Castellana, Noelia

Cometo, Esteban

Rapelli, Cecilia

García María del Carmen

Instituto de Investigaciones Teóricas y Aplicadas, de la Escuela de Estadística

Una introducción al problema de covariables que varían en el tiempo en modelos mixtos.

# Resumen

En los estudios longitudinales las unidades experimentales se observan repetidamente en varias ocasiones. Los modelos lineales mixtos permiten analizar este tipo de datos, modelando, por un lado, la evolución de la respuesta promedio en función del tiempo mediante efectos fijos (estructura media) y, por otro lado, la variación entre las respuestas repetidas dentro y entre sujetos por medio del error y los efectos aleatorios (estructura de covariancia). En este tipo de estudios es también frecuente contar con variables independientes que se desean incorporar al análisis. Estas variables pueden ser fijas a lo largo de todo el período (por ejemplo el sexo biológico de una persona) o bien puedan variar a lo largo del seguimiento (por ejemplo el valor de colesterol). Estas covariables que varían con el tiempo son variables independientes que incluyen tanto la variación intra-sujeto y la variación entre-sujetos. Esto significa que, para un individuo en particular, el valor de la covariable cambia a través del tiempo y puede cambiar también entre diferentes individuos. Este tipo de covariables pueden utilizarse para comparar poblaciones, para describir tendencias en el tiempo, y también para describir relaciones dinámicas con la variable respuesta. La relación entre la covariable que varía en el tiempo y la variable respuesta puede estar confundida por valores anteriores y/o posteriores de la covariable y en consecuencia esto puede conducir a inferencias engañosas sobre los parámetros del modelo. En este estudio se presentan diferentes definiciones y tipos de covariables que varían en el tiempo, así como también se discuten diferentes enfoques para incorporar este tipo de covariables en los modelos lineales mixtos.

**Palabras claves:** Datos longitudinales. Modelos lineales mixtos. Covariables que varían en el tiempo.

**Abstract PENDIENTE**

Keywords

# Objetivos

Los objetivos de este trabajo son:

* presentar definiciones y tipos de covariables que varían en el tiempo
* describir diferentes enfoques para incorporar este tipo de covariables en los modelos lineales mixtos.

1. **Introducción**

En los estudios longitudinales las unidades experimentales se observan repetidamente en varias ocasiones. Los modelos lineales mixtos permiten analizar este tipo de datos, modelando, por un lado, la evolución de la respuesta promedio en función del tiempo mediante efectos fijos (estructura media) y, por otro lado, la variación entre las respuestas repetidas dentro y entre sujetos por medio del error y los efectos aleatorios (estructura de covariancia). En este tipo de estudios es también frecuente contar con variables independientes que se desean incorporar al análisis. Estas variables pueden ser fijas a lo largo de todo el período (por ejemplo el sexo biológico de una persona) o bien puedan variar a lo largo del seguimiento (por ejemplo el valor de colesterol). A continuación se presentan diferentes definiciones de covariables y enfoques para incorporarlas en los modelos mixtos.

1. **Notación**

Se considera un estudio longitudinal en donde se observan n individuos a lo largo del tiempo, donde para cada i-ésimo individuo (i=1,2,..n) se cuenta con un conjunto de tiempos de seguimiento discretos u ocasiones de medición j, (j=1, 2….ni). Además, en cada ocasión j se observa la variable respuesta Y y una única variable explicativa X o covariable:

valor de la variable respuesta Y para el i-ésimo individuo en la ocasión j-ésima.

valor de la variable explicativa X para el i-ésimo individuo en la ocasión j-ésima.

Se asume que son simultáneamente medidas. En un análisis de corte transversal se correlacionan directamente, sin embargo, para un análisis longitudinal se debe asumir que existe un orden pre-establecido: , **,**  , , .

1. **Definiciones de covariables**

***Covariables que varían con el tiempo y fijas en el tiempo***

En los estudios longitudinales, las variables independientes pueden ser clasificadas en dos categorías: covariables fijas en el tiempo, es decir, que no varían con el tiempo (CNVT) o covariables que varían con el tiempo (CVT). La diferencia entre ellas puede conducir a diferentes enfoques de análisis así como también a diferentes conclusiones.

Las CVNT son variables independientes que no presentan variación intra- sujetos, es decir, los valores de estas covariables no cambian a lo largo del estudio para un individuo en particular. Por ejemplo, el sexo biológico de una persona, el grupo de tratamiento.

Las CVT son variables independientes que incluyen tanto la variación intra-sujeto y la variación entre-sujetos. Esto significa que, para un individuo en particular, el valor de la covariable cambia a través del tiempo y puede cambiar también entre diferentes individuos. Por ejemplo, valor de la presión arterial, condición de fumar (si/no), nivel de polución en el airte.

Tanto las CNVT y las CVT pueden ser utilizadas para realizar comparaciones entre poblaciones y describir diferentes tendencias a lo largo del tiempo. Sólo las CVT permiten describir una relación dinámica entre la covariable y la variable respuesta.

***Covariables estocásticas y no estocáticas***

Las CVT no estocásicas son covariables que varían sistemáticamente a través del tiempo pero son fijas por diseño del estudio o bien su valor puede predecirse. En cambio, las CVT estocásticas son covariables que varían aleatoriamente a través del tiempo es decir, los valores en cualquier ocasión no pueden ser estimados ya que son gobernados por un mecanismo aleatorio. Ejemplos de las primeras son: tiempo desde la visita basal, edad, grupo de tratamiento en los estudios cross-over. Ejemplos de las segundas son: valor del colesterol, ingesta de alcohol (si, no), ingesta de grasas, etc.

***Covariables exógenas y endógenas***

Una CVT estocástica se define como exógena, respecto a la variable respuesta, si el valor de la covariable en un determinado momento es condicionalmente independiente de todos los valores precedentes de la variable respuesta. Formalmente, para el individuo i en la ocasión j-esima:

Donde la es la función de densidad de la variable X (CVT).

Por lo tanto, mientras puede estar asociada con valores previos de la covariable (, no puede estar asociada con valores previos de la variable respuesta (.

En consecuencia: =

Esta definición implica que la respuesta en cualquier momento puede depender de valores previos de la variable respuesta y de la CVT, pero será independiente de todos los demás valores de la covariable. Por ejemplo, en un estudio longitudinal en donde se evalúa si el nivel de polución en el aire está asociado a la función pulmonar, es de esperar que el nivel de polución del aire en una determinada ocasión dependa de los niveles observados previamente, pero no se espera que dependa de los niveles de la función pulmonar observados previamente en un sujeto.

Una CVT que no es exógena se define como CVT endógena o interna. Por ejemplo, cuando se evalúa si la cantidad de actividad física está asociada al nivel de glicemia. El nivel de actividad física en un determinado momento puede estar (o no) asociado a niveles previos y también puede estar asociado a valores previos de glicemia (un paciente con valor de glicemia alto en una visita puede decidir aumentar su nivel de actividad fística para ver se este valor se reduce).

Se puede evaluar en forma empírica si una CVT es exógena ajustando un modelo de regresión en donde se considera como variable respuesta a la covariable en un momento determinado y como variables explicativas tanto a los valores previos de la covariable como a los valores previos de la variable respuesta. Si, después de controlar por los valores previos de la covariable, el valor actual de la covariable muestra una asociación con los valores previos de la variable respuesta, la covariable no es exógena.

1. **Medias condicionales parciales y total**

Cuando se consideran CVT estocásticas se pueden distinguir diferentes medias condicionales de Y dado la información temporal que se considera de la CVT. Si se considera todo el proceso de la CVT, se define la media condicional total como: . En cambio, cuando se considera la media condicional de Y dando un subconjunto de valores de la CVT se la define como media condicional parcial. A continuación se describen diferentes tipos de medias condicionales que pueden resultar de interés en los estudios longitudinales:

* Asociación concurrente: : describe la relación entre la variable respuesta en la ocasión j y la CVT medida en la misma ocasión.
* Asociación resagada t ocasiones: (t >0) : describe la relación entre la variable respuesta en la ocasión j y la CVT medida en t ocasiones previas. En particular, cuando t=1, describe la relación entre la variable respuesta en la ocasión j y la CVT medida en la ocasión anterior.
* Asociación histórica: (t≥1) : describe la relación entre la variable respuesta en la ocasión j y los valores de la CVT previos a la ocasión j.
* Asociación acumulada: : describe la relación entre la variable respuesta en la ocasión j y la exposición acumulada generada por la CVT hasta una ocasión previa determinada (s).

Dependiendo del objetivo de la investigación, la naturaleza de la CVT y de información previa que se conozca sobre el mecanismo de acción de la covariable es necesario identificar cuál o cuáles de las medias condicionales son de interés para luego crear nuevas variables (si fuera necesario) y plantear un modelo adecuado.

1. **Modelo lineal mixto**

El modelo lineal mixto se puede expresar como,

****

siendo, **Y**i={Yij}, i=1,... N, el vector () de medidas repetidas del i-ésimo individuo, **** y  son matrices conocidas de dimensión () y () respectivamente, **b**i es un vector aleatorio de dimensión (), cuyas componentes se denominan efectos aleatorios,  es un vector aleatorio () de mediciones de error (intra-sujeto),  un vector () de parámetros. El número total de observaciones es  y las componentes aleatorias tienen las siguientes distribuciones , .

**Variación intra y entre individuos**

Cuando se consideran CVT, el modelo lineal mixto puede ser ajustado considerando dos componentes que reflejen tanto la variación intra-sujeto y la variación entre-sujetos respecto a la CVT. Por lo tanto, el término del modelo que representa a la covariable se puede descomponer en dos términos:

Donde, representa el promedio de todos los valores observados en el tiempo de la CVT para el individuo i, representa el cambio esperado en la media de la variable respuesta asociado con cambios de la CVT dentro de los sujetos y representa el cambio esperado en la media de la variable respuesta asociado con cambios de la CVT entre sujetos

**Problemas en la estimación**

La estimación de los parámetros de este modelo mediante máxima verosimilitud asume, de manera implícita, que la media condicional de la variable respuesta en la j-ésima ocasión, dados todos los valores de la CVT, depende solamente del valor de la coviariable en esa ocasión. Es decir, (Pepe and Anderson 1994). Este supuesto se cumple tanto con CNVT como con CVT no estocásticas. Sin embargo, para las CVT estocásticas puede no necesariamente cumplirse: valores anteriores o posteriores de la CVT pueden confundir la relación en una determinada ocasión entre . En consecuencia, esto puede conducir a estimaciones sesgadas de . Por lo tanto, para poder incorporar CVT estocásticas hay que verificar que se cumpla esta condición. Cabe destacar que una CVT exógena puede no cumplir esta condición implícita del modelo ya que es más exigente que la propia definición de exogeneidad.

Frente a este escenario, Pepe y Anderson (1994) recomendaron plantear el modelo longitudinal marginal y realizar las estimaciones mediante GEE (ecuaciones de estimación generalizadas) con estructura de correlación independiente. Por otro lado, Laid y Small (2007) propusieron utilizar el “Método generalizado de los momentos“ (GMM) (Hansen, 1982). En este método de estimación se puede incorporar información sobre la naturaleza de la CVT que se está analizando. Laid y Small (2007) definieron 3 tipos de CVT y luego Lalonde et. al (2014) definieron un cuarto tipo de CVT.

1. **Aplicación**

Un programa de atención y control de pacientes hipertensos iniciado en el año 2014 en Rosario realiza un seguimiento exhaustivo de aproximadamente 1000 pacientes. Este programa contempla: efectores no médicos supervisados, tratamiento farmacológico genérico para la hipertensió y utilización de un algoritmo terapéutico sistematizado.. En cada visita se registran tanto características de los pacientes, del tratamiento y de los valores de la tensión arterial. En particular, se desea evaluar si la adherencia al tratamiento farmacológico influye en los valores de la tensión arterial sistólica a lo largo del seguimiento.

Como la variable “adherencia al tratamiento farmacológico” es una CVT estocástica se evaluaran diferentes enfoques para incluirla en un modelo longitudinal que pueda explicar el cambio en la tensión arterial sistólica media a lo largo del tiempo.

Referencias Bibliográficas

Fitzmaurice, G.M., Laird, N.M., Ware, J.H.: Applied Longitudinal Analysis, 2nd edn. Wiley, Hoboken (2012)

Diggle, P.J., Heagerty, P.J., Liang, K.-Y. and Zeger, S.L. (2002). The Analysis of Longitudinal Data, 2nd edition. New York: Oxford University Press.

Hansen, L.P. (1982). Large sample properties of generalized method of moments estimators. Econometrica 50: 1029-1054.

Lai, T.L, Small, D. : marginal regression analysis of longitudinal data with time-dependant covariates: a generalized method of moments approach. J. Roy. Stat.Soc.Ser.B 69(1), 79-99 (2007)

Lalonde, T.L, Wilson, J.R., Yin, J.:Gmm logistic regression models for longitudinal data with time dependant covariates and extended classifications (2014).

Lalonde, T. L. Modeling Time-Dependent Covariates in Longitudinal Data Analyses in D. G. Chen and J. R. Wilson (Ed.), Innovative Statistical Methods for Public Health Data (pp. 57-80). New York, N.Y.: Springer. (2015).

Pepe, M.S. and Anderson, G.L. (1994). A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. Communications in Statistics, Part B - Simulation and Computation 23: 939-951.