Universidad Nacional de Rosario

Facultad de Ciencias Económicas y Estadística

Anteproyecto de Tesina

Modelos longitudinales con covariables que varían en el tiempo

Autor: **Esteban Cometto**

Director: Maria Del Carmen García

Codirector: Noelia Castellana

17 de agosto de 2021

Índice

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 varía en el tiempo (CVT) 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.

Un programa de atención y control de pacientes hipertensos iniciado en el año 2014 en Rosario 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 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.

2. Objetivos

2.1. Objetivo Principal

Profundizar en el estudio de propuestas metodológicas para utilizar la información obtenida de la covariable que varía en el tiempo dentro de un modelo mixto.

2.2. Objetivos Específicos

- Específicar distintos tipos de CVT
- Transformaciones a realizar sobre la CVT antes de incluirla al modelo, incluyendo conversión a covariable fija
- Consideraciones sobre interpretación de los parámetros sobre las CVT
- Indagar sobre feedback entre la CVT y la variable respuesta

3. Metodología

En un modelo mixto, cada unidad tiene una trayectoria individual caracterizada por parámetros y un subconjunto de esos parámetros ahora se consideran aleatorios. 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).

Se consideran las dos fuentes de variación (intra y entre) presentes en los datos longitudinales. Entonces, este modelo va a ser similar al modelo lineal general con respecto a la parte media del mismo, pero se va a diferenciar en cuanto a la estructura de covariancia.

El modelo lineal mixto para la unidad i se puede expresar en forma matricial como:

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i \quad i = 1, ...n \tag{1}$$

Donde:

- $\boldsymbol{Y}_i$: Vector de la variable respuesta de la i-ésima unidad, de dimensión $(n_i * 1)$
- $\boldsymbol{X}_i$: Matriz de diseño de la i-ésima unidad, que caracteriza la parte sistematica de la respuesta, de dimensión $(n_i * p)$
- $\boldsymbol{\beta}$: Vector de parámetros de dimensión $(p * 1)$
- $\boldsymbol{Z_i}$: Matriz de diseño de la i-ésima unidad, que caracteriza la parte aleatoria de la respuesta, de dimensión $(n_i * k)$
- $\boldsymbol{b}_i$: Vector de efectos aleatorios de la i-ésima unidad, de dimensión $(k * 1)$
- $\boldsymbol{\varepsilon}_i$: Vector de errores aleatorios de la i-ésima unidad, de dimensión $(n_i * 1)$
- n: número de individuos
- p: número de parámetros

Además, $\boldsymbol{b}_i$ y $\boldsymbol{\varepsilon}_i$ son independientes.

$$\boldsymbol{b}_i \; N_{n_i}(0, R_i)$$

$$\boldsymbol{\varepsilon}_i \; N_{n_i}(0, D_i)$$

Las matrices D_i y R_i contienen las variancias y covariancias de los elementos de los vectores $\boldsymbol{b}_i$ y $\boldsymbol{\varepsilon}_i$ respectivamente. La media marginal de $\boldsymbol{Y}_i$ está dada por:

$$E(\boldsymbol{Y}_i) = \boldsymbol{X}_i\boldsymbol{\beta} \tag{2}$$

Y la media condicional de $\boldsymbol{Y}_i$, dado $\boldsymbol{b}_i$, por:

$$E(\boldsymbol{Y}_i|\boldsymbol{b}_i) = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b}_i \tag{3}$$

Dejando de lado la forma matricial, lo que a menudo se pasa por alto es la suposición implícita de que la media condicional de la respuesta j-ésima, dado $X_{i1}, ..., X_{in_i}$, depende solo de X_{ij}.

$$E(Y_{ij}|\boldsymbol{X}_i) = E(Y_{ij}|X_{i1}, ..., X_{in_i}) = E(Y_{ij}|X_{ij}) \tag{4}$$

3.1. Supuestos y problemas en la estimación

Con las covariables fijas, esta suposición se mantiene necesariamente desde $X_{ij} = X_{ik}$ para todas las ocasiones $k \neq j$. Además, 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 y estocástica, puede que no necesariamente se mantenga.

En general, cuando (4) 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 β

3.2. Variables Endógenas y Exógenas

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 $Y_{i1}, ..., Y_{ij}$, dados los valores actuales y precedentes de la CVT $X_{i1}, ..., X_{ij}$, no predice el valor posterior de X_{ij+1}. Más formalmente, una CVT es exógena cuando:

$$f(X_{ij+1}|X_{i1},...,X_{ij},Y_{i1},...,Y_{ij}) = f(X_{ij+1}|X_{i1},...,X_{ij}) \tag{5}$$

De lo contrario, se dice que la CVT es endógena. Cuando la CVT es exógena, se tiene que:

$$E(Y_{ij}|\boldsymbol{X}_i) = E(Y_{ij}|X_{i1},...,X_{in_i}) = E(Y_{ij}|X_{i1},...,X_{ij}) \tag{6}$$

que es una suposición más débil que 4.

3.3. Pruebas sobre la exogeneidad

En principio, es posible examinar la suposición de que una covariable que varía en el tiempo es exógena al considerar modelos de regresión para la dependencia de X_{ij} en $Y_{i1},...,Y_{ij}$ (o alguna función conocida de $Y_{i1},...,Y_{ij}$) y $X_{i1},...,X_{ij-1}$ (o alguna función conocida de $X_{i1},...,X_{ij-1}$). La ausencia de cualquier relación entre X_{ij} e $Y_{i1},...,Y_{ij-1}$, dado el perfil de covariable anterior, $X_{i1},...,X_{ij-1}$, proporciona soporte para la validez de la suposición de que el proceso de covariable es exógeno.

3.4. Conclusiones sobre la exogeneidad

En conclusión, cuando las covariables son variables en el tiempo y estocásticas, los parámetros de regresión no necesariamente tienen la interpretación causal implícita incluso cuando 4 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 (es decir, cuando 5 se cumple).

4. Aplicación

Se propone estudiar el comportamiento de la adherencia al tratamiento farmacológico sobre la tensión arterial sistólica (TAS) a través de distintas formas de introducirla al modelo y también la comprobación de ciertos supuestos para poder obtener interpretación causales sobre sus parámetros. 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 TAS, 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.