

INTRODUCCIÓN A LA TERCERA PARTE: UNA CLASIFICACIÓN DE LOS MÉTODOS DEL ANÁLISIS MULTIVARIADO

En un sentido amplio, el análisis multivariado designa un conjunto de métodos de análisis que tratan más de una variable al mismo tiempo. En particular, se requiere del análisis multivariado para:

- Medir el grado de asociación entre dos o varias variables.
- Estimar los parámetros de una relación entre dos o varias variables.
- Evaluar hasta qué punto las diferencias entre dos o varios grupos de observaciones son significativas.
- Intentar predecir a cuál grupo pertenece un individuo a partir de sus demás características.
- Buscar discernir una estructura en un conjunto de datos.

Varias técnicas de análisis multivariado necesitan distinguir entre las variables dependientes y las variables independientes. Las variables dependientes son aquellas cuyo valor se quiere predecir; se conocen las demás como las variables independientes.¹³⁵ Es posible clasificar los métodos de análisis

¹³⁵ Se encontra abajo una discusión de los términos “variable dependiente” y “variable independiente”.

sis en función, por una parte, del número de variables dependientes e independientes y, por otra parte, de que sean, las unas o las otras, variables discretas o continuas.¹³⁶

La tabla que sigue presenta una clasificación de algunos métodos de análisis multivariado.

Variable dependiente		Variables independientes	Método	
Ninguna		2 variables categóricas	Análisis de tabla de contingencia	... con 2 dimensiones
		Más de 2 variables categóricas		... con más de 2 dimensiones
Continua		Discretas (categóricas)	Análisis de varianza o Regresión múltiple	
		Continuas y/o discretas	Regresión múltiple	
Categórica	2 categorías	Continuas y/o discretas	Logit o probit	... binomial
	Más de 2 categorías			... multinomial

Esta parte de la obra trata del análisis de regresión. El análisis de regresión es un método de análisis de los datos que se aplica cuando nos basamos en un modelo teórico formalizado con una relación entre una variable dependiente continua y una o varias variables independientes continuas o discretas.

¹³⁶ Se infiere esto de la escala de mediciones de cada variable (ver cap. 1-1): las variables categóricas son discretas mientras que las variables racionales y las variables de intervalo se consideran, comúnmente, como continuas. En cuanto a las variables ordinales, pocos modelos son, de manera específica, idóneos para su tratamiento; en la práctica, se tratan como si fueran continuas pero, al momento de interpretar los resultados, es importante tomar en cuenta el hecho de que son variables ordinales.

La regresión es lineal siempre y cuando la forma funcional de la relación sea lineal.¹³⁷

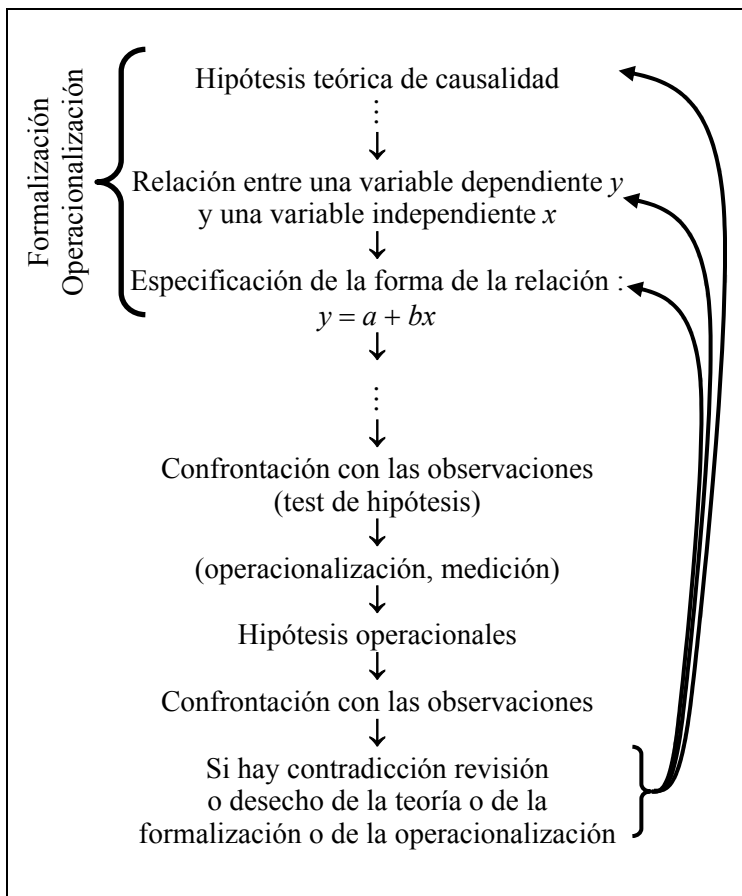
Antes de empezar de pleno el estudio del análisis de regresión, examinemos con más detenimiento la distinción entre las variables dependientes e independientes, particularmente desde el punto de vista del examen de las relaciones de causalidad. Los términos “variable dependiente” y “variable independiente” provienen de las ciencias experimentales cuando el investigador fija de manera “independiente” el valor de algunas variables (por ejemplo, la dosificación de un tratamiento) con el objetivo de observar luego el efecto sobre la variable “dependiente”. En un modelo con una única ecuación, la variable “dependiente” es conocida también como “endógena”, es decir que se determina en el interior del modelo, cuando las variables independientes son exógenas puesto que se determinan en el exterior del modelo. Estas mismas variables independientes son denominadas, también, “estímulos”; en este caso, las variables dependientes son “respuestas”. En inglés, se emplean los términos *predictor/criterion*, *stimulus/response*, *task/performance*, *input/output*.

Esta multitud de términos es sintomática de los numerosos significados conceptuales o teóricos que puede representar la relación entre una variable dependiente y una variable independiente. Las variables independientes son, a veces, calificadas de “explicativas”. Antes bien, esta expresión requiere emplearse con prudencia por la fuerte connotación de causalidad que implica. Es posible que la relación sea puramente estadística, de tal manera que la variable independiente pueda permitir “predecir” el valor de variable dependiente sin “explicar” este valor.

¹³⁷ Cuando no es posible traducir un modelo teórico con una relación lineal, hay que usar la regresión no lineal a la cual se aplica el método de estimación del máximo de verosimilitud.

Por ejemplo, sería posible observar una relación inversa entre la variable dependiente “variación de las ventas de menudeo de las tiendas de regalos con relación al mes anterior” y la variable “variación de la temperatura promedio con relación al mes anterior”. Parece más que evidente que no es el frío el que incita a la compra de regalos sino, más bien, la llegada de la Navidad. En nuestro hemisferio, está la casualidad que este acontecimiento coincide con la llegada del invierno, pero en Australia, pasa lo contrario. Ahora bien, existe efectivamente una relación causal entre la tendencia a vestir ropa más caliente (variable dependiente) y la tendencia a la baja estacional de la temperatura (variable independiente). Lo que nos permite distinguir una relación causal (la segunda) de una relación no causal, es el modelo teórico que poseemos del fenómeno.

En particular, esto implica que los tests de hipótesis que podríamos aplicar a la relación entre variable dependiente y variable independiente no serían aptos para demostrar una relación de causalidad. Lo más que estos tests de hipótesis nos permitirían, sería constatar si la hipótesis de causalidad es o no rechazada por las observaciones. Ilustramos este punto con el diagrama que presentamos a continuación, el cual puede compararse con el diagrama del método hipotético-deductivo del capítulo 2-2.



Para resumir, la relación entre una variable dependiente y una variable independiente no es necesariamente una relación causal. La interpretación que hagamos de esta relación, el significado que le queramos dar dependen del modelo teórico que sirve de punto de partida. Además de la relación causal, es posible distinguir, en orden decreciente del contenido teórico:

- Modelos de simulación, algunas veces llamados modelos de previsión condicional, los cuales se apoyan en un modelo teórico que representa el funcionamiento del fenómeno¹³⁸ y que se conciben para contestar a las preguntas del tipo “¿qué pasaría (o qué hubiera pasado) si...?”
- Modelos de proyección que buscan contestar a la pregunta “¿qué va a pasar si la tendencia se mantiene?” o, a partir de un modelo más desarrollado del fenómeno, “¿qué va a pasar si los parámetros de las relaciones entre las variables siguen idénticas?”
- Modelos de previsión que, de manera puramente pragmática, buscan contestar a la pregunta “¿qué va a pasar?” y que, en función de este objetivo, pueden, con todo derecho, explotar relaciones de simple asociación estadística.

¹³⁸ El cómo más que el porqué de los modelos de relación causal.