# Errores comunes al interpretar una regresión

En realidad, ver una correlación entre su predictor y su variable objetivo, y de una regresión lineal, no significa realmente que exista una relación causal entre ambas (aunque su regresión puede funcionar muy bien, e incluso de forma óptima).

Aunque utilizar un enfoque de ciencia de datos, en lugar de uno estadístico, garantizará una cierta eficacia en su modelo, es fácil caer en algunos errores al no tener idea de por qué su variable objetivo está correlacionada con un predictor.

Le hablaremos de seis razones diferentes y le ofreceremos una advertencia para ayudarle a manejar esos predictores sin dificultad. En lo subsecuente, **x** denotará un predictor y **y** la respuesta.

## Causalidad directa

x causa y; por ejemplo, en el negocio inmobiliario el valor es directamente proporcional al tamaño de la casa en metros cuadrados.

## Efectos recíprocos

x causa y, pero también está influida por y. Esto es bastante típico de muchas dinámicas macroeconómicas en las que el efecto de una política aumenta o disminuye sus efectos. Por ejemplo, en el sector inmobiliario, los altos índices de delincuencia en una zona pueden reducir sus precios, pero unos precios más bajos significan que la zona podría convertirse rápidamente en una zona aún más degradada y peligrosa.

## Causalidad espuria

Esto ocurre cuando la causa real es la variable z, que causa tanto a la variable x como a la variable y; en consecuencia, es solo una ilusión falaz que x implica a y porque es z detrás de la escena. Por ejemplo, la presencia de tiendas y galerías de arte caras de arte pueden parecer correlacionadas con los precios de la vivienda; en realidad, ambas están determinadas por la presencia de residentes acomodados.

## Causalidad indirecta

En realidad, x no está causando y, sino que está causando otra cosa, que a su vez causa y. Un buen municipio que invierte en infraestructuras tras subir impuestos puede afectar indirectamente al precio de la vivienda porque la zona se vuelve más cómoda para vivir, atrayendo así más demanda. Los impuestos más altos, y por tanto más inversiones afectan indirectamente al precio de la vivienda.

## Efecto condicional

x causa y respecto a los valores de otra variable z; por ejemplo, cuando z tiene determinados valores x no influye en y, pero, cuando z toma determinados valores, la x empieza a impactar en y. A esta situación también la llamamos interacción. Por ejemplo, la presencia de escuelas en una zona puede convertirse en un factor de atracción cuando el índice de delincuencia es bajo, de modo que afecta al precio de la vivienda solo cuando hay poca delincuencia.

## Efecto aleatorio

Cualquier correlación registrada entre x e y se ha debido a una selección de muestreo afortunada; en realidad no hay ninguna relación con y.

# Observaciones relevantes

El caso ideal es cuando se tiene una causalidad directa; entonces, tendrá un predictor en su modelo que siempre le proporcionará los mejores valores para derivar sus respuestas.

En los demás casos, es probable que la relación causa-efecto imperfecta con la variable de respuesta dé lugar a estimaciones más ruidosas, sobre todo en la producción, cuando tendrá que trabajar con datos no vistos antes por el modelo.

Los efectos recíprocos son más típicos de los modelos econométricos. Requieren tipos especiales de análisis de regresión. Incluirlos en su análisis de regresión puede mejorar su modelo. Sin embargo, su papel puede subestimarse.

Las causas espurias e indirectas añadirán algo de ruido a su relación x e y; esto podría dar lugar a estimaciones más ruidosas (errores estándar más grandes). A menudo, la solución es obtener más observaciones para su análisis.

Los efectos condicionales, si no se detectan, pueden limitar la capacidad de su modelo para producir estimaciones precisas. Si no es consciente de ninguno de ellos, dado su conocimiento del dominio del problema, es un buen paso comprobar la existencia de alguno de ellos mediante algún procedimiento automático para comprobar posibles interacciones entre las variables.

Los efectos aleatorios son lo peor que le puede pasar a su modelo, pero se pueden evitar fácilmente si se sigue un procedimiento adecuado de limpieza de datos.

# Referencias

Massaron, L., & Boschetti, A. (2016). *Regression Analysis with Python.* Birmingham, UK: Packt Publishing.