

UNIVERSIDAD NACIONAL DE ROSARIO



FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA

ANTEPROYECTO DE TESINA

Incorporación de covariables que varían en el tiempo a un modelo mixto

Autor: Esteban Cometto

Directora: Noelia Castellana

Codirectora: Cecilia Rapelli

13 de octubre de 2021

Índice

1. Introducción	2
2. Objetivos	3
2.1. Objetivo Principal	3
2.2. Objetivos Específicos	3
3. Metodología	4
4. Materiales	5
5. Cronograma	6

1. Introducción

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. Esta tesina 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 aplicando los conceptos en un ejemplo real.

2. Objetivos

2.1. Objetivo Principal

Profundizar en el estudio de propuestas metodológicas respecto a la incorporación de covariables que varían con el tiempo en modelos mixtos para datos longitudinales.

2.2. Objetivos Específicos

- Profundizar en las definiciones de los tipos de covariables
- Describir diferentes transformaciones que pueden realizarse sobre las covariables que varían en el tiempo
- Evaluar propuestas de incorporación de covariables que varían en el tiempo en los modelos mixtos
- Aplicar los conceptos vistos a una problemática con datos que surgen de un estudio sobre hipertensión arterial

3. Metodología

En un modelo mixto, cada unidad tiene una trayectoria individual caracterizada por parámetros individuales. La respuesta media es modelada como una combinación de características poblacionales que son comunes a todos los individuos (efectos fijos) y efectos específicos de la unidad que son únicos de ella (efectos aleatorios), considerando las dos fuentes de variación (intra y entre) presentes en los datos longitudinales.

Este modelo supone implícitamente que la media condicional de la respuesta j -ésima, dado X_{i1}, \dots, X_{in_i} , depende solo de X_{ij} .

$$E(Y_{ij}|\mathbf{X}_i) = E(Y_{ij}|X_{i1}, \dots, X_{in_i}) = E(Y_{ij}|X_{ij}) \quad (1)$$

Con las covariables fijas, esta suposición se mantiene necesariamente desde $X_{ij} = X_{ik}$ para todas las ocasiones $k \neq j$. Con covariables variables en el tiempo que se fijan por diseño del estudio (por ejemplo, indicador de grupo de tratamiento en una prueba cruzada), la suposición también se cumple ya que los valores de las covariables en cualquier ocasión se determinan a priori por diseño del estudio y de manera completamente no relacionado con la respuesta longitudinal. Sin embargo, cuando una covariable es variable en el tiempo, puede que no necesariamente se mantenga.

En general, cuando (1) no se cumple, los valores precedentes y/o posteriores de la CVT confunden la relación entre Y_{ij} y X_{ij} , esto puede llevar a estimaciones sesgadas de los parámetros del modelo.

Se dice que una covariable que varía en el tiempo es exógena cuando los valores actuales y anteriores de la respuesta en la ocasión j (Y_{i1}, \dots, Y_{ij}), dados los valores actuales y precedentes de la CVT (X_{i1}, \dots, X_{ij}), no predice el valor posterior de X_{ij+1} . De lo contrario, se dice que la CVT es endógena.

Cuando las covariables son variables en el tiempo, los parámetros de regresión no necesariamente tienen la interpretación causal implícita incluso cuando (1) se cumple. A los parámetros de regresión se les puede dar una interpretación causal solo cuando se puede asumir además que las CVT son exógenas con respecto a la variable de respuesta.

4. Materiales

Los datos que se utilizarán provienen de un programa de atención y control de pacientes hipertensos iniciado en el año 2014 en Rosario que realiza un seguimiento exhaustivo de 560 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 variable "adherencia al tratamiento farmacológico" influye en los valores de la tensión arterial sistólica a lo largo del seguimiento. Dicha variable se evaluó a través de un cuestionario validado que otorga como resultado una respuesta dicotómica: adhiere - no adhiere. Este cuestionario se realizó en cada visita. Como la adherencia al tratamiento farmacológico es una CVT, se evaluarán diferentes enfoques para incluirla en un modelo mixto que pueda explicar el cambio en la tensión arterial sistólica media a lo largo del tiempo.

Un aspecto a tener en cuenta en este trabajo es que, si bien contamos con mucha otra información para obtener modelos que describan de mejor manera el comportamiento de la tensión arterial sistólica, nos centraremos en modelos más simples con respecto a las covariables fijas con el fin de no perder de vista la relación entre la variable respuesta y la CVT.

Para el desarrollo del análisis de los datos se utilizará el lenguaje Python.

5. Cronograma

Octubre 2020 - Julio 2021

- Revisión bibliográfica sobre covariables que varían en el tiempo para modelos longitudinales

Agosto 2021 - Octubre 2021

- Procesamiento y análisis de la información

Noviembre 2021 - Abril 2022

- Elaboración de informe

Mayo 2022 - Junio 2022

- Revisión de informe

Julio 2022

- Fecha tentativa de presentación

Bibliografía

- [1] G. M. Fitzmaurice, N. M. Laird, and J. H. Ware, *Applied Longitudinal Analysis*. John Wiley & Sons, 2004.
- [2] P. J. Diggle, P. Heagerty, K.-Y. Liang, and S. L. Zeger, *Analysis of Longitudinal Data*. Oxford University Press, 2002.
- [3] D.-G. Chen and J. R. Wilson, *Innovative Statistical Methods for Public Health Data*. Springer International Publishing, 2015.
- [4] L. Hoffman, *Longitudinal Analysis, Modeling Within-Person Fluctuation and Change*. Routledge, 2015.