



UNIVERSIDAD NACIONAL DE ROSARIO
FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA
SECRETARIA DE CIENCIA Y TECNOLOGIA E INSTITUTOS DE INVESTIGACIONES

Resumen Ampliado

Jornadas Anuales

“Investigaciones en la Facultad”

Ciencias Económicas y Estadística



Castellana, Noelia

Cometo, Esteban

Rapelli, Cecilia

García María del Carmen

Instituto de Investigaciones Teóricas y Aplicadas, Escuela de Estadística

UNA INTRODUCCIÓN AL PROBLEMA DE COVARIABLES QUE VARÍAN EN EL TIEMPO EN MODELOS MIXTOS.

Resumen

Los estudios longitudinales están conformados por datos obtenidos midiendo repetidamente una variable respuesta a la misma unidad. En este tipo de estudios es también frecuente contar con variables explicativas que se desean incorporar al análisis. Estas covariables 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). Los modelos lineales mixtos permiten analizar este tipo de datos, modelando, por un lado, la evolución de la respuesta promedio en función del tiempo y las covariables, mediante efectos fijos, y, por otro lado, la variación entre las respuestas repetidas dentro y entre sujetos por medio del error y los efectos aleatorios, respectivamente. Este tipo de covariables que varían en el tiempo 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. Este estudio realiza una introducción a la problemática de incorporar covariables que varían con el tiempo en modelos para datos longitudinales, presentando diferentes definiciones de las mismas y enfoques metodológicos.

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

Abstract

Longitudinal studies involve the collection of the same response variable repeatedly over time from the same subject/unit. In these types of studies it is also frequent to have explanatory variables that could be incorporated into the analysis. These covariates can be time-independent (for example, the biological sex of a person) or they can vary throughout the follow-up (for example, the cholesterol value) and are denominated as time-varying covariates. Mixed effects linear models allow analyzing this type of data, modeling separately the evolution of the average response (as a function of time and covariates through fixed effects) and the repeated response variation between and within subjects (through the error and random effects). Time-varying covariates can be used to make comparisons across populations, to describe different time trends, and also to describe dynamic relationship with the response variable. The relationship between the time-varying covariate and the response variable may be confused by preceding and/or subsequent covariate values and therefore, misleading inferences about the model parameters can result. This study makes an introduction to the problem of time-varying covariates in longitudinal studies, presenting different definitions and methodological approaches.

Keywords: Longitudinal data, Mixed linear models, time-varying covariates



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, y, por otro lado, la variación entre las respuestas repetidas dentro y entre sujetos por medio del error y los efectos aleatorios, respectivamente. En este tipo de estudios es también frecuente contar con variables explicativas que se desean incorporar al análisis. Estas variables pueden ser fijas a lo largo de todo el período o bien pueden variar a lo largo del seguimiento. La introducción de una covariable que cambia con el tiempo al modelo produce un importante desafío conceptual. Se consideran algunos aspectos de la interpretación de este tipo de covariables, presentando diferentes definiciones y enfoques para incorporarlas en los modelos mixtos.

2. Tipos de covariables

En los estudios longitudinales se pueden distinguir covariables fijas en el tiempo, es decir, que no varían a lo largo del estudio (CNVT) o covariables que varían con el tiempo (CVT). Por ejemplo: el grupo de tratamiento y condición de fumar (si/no) respectivamente. 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.

Las CVT pueden ser estocásticas o no. Las CVT no estocásticas 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. Ejemplos de las segundas son: valor del colesterol, ingesta de alcohol (sí, no).

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. 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. Una CVT que no es exógena se denomina CVT endógena o interna.

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.

3. Modelo lineal mixto

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 representando a cada componente. De esta manera es posible estimar el cambio esperado en la media de la variable respuesta asociado con cambios de la CVT dentro de los sujetos y el cambio esperado en la media de la variable respuesta asociado con cambios de la CVT entre sujetos.



La estimación de los parámetros de este modelo mediante máxima verosimilitud asume, de manera implícita, que la media condicional de la variable respuesta en una determinada ocasión, dados todos los valores de la CVT, depende solamente del valor de la covariable en esa ocasión. 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 entre la variable respuesta y la CVT en una determinada ocasión. En consecuencia, esto puede conducir a estimaciones sesgadas de los efectos fijos del modelo. 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, varios autores recomendaron plantear el modelo longitudinal marginal y realizar las estimaciones mediante GEE (ecuaciones de estimación generalizadas) con estructura de correlación independiente o bien utilizar el "Método generalizado de los momentos"(GMM), donde es posible incorporar información sobre la naturaleza de la CVT que se está analizando.

4. 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 hipertensión 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 evaluarán 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. 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.