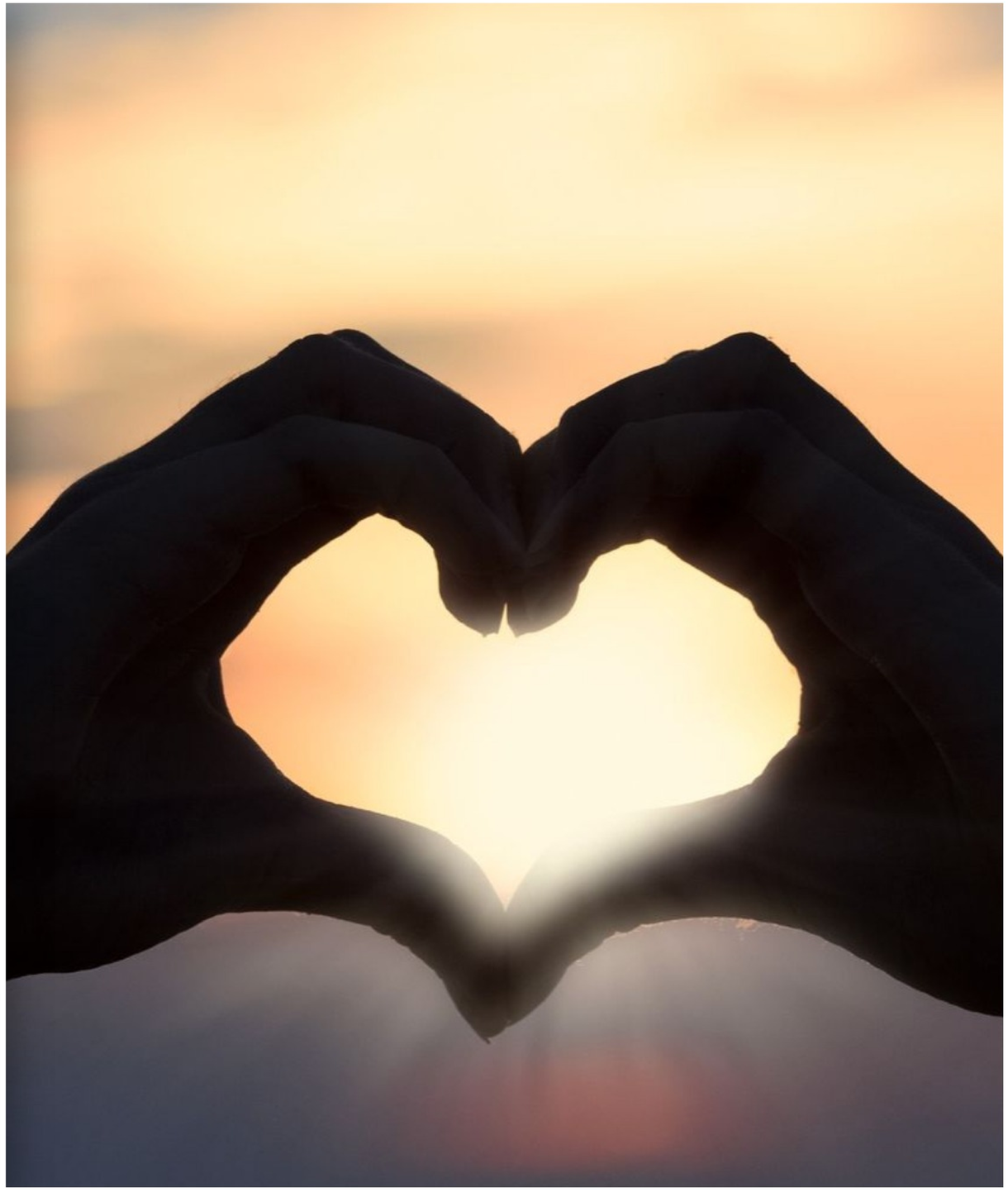


¡Recordatorio! ¿Cómo saben los autores que la edad es un posible factor de confusión?

Debido a la construcción de la hipótesis basada en el fondo científico, es decir, al diseñar la planificación del estudio, es importante hacerlo basándose en la literatura científica. Resalta el recordatorio como un ícono.

Como se sabía que la edad estaba directamente relacionada con el resultado (riesgo de infarto de miocardio) e inversamente relacionada con la exposición (disminución del uso de anticonceptivos orales entre mujeres mayores), se infirió que la edad podría actuar como un factor de confusión negativo.

En un análisis estratificado por edad, todos los estratos menos uno, tenían odds ratio que eran más distantes que 1,0 que la estimación bruta global, lo que confirma la expectativa de confusión negativa (es decir, la edad conduce a la asociación bruta estimada hacia cero).



El odds ratio ajustada para los datos en la Tabla 1 fue de 3.9, más del doble de la estimación bruta. Como se mencionó anteriormente, implícito en el cálculo de cualquier promedio, esta estimación de la razón de probabilidades ajustada requiere que las asociaciones sean homogéneas (es decir, que las diferencias observadas entre los estratos en los odds ratio resultan de una variación aleatoria).

En este ejemplo, esta suposición es probablemente razonable, dado el pequeño número de casos en algunas celdas y el hecho de que todas los odds ratio están en la misma dirección (lo que denota la ausencia de interacción cualitativa).

Por otro lado, se puede argumentar que las diferencias cuantitativas son muy grandes y, por lo tanto, no se justifica la estimación de un promedio de estimación único (ajustado) supuestamente representativo de todos los grupos de edad.

Por ejemplo, se puede argumentar que la asociación parece ser más fuerte en mujeres menores de 35 años (odds ratio de 7.2 y 8.9) que en mujeres mayores (odds ratio que van de 1.5 a 3, 9).

La aceptación de esta heterogeneidad de las estimaciones sugiere un enfoque alternativo que consiste en calcular dos razones de probabilidad ajustadas por edad: una para las mujeres de 25 a 34 años (es decir, la probabilidad promedio ponderada de la relación entre los dos grupos más jóvenes) y otra para mujeres de 35 a 49 años (es decir, el promedio ponderado de los tres grupos más antiguos).

Este ejemplo ilustra las ventajas de la estratificación para evaluar la presencia de confusión y / o interacción y para decidir cuándo el ajuste es apropiado y cómo debe realizarse.

También ilustra una situación común en el análisis epidemiológico: la exposición de interés parece tener efectos heterogéneos de acuerdo con una cierta agrupación de la tercera variable, a veces no considerada antes del análisis de datos.

Dada la gran cantidad de posibilidades para agrupar variables al realizar análisis estratificados y los posibles factores aleatorios de variabilidad de los efectos aparentes en los subgrupos, este tipo de análisis, si no se señala en hipótesis a priori biológicamente plausibles, debe considerarse simplemente exploratorio.

Tabla 1: Ejemplo de análisis estratificado: estudio de casos y controles de anticonceptivos orales (AC) e infarto de miocardio en mujeres.

Análisis crudo				
Todos los casos y controles				
		Casos	Controles	
	AC	29	135	OR = 1.7
	No AC	205	1607	
	234		1742	

Análisis estratificado por edad							
25–29 años				30–34 años			
	Casos	Controles			Casos	Controles	
AO	4	62	OR = 7.2	AO	9	33	OR = 8.9
Sin AO	2	224		Sem AO	12	390	
35–39 años				40–44 años			
	Casos	Controles			Casos	Controles	
AO	4	26	OR = 1.5	AO	6	9	OR = 3.7
Sin AO	33	330		Sin AO	65	362	
45–49 años							
	Casos	Controles					
AO	6	5	OR = 3.9				
Sin AO	93	301					

Tabla reproducida de (Szklo y Nieto, 2019), p.263; datos de un estudio publicado (Shapiro et al., 1979).

Suposiciones (premisas) para estratificación y ajuste

En comparación con el ajuste, la *estratificación* está virtualmente (pero no completamente) libre de supuestos.

La *estratificación* requiere asumir que los estratos son significativos y están adecuadamente definidos. Esto significa que *debe haber homogeneidad dentro de cada estrato*. Por ejemplo, en la Tabla 1, se supone que la asociación de los anticonceptivos orales con el infarto de miocardio es homogénea cada 5 años en el grupo de edad. Si este supuesto no fuera apropiado, se deberían definir otras categorías más precisas para el análisis estratificado.

Para calcular un odds ratio promedio (ponderada), se supone que los estratos son homogéneos (es decir, que no hay interacción [multiplicativa]).

El análisis multivariante se refiere a una serie de técnicas analíticas, cada una basada en un modelo matemático, utilizada para realizar ajustes estadísticos, es decir, la estimación de una determinada medida de asociación entre una exposición y un resultado cuando se controlan una o más variables de confusión. La modificación del efecto también se puede evaluar en el contexto del análisis multivariado.

Para el ajuste, se deben cumplir otras premisas.

Todas las técnicas de ajuste se basan en asumir algún tipo de suposición en el modelo estadístico, que resume la asociación entre las variables bajo investigación. Es decir, las premisas dependerán del modelo estadístico adoptado.

Los métodos de ajuste basados en datos se usan con mayor frecuencia cuando se controla un número limitado de posibles factores de confusión que son categóricos o que se pueden clasificar. Por otro lado, los modelos matemáticamente más complejos son la base de los métodos de ajuste multivariados, como los métodos de regresión.

Las técnicas más utilizadas para el ajuste y el análisis multivariado de datos epidemiológicos.

¡Cuidado! Cada una de las técnicas analíticas descritas a continuación se basa tanto en un modelo conceptual como en un modelo matemático.

2 - Métodos de ajuste basados en la estratificación

2.1. Ajuste directo

El método directo es el procedimiento que compara las tasas de incidencia entre dos grupos, A y B (por ejemplo, expuestos y no expuestos), estratificados según la variable de sospecha de confusión.
El ajuste directo se ha utilizado tradicionalmente para ajustar la edad al comparar la morbilidad y la mortalidad entre países o regiones en diferentes momentos, aunque ajustar la edad no es su única aplicación.

La popularidad de las estadísticas más sofisticadas ha limitado el uso del método de ajuste directo, pero sigue siendo una técnica simple que es particularmente útil para ilustrar los principios básicos del ajuste estadístico.

El método directo se puede usar para ajustar cualquier tasa o proporción (mortalidad, incidencia por persona-tiempo, prevalencia), por lo que este método también se puede usar en el contexto de un estudio de casos y controles para obtener las proporciones ajustadas de casos expuestos y controles, que a su vez se pueden usar para calcular u odds ratio ajustado.

2.2. Ajuste indirecto

Al igual que el método de ajuste directo, el ajuste indirecto se ha utilizado tradicionalmente para ajustar la edad de los datos de mortalidad y morbilidad. En el método indirecto de ajuste, particularmente popular en el campo de la epidemiología ocupacional, el número esperado de eventos (por ejemplo, muertes) en un grupo de estudio (por ejemplo, una cohorte ocupacional) se calcula aplicando tasas de referencia (tasas "estándar") al número de individuos en cada estrato de los grupos de estudio.

Para cada grupo en estudio, la relación entre el número total de eventos observados y el número de eventos esperados (tasas estandarizadas) proporciona una estimación del riesgo relativo ajustado por factores o tasa de comparación entre el grupo de estudio y la población que sirvió como fuente De referencia.

Cuando se usa en el contexto de los datos de mortalidad, esta proporción se conoce como ***la tasa de mortalidad estandarizada (TME)***, términos similares se aplican a los datos de morbilidad, ***como la incidencia estandarizada y la razón de prevalencia estandarizada.***

2.3. Método de Mantel-Haenszel para estimar una medida de asociación ajustada

Cuando la medida de asociación de interés es la razón de posibilidades (por ejemplo, al analizar los resultados del estudio de casos y controles), el método descrito por Mantel y Haenszel a menudo se aplica para ajustar uno o más posibles factores de confusión. El supuesto implícito en el procedimiento de Mantel-Haenszel para calcular un odds ratio ajustado es que existe una homogeneidad de efectos en las categorías de variables estratificadas. En otras palabras, se supone que no hay interacción multiplicativa entre la exposición y el posible factor de confusión.

Las técnicas descritas anteriormente (ajuste directo e indirecto, Mantel-Haenszel) pueden aplicarse para análisis multivariados, por lo tanto, para controlar simultáneamente más de una covariable.

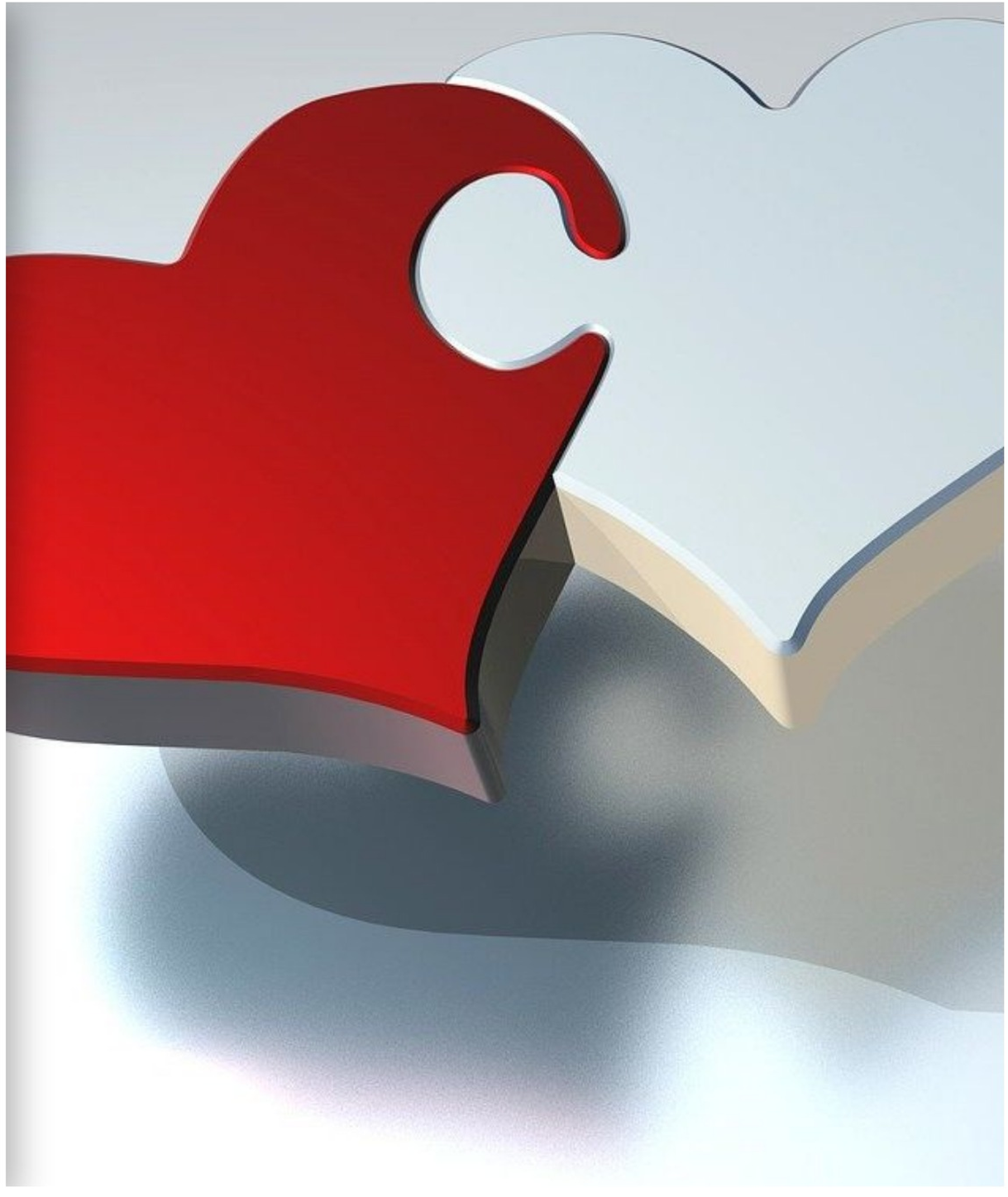
Esto puede hacerse simplemente construyendo los estratos basados en todas las combinaciones de las variables de ajuste.

Sin embargo, estos métodos basados en estratificación tienen limitaciones prácticas para el ajuste multivariado:

- Aunque pueden usarse para ajustar varias covariables simultáneamente, el método de ajuste se realiza solo para la asociación entre una variable independiente y un resultado a la vez. Por ejemplo, para evaluar la asociación de los anticonceptivos orales con el infarto de miocardio, al controlar la edad y el nivel educativo, sería necesario crear una tabla 2 × 2 para los anticonceptivos orales en relación con el infarto de miocardio para cada estrato definido por la combinación de grupos de edad y niveles educativos. Si la exposición de interés es la escolarización y las covariables a ajustar, la edad y los anticonceptivos orales, se tendría que crear un nuevo conjunto de tablas 2 × 2 (que representa la educación versus el infarto de miocardio para cada estrato definido por edad y categorías de anticonceptivos orales)

- Estos métodos permiten ajustes solo para covariables categóricas (por ejemplo, sexo); Las covariables continuas deben clasificarse. Las diferencias residuales en las categorías definidas arbitrariamente pueden, a su vez, generar confusión residual.

- Técnicas de regresión múltiple para el ajuste. Las técnicas de regresión múltiple son más apropiadas para tratar las limitaciones de las técnicas más simples discutidas hasta ahora, ya que permiten el análisis de los efectos de todas las variables de exposición, con ajustes recíprocos y simultáneos para todas las demás variables del modelo. Además, permiten el ajuste de covariables continuas y, dentro de límites razonables, generalmente son más eficientes que los métodos estratificados. Además de las técnicas de regresión múltiple y múltiple, son útiles para pronosticar (es decir, para estimar el valor pronosticado de un resultado dado de acuerdo con ciertos valores de variables independientes, como en el caso de las ecuaciones de predicción de riesgo coronario que se desarrollaron de Framingham).



Los cuatro modelos de regresión más frecuentemente aplicados para modelos de ajuste en epidemiología:

1. Regresión lineal, utilizada cuando el resultad es continuo (por ejemplo, presión arterial).

2. Regresión logística, preferiblemente utilizada cuando el resultado es categórico (incidencia acumulada, prevalencia).

3. Regresión proporcional de Cox, utilizada en e análisis de supervivencia.

4. Regresión de Poisson, utilizada cuando la tas de incidencia (basada en el tiempo de la persona) es el resultado.

La siguiente tabla muestra los modelos matemáticos de regresión múltiple.

Tabla 2: Modelos matemáticos de regresión múltiple

	Modelo	Interpretando b_1
Lineal	$y = b_0 + b_1x_1 + b_2x_2 + \ldots + b_kx_k$	Aumenta el valor de resultado promedio (variable continua) con cada unidad de aumento en x1, ajustado para todas las demás variables en el modelo
Logística	$\log(\text{odds}) = b_0 + b_1x_1 + b_2x_2 + \ldots + b_kx_k$	Aumento en el registro de las probabilidades del resultado para cada unidad de aumento en x1, ajustado para todas las demás variables en el modelo
Cox	$\log(\text{risco}) = b_0 + b_1x_1 + b_2x_2 + \ldots + b_kx_k$	Aumento en el registro de riesgos del resultado para cada unidad de aumento en x1, ajustado para todas las demás variables en el modelo
Poisson	$\log(\text{rate}) = b_0 + b_1x_1 + b_2x_2 + \ldots + b_kx_k$	Aumento de registro en la tasa de resultado para cada unidad de aumento en x1, ajustado para todas las demás variables en el modelo

Tabla reproducida de (Szklo y Nieto, 2019), p.280; datos de un estudio publicado (Shapiro et al., 1979).

Pruebas estadísticas e intervalos de confianza de estimaciones de regresión

Los valores de los parámetros en regresión lineal, ya sea regresión lineal simple o múltiple, se estiman utilizando el método de mínimos cuadrados comunes.
Este método de estimación es bastante simple y, en el caso de una regresión lineal simple, se puede realizar con una calculadora de bolsillo. Sin embargo, para la regresión lineal multivariante, el cálculo es más complejo. Por otro lado, el método simple de mínimos cuadrados no se puede aplicar a los otros tres métodos de regresión descritos anteriormente.

La regresión logística, la regresión de Cox y la regresión de Poisson requieren un proceso de estimación basado en el método de estimación de probabilidad, una tarea que se ve facilitada en gran medida por el uso de computadoras modernas y software estadístico.

Con todo, incluso con la facilidad del software para calcular las estimaciones, es necesario comprender el modelo y saber cuál es el más adecuado para el estudio.

Ajuste incompleto: confusión residual

La confusión residual ocurre cuando el ajuste no elimina por completo el efecto del factor de confusión debido a una variable dada o un conjunto de variables introducidas en el estudio. Las fuentes de ajustes incompletos y / o confusión residual son diversas, siendo las más importantes:

- Las categorías de la variable de confusión no están definidas correctamente.
- La variable utilizada para el ajuste es un sustituto imperfecto de la condición o característica que el investigador desea ajustar.
- Otros factores de confusión relevantes para el estudio no están incluidos en el modelo.
- Variables de confusión clasificadas incorrectamente.

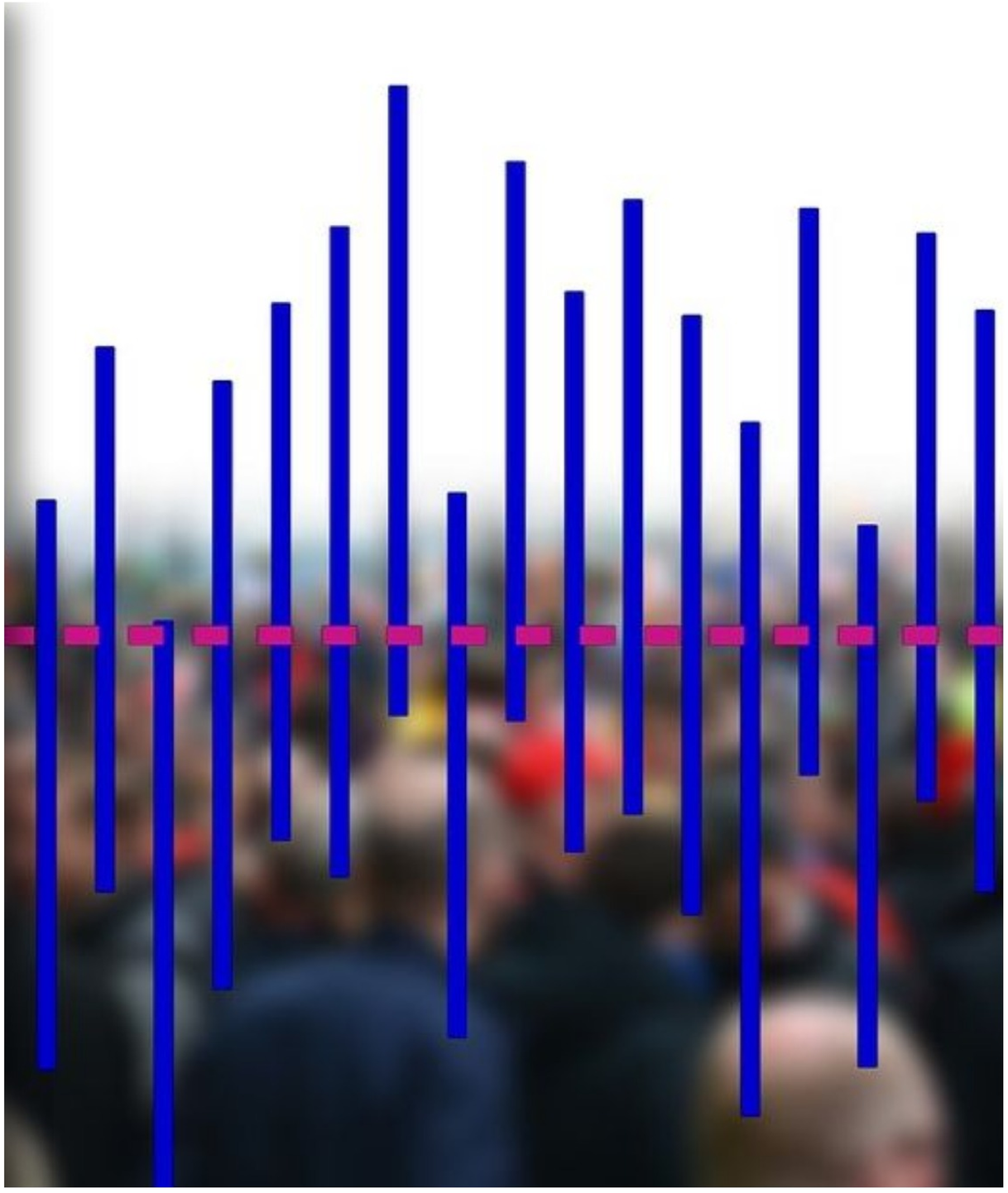
5- Sobreajuste

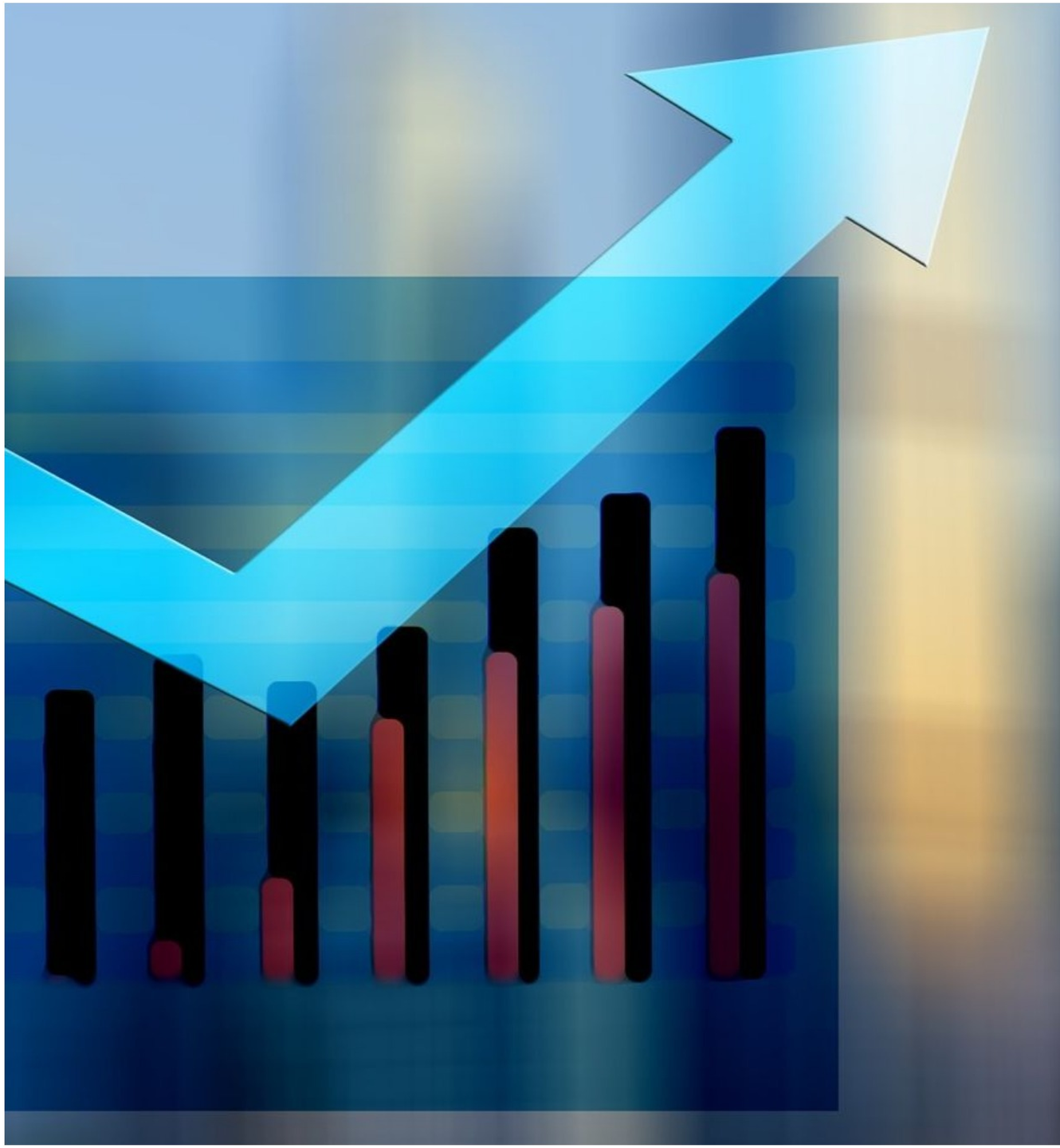
El ajuste para una variable dada implica, al menc parcialmente, el ajuste para otras variables relacionadas. Por lo tanto, cuando se ajusta por e nivel educativo, el ajuste por ingresos también se lleva a cabo, hasta cierto punto. Dependiendo de las características específicas de la población de estudio, el ajuste por residencia dará como resultado el ajuste de variables relacionadas, cor el estado socioeconómico o el origen étnico.

El sobreajuste ocurre cuando el ajuste se realiza inadvertidamente para una variable que está en l ruta causal entre la exposición y el resultado (causa intermedia) o tan fuertemente relacionada con la exposición o el resultado que su verdader asociación está distorsionada. Por lo tanto, un ajuste excesivo puede ocultar un efecto real o crear un efecto aparente cuando no existe ningu

¡Seleccionando el modelo estadístico correcto!

Ya entendemos cómo modelar problemas relacionados con la asociación entre una determinada variable independiente y una variable dependiente. Obviamente, sin embargo, este no es el único problema en el modelado estadístico.





Otras decisiones fundamentales en cualquier análisis multivariado implican:

1. La elección de una técnica de regresión o estratificación específica .

2. La elección de las variables que se incluirán la ecuación o estratificación .

La regresión lineal se indica cuando el resultado continuo. Cuando se trata de una variable binaria las técnicas de ajuste basadas en modelos multiplicativos populares (por ejemplo, odds rati regresión logística, regresión de Cox, regresión d Poisson) tienden a proporcionar resultados similares.

¿Debemos entender mejor en la práctica la idea de que los modelos multiplicativos producen resultados similares?

Deténgase y lea el estudio de Coutinho et al., 2008 , donde se utilizaron diferentes métodos de ajuste basados para obtener estimaciones ajustadas de las asociaciones de factores de riesgo (nivel educacional, edad y género) self-rated poor health y trastornos mentales en un conjunto de personas mayores en Sao Paulo (sureste de Brasil). Todos los métodos dieron como resultado estimaciones comparables, lo que llevó a conclusiones similares.

Muchos de estos métodos utilizados en epidemiología (por ejemplo, Mantel-Haenszel, logística, Poisson, Cox) se basan en supuestos similares (es decir, relaciones multiplicativas entr exposición y resultado) y tienen significados epidemiológicos similares.
No es sorprendente que, con la excepción de circunstancias extraordinariamente extremas (po ejemplo, violaciones graves de las premisas básicas del modelo, valores extremos), estos métodos generalmente producen resultados similares.

Para un ajuste y elección de un modelo apropiad la pregunta "¿qué variables independientes (factores de confusión) deben incluirse en el modelo?" se centra en el núcleo de la discusión para garantizar datos adecuadamente ajustados. Finalmente, una recomendación importante al elegir un modelo estadístico específico es realiza un análisis de sensibilidad (es decir, verificar que obtengan resultados similares cuando se utilizan diferentes modelos o suposiciones para el anális

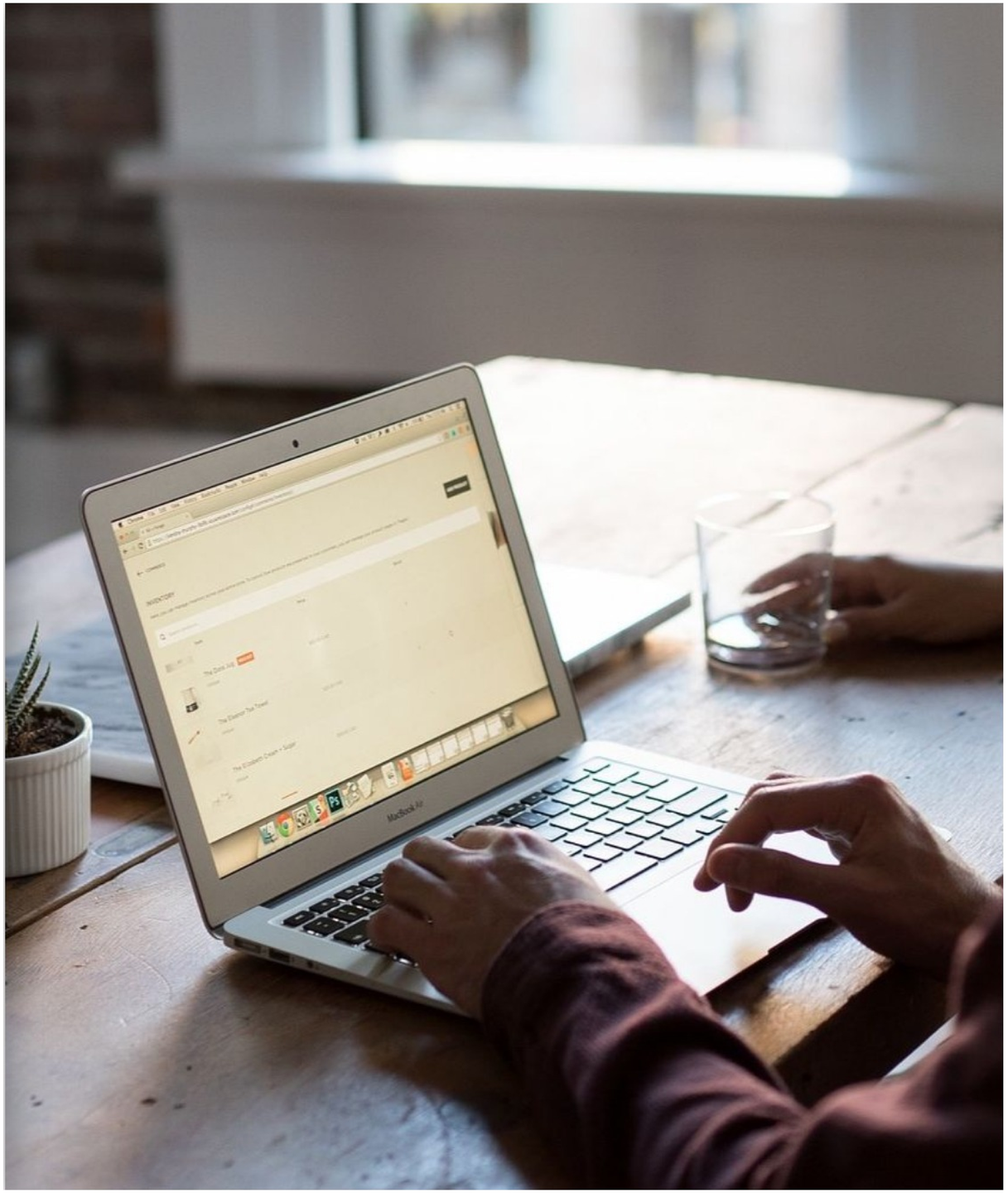
Referencias

Bhopal, R. S. (2016). Concepts of epidemiology: Integrating the ideas, theories, principles, and methods of epidemiology (Third edition). Oxford University Press.

Coutinho, L. M. S., Scazufca, M., & Menezes, P. R. (2008). Methods for estimating prevalence ratios in cross-sectional studies. Revista De Saude Publica, 42(6), 992–998.

Shapiro, S., Rosenberg, L., Slone, D., Kaufman, DavidW., Stolley, PaulD., & Miettinen, OlliS. (1979). Oral-contraceptive use in relation to myocardial infarction. The Lancet, 313(8119), 743–747. https://doi.org/10.1016/S0140-6736(79)91205-4

Szklo, M., & Nieto, F. J. (2019). Epidemio



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