

Análisis de Datos Longitudinales

Maestría en Demografía y Estudios de Población

Alejandra Marroig

Instituto de Estadística, Universidad de la República, Uruguay

August 2025

Contenidos

En esta clase veremos

- 1 Contexto de investigaciones del cambio en el tiempo
- 2 Análisis exploratorio del cambio con datos longitudinales
- 3 **Análisis de datos con el modelo multinivel**
- 4 **Aspectos avanzados de análisis de datos longitudinales (no linealidades, flexibilización de la variable tiempo, otros)**

Estrategias prácticas para el análisis de datos longitudinales con el modelo multinivel

- variables a considerar en efectos fijos
- interpretabilidad del modelo
- comparación de especificaciones del modelo e inferencia
- chequeo de supuestos del modelo

Variables en el modelo multinivel

- Basar decisiones para introducir, conservar y eliminar predictores en una combinación de lógica, teoría e investigaciones previas, complementadas con pruebas de hipótesis y comparaciones del ajuste del modelo.
- Antes de ajustar los modelos es deseable considerar:
 - 1 covariables (predictores) de interés en las preguntas de investigación
 - 2 covariables (predictores) de control, cuyos efectos se desean “eliminar” o “controlar”.

Interpretabilidad del modelo

- Considerar la posibilidad de centrar las variables:
 - 1 la variable de tiempo
 - 2 covariables (mejora interpretabilidad del intercepto)

Comparar modelos y hacer contrastes de hipótesis

- Para comparar modelos se puede usar LR (para modelos anidados), AIC, BIC. En R se usa el `anova()`.
- Se pueden hacer contrastes de hipótesis conjuntas con test de Wald. Los paquetes `lmer` y `lmerTest` tienen desarrollos muy recientes en estos temas.

Chequeo de los supuestos del modelo

Chequeos usuales sobre los supuestos del modelo:

- forma funcional
- normalidad de los errores
- homoscedasticidad

Chequeo de la forma funcional propuesta

- Hacer un análisis exploratorio del *outcome* versus las covariables de interés
- Nivel 1: para cada (algunos) individuo hacer visualizaciones de las trayectorias observadas y estimadas por MCO
- Nivel 2: hacer visualizaciones de las trayectorias estimadas por MCO para los grupos determinados por las covariables

Chequeo de normalidad de los residuos y homoscedasticidad

Para la normalidad hacer análisis exploratorios a partir de visualizaciones de los residuos estimados del modelo ($Y_{ij} - \hat{Y}_{ij} = \hat{\epsilon}_{ij}$)

- Para normalidad: Q-Q plots, etc.
- Para homoscedasticidad: covariables vs residuos, etc.

Tratando el tiempo de forma más flexible

Antes teníamos datos

- 1 balanceados: todos los individuos se “evalúan” el mismo número de ocasiones
- 2 tiempo estructurado: cada set de ocasiones es idéntico entre individuos

Ocasiones de medición en diferentes edades

- Es muy usual en los estudios longitudinales
- La edad medida en años de cada participante puede ser 50 o 50.3 o 60.5 o 74.35 o 75.01 etc, etc, etc,
- Es conveniente utilizar la métrica de tiempo en la que se evaluaron los individuos aunque queden espaciados de forma diferente entre individuos
- Es conveniente centrar la variable tiempo para que tenga sentido para la investigación (la edad al inicio del estudio de participantes suele ser buena elección)

Datos no balanceados

Tenemos diferente número de ocasiones de medición para cada individuo

- El modelo multinivel se puede estimar de todas formas sin mayores complicaciones
- Cada individuo aportará en mayor o menor medida a la verosimilitud según la cantidad de ocasiones de medición
- Individuos con menos ocasiones de evaluación no contribuyen a la estimación del componente de variabilidad individual
- La trayectoria ajustada de cada persona se basa en una combinación de: (1) su trayectoria observada y (2) una trayectoria basada en el modelo determinada por los valores de las covariables.

Problemas prácticos de la estimación con datos no balanceados

- Puede haber un problema computacional y de los métodos de optimización para encontrar estimaciones MV
- Lo más usual es que si los datos están muy “desbalanceados”, los algoritmos no convergen y no se obtienen las estimaciones MV
- Esto afecta únicamente a los componentes de varianza, pero no a la estimación de los efectos fijos (esto es igual que la regresión lineal):
 - se requieren suficientes personas con datos suficientes para permitir cuantificar la variación residual intraindividual

Datos faltantes en estudios longitudinales

- Lo clave es entender la razón por la cual algunos individuos tienen menos ocasiones de medición
- Los datos faltantes en estudios longitudinales son un tema de importancia para obtener estimaciones insesgadas
- Se distinguen tres casos:
 - a. Datos faltantes completamente al azar (*MCAR*)
 - b. Datos faltantes al azar (*MAR*)
 - c. Datos faltantes no al azar (*MNAR*)

Pregunta clave

¿es posible asumir que la probabilidad del dato faltante no está relacionada con la variable *outcome* no observada (condicionado por todos los resultados observados, por ejemplo, en el pasado)?

Covariables que cambian en el tiempo

Conceptualmente, no se necesitan estrategias especiales para incluir una covariable cuyo valor cambia en el tiempo en un modelo multinivel para el cambio.

Ejemplos:

- Estado conyugal y su efecto sobre la trayectoria de la satisfacción con la vida
- Desempleo y su efecto sobre la trayectoria de la depresión (escala CES-D)

Lo que ahora estas covariables entran en el nivel 1 del modelo y aportan a la trayectoria *within-person*

Componentes de varianza con covariables que varían en el tiempo

- Al incluir covariables invariantes en el tiempo los componentes de varianza cambian:
 - 1 variabilidad en el nivel 1 permanece relativamente estable porque las covariables invariantes no pueden explicar gran parte de la variación intraindividual
 - 2 variabilidad en el nivel 2 disminuyen si las covariables invariantes “explican” parte de la variación entre personas (intercepto o pendiente).

Variables que varían en el tiempo

Pueden afectar a los tres componentes de la varianza porque varían tanto dentro como entre personas.

Modelos para cambios discontinuos y no lineales en el tiempo

- 1 Discontinuidades abruptas en la trayectoria de cambio individual
- 2 Transformación del *outcome* o del tiempo para transformar en modelo multinivel lineal
- 3 Trayectorias que son funciones polinómicas del tiempo
- 4 La variable *outcome* es una función no lineal explícita de los parámetros de crecimiento

Comentario

Lo anterior asume que la trayectoria individual no tiene discontinuidades y es lineal. Pero se puede relajar este supuesto

¡¡Gracias!!

alejandra.marroig@fcea.edu.uy