

"Técnicas Avanzadas de Investigación de Mercados: Aplicaciones de los Modelos Multinivel en Estudios Longitudinales con R"

Lily (Xuehui) Gao
Universidad de Zaragoza



Lily (Xuehui) Gao, doctora de la Universidad de Zaragoza. Es miembro del grupo de investigación Generés reconocido por el Gobierno de Aragón. Sus intereses de investigación se centran en la gestión de la experiencia del consumidor, de estrategias multicanal, así como en el análisis de las consecuencias que se derivan de las experiencias cliente-empresa. Ha publicado artículos de investigación en revistas académicas como Journal of Service Research, Journal of Business Research, Electronic Commerce Research and Applications y The Service Industries Journal. Ha asistido y presentado trabajos de investigación en varias conferencias internacionales, como AEMARK, CMC o EMAC.

Los modelos multinivel o de estructuras jerárquicas constituyen la metodología de análisis más adecuada para tratar datos “jerarquizados” o “anidados”.

Por ejemplo, en salud podemos considerar los pacientes agrupados por médicos, los médicos en centros de atención primaria y éstos a su vez en áreas de salud. Del mismo modo, los estudiantes se pueden considerar agrupados en aulas, las aulas en escuelas y las escuelas en zonas de una ciudad. De este modo, el análisis multinivel permite analizar la influencia de una serie de variables independientes sobre la variable dependiente cuando los casos se pueden agrupar en unidades de información de diferentes niveles (paciente/estudiante, médico/aula, centro salud/escuela...) y las medidas se pueden tomar tanto en los niveles más bajos (los individuos) como en los niveles superiores en que pueden agruparse los individuos. De este modo, además de mejorar la calidad de los resultados, el análisis multinivel posibilita realizar análisis novedosos, tales como estimar la aportación de cada nivel de análisis o las interacciones entre variables de distintos niveles.

Tomando como ejemplo el caso de las escuelas, cuando los datos están jerarquizados tradicionalmente se han empleado dos alternativas:

- 1) Considerar que la unidad de análisis es el estudiante individual, con lo cual se recogían los datos de cada sujeto de manera independiente. Con ello, la base de datos estaría formada por un conjunto amplio de sujetos de los cuales tenemos una serie de variables, entre las que se pueden medir los datos referidos al aula, por ejemplo: las características del profesor y las del centro donde cursa sus estudios. En este caso se puede caer en la llamada falacia atomística, por la cual se atribuyen las diferencias en las variables de los sujetos a las aulas o los centros. Por añadidura, los modelos de regresión lineales descansan en el supuesto de independencia de las observaciones y, como compartir el mismo contexto causa su dependencia, los errores estándar estimados de las pruebas estadísticas tradicionales aparecerán claramente subestimados y ello conducirá irremisiblemente a que la mayoría de los resultados sean significativamente espurios.
- 2) La otra posibilidad es que la unidad de análisis sea el centro o el aula. En ese caso los datos de los alumnos se agrupan (normalmente promediándolos) y se incluyen en los datos de cada aula o centro. En este caso, se pierde una gran cantidad de información, con lo que disminuye la potencia del análisis estadístico. Además, se puede caer en la llamada falacia ecológica por la cual se otorgan incorrectamente las características del contexto a los sujetos.

Los modelos multinivel trabajan con varias unidades de análisis de forma simultánea. Proponen una estructura de análisis dentro de la cual se pueden reconocer los distintos niveles en que se articulan los datos, pues cada subnivel está representado por su propio modelo. Con ello, los modelos multinivel respetan la organización jerárquica que presentan los datos de forma natural. Cada uno de estos submodelos expresa la relación entre las variables dentro de un determinado nivel y especifica cómo las variables de ese nivel influyen en las relaciones que se establecen en otros niveles.

Los modelos multinivel son, en esencia, ampliaciones de los modelos de regresión lineal clásicos; ampliaciones mediante las cuales se elaboran varios modelos de regresión para cada nivel de análisis. Con ello los modelos del primer nivel están relacionados por un modelo de segundo nivel en el que los coeficientes de regresión del nivel 1 se regresan en un segundo nivel de variables explicativas, y así sucesivamente para los diferentes niveles.

Pero antes de profundizar mínimamente en el desarrollo formal de los modelos multinivel vamos a prestar atención a tres conceptos fundamentales y sus implicaciones: correlación intraclase, coeficiente fijo y aleatorio, e interacción internivel.

Se entiende por correlación intraclase o autocorrelación la medida del grado de dependencia de los individuos. Es decir, es una estimación de lo que comparten los alumnos por estudiar en una misma clase o centro. Una correlación baja o cercana a cero significará que los sujetos dentro del mismo grupo son tan diferentes entre sí como los que pertenecen a otros grupos. En ese caso, la agrupación no tiene consecuencias, los grupos no son homogéneos internamente y las observaciones son independientes (requisito necesario dentro de los modelos lineales tradicionales). Si se ignora la presencia de esta correlación intraclase, los modelos resultantes son innecesaria y falsamente complejos, dado que aparecen relaciones significativas inexistentes.

Otro concepto fundamental, y que supone la gran aportación de los modelos multinivel, es el de coeficiente fijo y coeficiente aleatorio. En los modelos de regresión clásicos los parámetros que se estiman son el intercepto (o punto de corte) y las pendientes. Desde una perspectiva clásica, estos coeficientes se asumen como fijos, es decir, comunes a todos los sujetos y son estimados a partir de los datos. Los coeficientes aleatorios, sin embargo, son variables y se distribuyen según una función de probabilidad. En una estructura multinivel los coeficientes del primer nivel (alumnos) son tratados como aleatorios en el segundo nivel (centros o aulas). En los modelos multinivel están compuestos por dos partes, una general, común a todos los contextos, que es la llamada parte fija, y otra que representa lo específico de cada contexto, que varía y que se estima a través de la varianza en los distintos niveles.

Un tercer concepto importante es la interacción internivel o la interacción entre variables que están medidas en diferentes niveles de una estructura jerárquica de datos. Ello hace referencia a la interacción que puede haber entre variables de diferentes niveles; por ejemplo, determinada metodología docente puede ser mejor con ciertos estudiantes (el llamado efecto Aptitude Treatment Interaction-ATI), o un estilo directivo con profesores de determinadas características. La comprobación de este tipo de hipótesis necesita un modelo de análisis que no sólo dé cuenta de la estructura jerárquica de los datos, sino que también permita estimar las interacciones interniveles.

La estimación con R de modelos multinivel permite tratar un gran volumen de datos utilizando un software libre y estimar modelos de gran complejidad.

Bibliografía:

Murillo Torrecilla, F. J. (2008). Los modelos multinivel como herramienta para la investigación educativa.