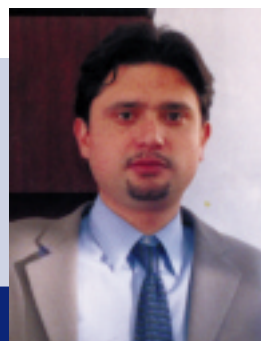


Modelos de Ecuaciones Estructurales (una primera aproximación)



Iván Castro Rivadeneyra

BIMSA-Ipsos

A continuación se presenta un acercamiento a lo que se conoce como Modelos de Ecuaciones Estructurales (M.E.E.) o modelos LISREL (del inglés *Linear Structural RELationships*). Como motivación al tema se puede señalar que, a pesar de ser una herramienta envuelta en un cierto halo de misterio, el número de sus aplicaciones en la investigación de mercados ha crecido de manera importante en años recientes. Con la intención de hacer un recorrido por la lógica detrás de la técnica, se desarrolla en primer lugar de forma sucinta el concepto de causalidad y modelación. Más adelante se identifican las similitudes, así como diferencias que guardan los M.E.E. con el análisis de regresión y el análisis de factores. Después, se presenta propiamente el concepto estadístico, apoyándose de una visión global de las dos técnicas anteriores. En el trayecto se incluyen varios ejemplos que ayudan a dimensionar su uso y se revisan también sus limitaciones.

Es importante advertir que los aspectos técnicos de la herramienta no se consideran en este momento. De este modo, se desea dar oportunidad a que las ideas principales logren su espacio en el “*share of mind*” de conceptos estadísticos de los usuarios de la investigación de mercados y constituya éste el objetivo central del documento.

Finalmente, es oportuno señalar que los ejemplos incluidos en el presente material deben considerarse sólo ilustrativos y no pretenden en lo absoluto ser modelos probados.

1 Una Posible Causa Casual

El propósito de los procedimientos estadísticos es ayudar a establecer la factibilidad de modelos teóricos y estimar en qué grado las diferentes variables explicativas parecen influir en las variables dependientes.

Cooley W.

Esta cita rescata de manera ejemplar la esencia de la modelación, sin embargo, en la práctica existe un

nivel de complejidad importante al:

- Definir un modelo de la realidad.
- Probarlo a partir de datos provenientes de observaciones no experimentales.
- Concluir causalidad a partir de correlación estadística.

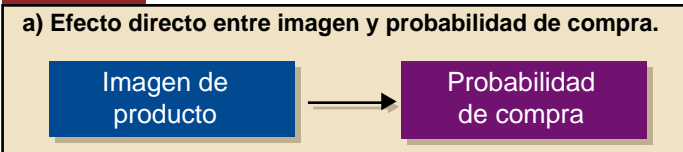
Para este momento ya hace falta un ejemplo. ¿Por qué los consumidores seleccionan productos determinados? ¿Cuáles son los factores que influyen en esa decisión? ¿Cómo interactúan estos factores entre ellos mismos y en qué medida participa cada uno en la selección final de compra? Al respecto se tienen teorías que enumeran variables y establecen relaciones entre ellas. El investigador cuenta con datos provenientes de encuestas en las que se tuvo cuidado de incluir pre-guntas relacionadas con los factores hipotéticos de influencia. Lo que resta, entonces, es probar si el modelo efectivamente trabaja.

En el espacio extraterrenal donde habitan las causas, éstas se ligan a sus efectos por medio de relaciones que se interpretan como correlaciones¹. Es decir, “si una variable es causa de otra entonces estarán correlacionadas”, pero el otro sentido de esta sentencia lógica no es necesariamente cierto, esto es: “si dos variables están correlacionadas entonces una es causa de otra”. Lo anterior da lugar inmediatamente a aceptar que existen variables correlacionadas que no están asociadas a un esquema de causa y efecto. En este tipo de situaciones se dice que hay correlación espuria o no causal.

¿Cómo identificar entonces una causa real de una ficticia usando correlación estadística? La respuesta es: dado que sólo contamos con correlaciones para avalar modelos de causalidad, se debe tener cuidado en interpretar la correlación, separando en lo posible la parte de naturaleza plenamente “causal” (por favor notar las comillas) de la espuria. Lo anterior será factible en la medida en que existan modelos coherentes con la realidad.

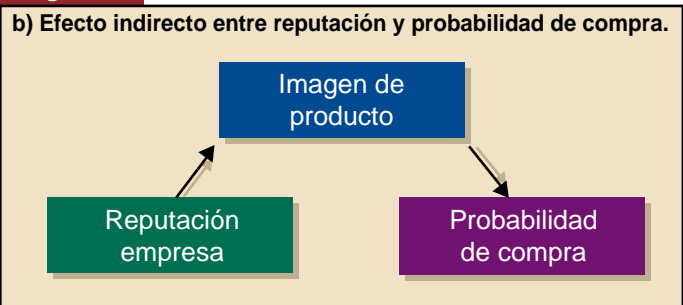
Ahora, al tratar de entender aún más la parte causal, ésta se puede dividir a su vez en efectos directos e indirectos. Ejemplo (diagrama 1) :

Diagrama 1



El diagrama de rutas anterior, muestra una relación de causa y efecto directa entre la imagen de un producto y la posibilidad de ser adquirido (la flecha indica la causa y el efecto).

Diagrama 2

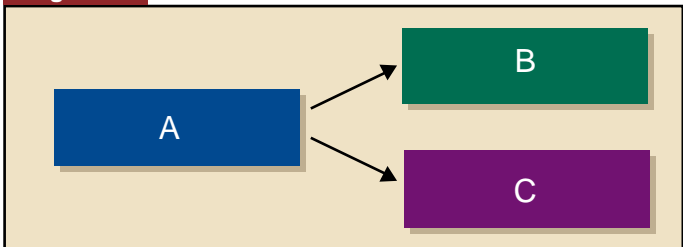


Se podría hipotetizar(ver diagrama 2), que la reputación de la empresa tiene un efecto indirecto sobre la probabilidad de compra por medio de la imagen del producto. La reputación afecta la imagen y la imagen la probabilidad. La forma matemática de representar los efectos indirectos es la multiplicación de los efectos directos involucrados.

Por otra parte, la correlación no causal o espuria también se divide en:

(a) Aquella que se debe a la presencia de antecedentes comunes. El diagrama de correlación espuria para este caso entre B y C es del tipo (diagrama 3) .

Diagrama 3



Es decir, la correlación que registren B y C se deberá al hecho de que comparten a A como causa y no a que B sea causa de C o viceversa.

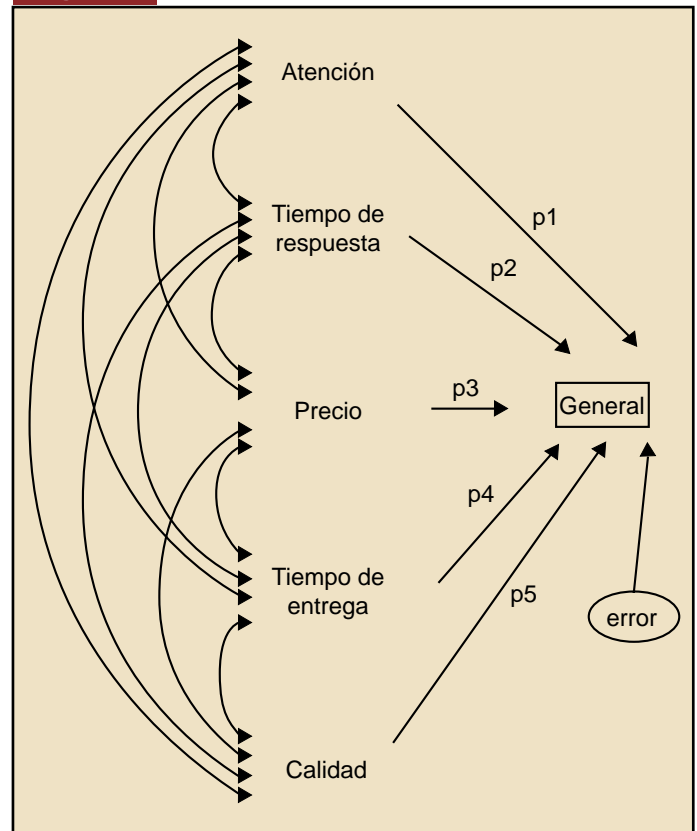
(b) Y en segundo lugar, en la debida a asociaciones previas no incluidas en el modelo.

Sumo cuidado exige entonces derivar conclusiones de causalidad cuando se tiene como herramienta la correlación, ya que podría llevar a resultados falsos. Pero entonces ¿qué alternativas sensatas de análisis hay?

2 ¿Qué Pesa Más? : Análisis Exploratorio con Regresión

Evidentemente, no basta con establecer una posible relación causa efecto, sino que también se debe determinar su fuerza. Al respecto, los estudios de satisfacción proporcionan un buen ejemplo. Los niveles de satisfacción para diferentes rubros se podrían pensar relacionados con la satisfacción general de la siguiente manera (diagrama 4).

Diagrama 4



Esto es, el modelo establece que el grado de satisfacción en la atención recibida; en el tiempo de respuesta; en el precio; en el tiempo de entrega y por último en la percepción de calidad, condicionan la evaluación general de satisfacción en la compra de un bien o servicio. La diferencia entre lo que explica el modelo y los datos observados se representa por la variable de error añadida en el diagrama. La pregunta luego es: ¿cuál es el rubro más importante en términos de su influencia sobre la satisfacción global (es decir el valor p o peso más grande)? ¿Y el siguiente? ¿Y el siguiente?...

Un análisis de regresión² proporciona elementos valiosos para hacer inferencia sobre los pesos de estas influencias. Sin embargo, en pocas ocasiones se consideran las posibles correlaciones (indicadas en el diagrama 4 por líneas curvas de doble sentido) que existen entre las variables independientes (en este caso niveles de satisfacción por rubros). Lo anterior podría originar que los coeficientes p estimados sean para este ejemplo incluso negativos. Es decir se tendría que concluir que: ¡el nivel de satisfacción de los rubros con coeficientes negativos influyen de manera contraria a la satisfacción global!

En resumen, la técnica de regresión en su versión explicativa (no predictiva), evalúa la fuerza de las relaciones que mantienen de manera conjunta un grupo de variables independientes con respecto a una variable dependiente. La lectura de las relaciones se basa en los coeficientes de regresión estimados (valores p en el diagrama 4) pero que pueden estar afectados por una posible correlación ignorada entre las variables explicativas.

Se debe recordar, además, que una de las restricciones del análisis de regresión es que la variable dependiente (y en general las independientes) debe estar medida en escalas intervalares³.

Entonces ¿qué sigue?

3 Un Mundo Complejo a la Mano: Análisis de Factores.

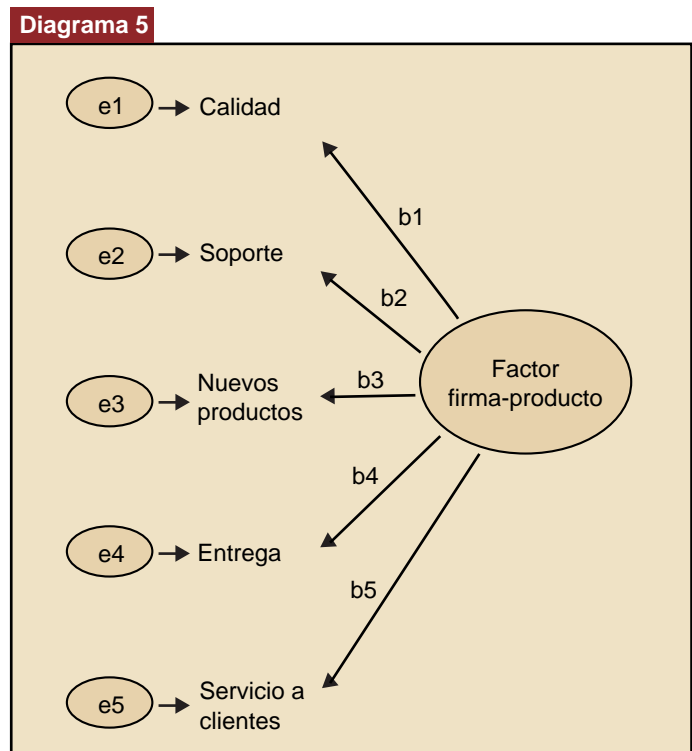
Una manera de atacar el problema de correlación no deseada entre variables es llevando a cabo, de manera previa al análisis de regresión, un análisis de factores exploratorio. Esta técnica buscará conformar objetos abstractos o *variables latentes* que

agrupen o representen a un conjunto de variables originales (en general *variables manifiestas*) que se encuentren correlacionadas.

Dos son las grandes ganancias de la aplicación de análisis de factores:

- Reducción de dimensionalidad. Esto significa la posibilidad de trabajar con un número relativamente manejable de objetos abstractos, los cuales están articulados por las variables originales y que en lo posible reflejan la mayor parte de su variabilidad.
- Ortogonalidad de las variables latentes. La herramienta permite la definición de objetos abstractos que cumplan con el requisito de estar no correlacionadas.

Por ejemplo, un clásico estudio de percepción, podría incluir una serie de evaluaciones al respecto de diferentes atributos posiblemente correlacionados, como son: la calidad del producto; el soporte técnico; el desarrollo de nuevos productos; la confianza en la entrega y el servicio a clientes. Un factor o variable latente que podría llamarse “factor percepción firma-producto” se puede esquematizar para el conjunto de variables originales como (diagrama 5).



Donde los parámetros b representan las cargas factoriales que asocian el peso de las variables manifiestas con la variable latente, la cual resulta entonces, una representante abstracta de las cinco dimensiones iniciales. Observar que la variable latente se representa por elipses y no por cajas que se usan para variables manifiestas. Las variables e son errores de medición no explicables por el modelo.

Generalmente, se parte de una lista importante de variables originales y se genera un número adecuado de factores. La diferencia entre el análisis de factores tradicional y el que se relaciona a los M.E.E. (que se llama *análisis de factores confirmatorio*) es que: en el primer caso no existen restricciones *a priori* entre las cargas factoriales de las variables manifiestas y los factores, los que una vez obtenidos, se tienen que nombrar. Por otra parte, el análisis factorial confirmatorio parte de la definición *a priori* de las variables latentes y de las relaciones que guardan con las variables manifiestas. Entonces, el análisis factorial exploratorio agrupa a las variables originales más correlacionadas y el confirmatorio espera que el modelo lo establezca para ser corroborado. Por último, en ambos casos, se requiere que las variables estén medidas también en escalas al menos de intervalo.

Recapitulando, el análisis de regresión ofrece una manera de establecer los pesos que tienen “las” variables “causales” sobre un efecto. El análisis factorial permite controlar la correlación que existe entre variables manifiestas y generar pocos representantes llamados variables latentes sobre los que se puede probar la relación hipotética de causalidad usando regresión.

Con esto en mente, se está a un paso de una posible mirada lógica detrás de los modelos de ecuaciones estructurales.

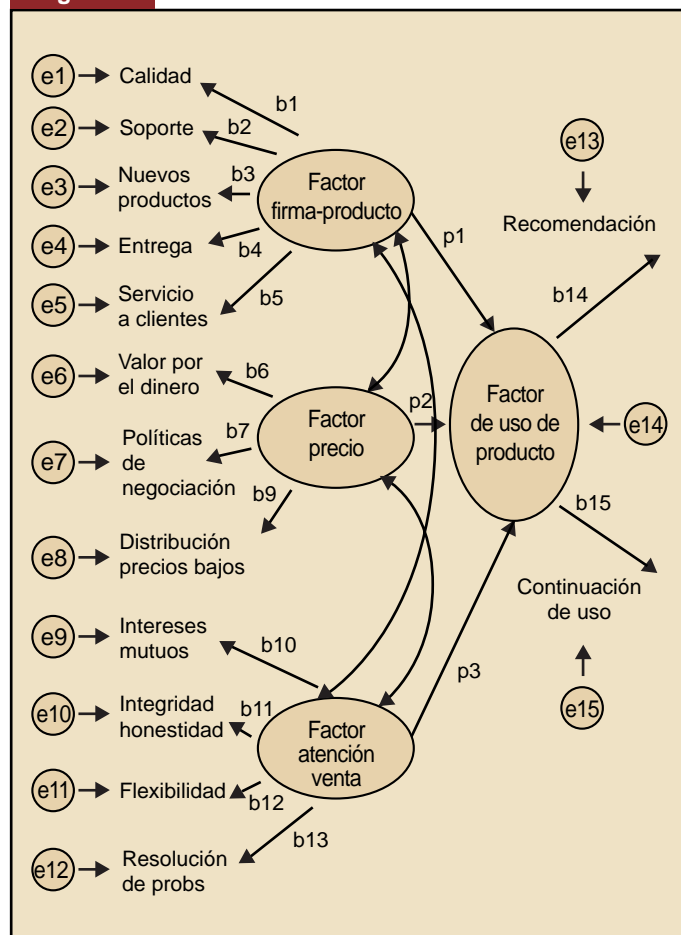
4 Poniendo Todo Junto y Algo Más

Revisar el siguiente diagrama de rutas de un ejemplo de M.E.E. (diagrama 6).

Esto es, tres variables latentes, la primera referida a la percepción firma-producto, la segunda a la percepción de precio y la tercera a la percepción de atención de venta influyen sobre un cuarto factor denominado “uso de producto”. Las tres primeras varia-

bles latentes se nombran exógenas (causas) y la cuarta se identifica como endógena (efecto).

Diagrama 6



Las magnitudes de influencia de las variables exógenas sobre la endógena se miden por los parámetros p . Recordar que el modelo total es una propuesta de comportamiento para la realidad y para estar completamente definido se consideran las posibles correlaciones entre las variables exógenas. Lo anterior no es inconsistente con el requerimiento de no correlación entre factores que se planteaba con anterioridad. En los M.E.E., si existe evidencia de correlación entre variables exógenas, se puede incluir dicha hipótesis en el modelo. Estas correlaciones se representan por medio de las líneas curvas de doble sentido. A la parte del modelo donde se establecen todas las relaciones entre las variables latentes se le conoce como el *modelo estructural*.

Por otro lado, cada una de las variables latentes se pone en operación por medio de la parte del modelo que se denomina *modelo de medición* y en el que se especifican las variables manifiestas y su relación con la parte del modelo estructural. Por ejemplo la variable latente “factor de uso de producto” representa a dos variables manifiestas que son continuidad de uso y recomendación de producto. Los parámetros b en el diagrama 6 indican la importancia que los factores tienen sobre sus correspondientes variables originales. Las variables manifiestas de las latentes exógenas son manifiestas independientes. Para las variables latentes endógenas, sus correspondientes variables manifiestas son dependientes.

De esta forma queda especificado un modelo de ecuación estructural.

5 El Procedimiento

A continuación se narra el procedimiento general para la aplicación de los M.E.E.

5.1 Conceptualización. Se refiere a la etapa en la que se establecen las ligas o relaciones entre las variables latentes, reflejándose así la parte substancial de la hipótesis de interés. Asimismo, en este paso se deberá indicar la manera en cómo las variables manifiestas ponen en operación a las latentes. Es oportuno señalar que las estructuras de los M.E.E. en la práctica son muy versátiles y no necesariamente se ajustan a la estructura del ejemplo usado en la sección anterior.

5.2 Diagrama de rutas. Es la representación gráfica del modelo conceptualizado. Es una herramienta útil que ayuda a la mejor comprensión del modelo.

5.3 Especificación. En esta etapa el modelo se traduce a su forma matemática para su procesamiento en la computadora.

5.4 Identificación del modelo. A este nivel se investiga si la información con la que se cuenta de manera empírica (datos) es suficiente para proveer de una solución al modelo. En ocasiones, para lograr estimar la mayoría de parámetros, son necesarias algunas restricciones extras sobre ellos.

5.5 Estimación de parámetros. Es una etapa muy

importante y en la que se usan criterios estadísticos para seleccionar los parámetros (las p 's, las b 's y otros) que mejor ajustan a los datos. Esta selección la llevan a cabo los programas de computación de manera conjunta para todos los parámetros involucrados siguiendo algoritmos de optimización numérica.

5.6 Evaluación del ajuste. Son pruebas que ayudan a verificar la congruencia entre las correlaciones estimadas y las observadas. Es decir, se prueba si el modelo es congruente con los datos.

5.7 Modificación del modelo. Es el paso donde el análisis deja de ser confirmatorio y se convierte en exploratorio al probar cambios sobre la definición del modelo que ayuden a un mejor ajuste.

5.8 Validación. Para un conjunto nuevo de datos se verifica si el modelo resulta consistente.

5.9 Programas de computación para M.E.E. LISREL es el más famoso y que origina un sinónimo de los M.E.E. Existen otros con módulos gráficos muy amigables como AMOS y EQS.

6 ¿Hasta Dónde?

Algunas de las críticas al uso de los M.E.E. se refieren a la fase interpretativa en los siguientes aspectos:

- a) *Los datos nunca confirman un modelo.* En el mejor de los casos es imposible descalificarlo, por lo que puede existir más de un modelo aceptable para el mismo conjunto de datos.
- b) *La secuencia temporal de variables no es una guía para inferir causalidad.* En ocasiones las variables recolectadas en el tiempo se piensan en el marco de causa-efecto, cuando no necesariamente es cierto.
- c) *Falacia nominalística.* Darle un nombre a alguna variable latente no la hace necesariamente igual al concepto que se nombra. Esto podría llevar a interpretaciones erróneas.
- d) *Los M.E.E. son una herramienta confirmatoria no de desarrollo de modelos.* Esta observación puede verse como desventaja ante otras técnicas exploratorias.

- e) *Violaciones a supuestos distribucionales de los datos.* Tiene que ver con hipótesis de comportamiento de los datos que pocas veces son confirmadas (como normalidad) y sin embargo las técnicas se aplican.
- f) *El ajuste ante todo.* Buscar un buen ajuste del modelo general puede llevar a rechazar modelos que especifican relaciones importantes individuales entre las variables.
- g) Finalmente: *¡Cuidado!, correlación no implica causalidad.*

7 Conclusiones

Se ha presentado una técnica que reúne conceptos tradicionales de análisis multivariado (regresión y factores) para llevar a cabo un análisis confirmatorio de modelos en los que se establecen posibles relaciones de causalidad entre variables. Sin embargo, a pesar de que resulta una herramienta útil y completa, su éxito recae en el planteamiento de hipótesis o modelos congruentes en lo posible con la realidad.

Recordar que el presente es sólo un acercamiento lateral y en cámara lenta al concepto general de modelos de ecuaciones estructurales pero que se espera resulte valioso para su comprensión y difusión.

8 Bibliografía

Hooley Graham J. & Michel K. Hussey. *Quantitative*

Methods in Marketing. Thomson Business Press (1999).

Maruyama Geoffrey M. *Basics of Structural Equation Modeling.* Sage Publications (1998).

Hair Joseph F. Et al. *Multivariate Data Analysis.* Prentice Hall (1998).

Arbuckle James L. & Werner Wothke. *Amos 4.0 User's Guide.* SmallWaters Corporation (1999).

Hoyle Rick H. *Structural Equation Modeling.* Sage Publications (1995).

Johnson Richard A. & Dean W. Wichern. *Applied Multivariate Statistical Analysis.* Prentice Hall (1999).

9 Notas

¹ La medida estadística de correlación es cercana a uno cuando valores grandes de una variable indican valores grandes de la otra (peso y estatura), cercana a menos uno cuando valores grandes de una se asocian a valores pequeños de la otra (satisfacción y cambio de trabajo) y próxima a cero cuando no hay relación (inteligencia y estatura).

² El análisis de regresión busca, bajo la hipótesis de una relación lineal entre las "causas" y el efecto, los pesos p que minimizan la diferencia entre los valores estimados y observados de la variable dependiente o efecto.

³ Las escalas intervalares son aquellas en las cuales las distancias entre los niveles de la escala tienen sentido (por ejemplo metros, kilos, conteos, IQ, etcétera).

**Si desea continuar
recibiendo**



**Datos,
Diagnósticos y
Tendencias**

¡Envíenos sus datos!