

Los Modelos Mixtos de Clases Latentes en Ciencias de la Salud.

Ruth Fuentes García
Facultad de Ciencias, UNAM
Septiembre, 2021

Modelos Mixtos de Clases Latentes

Los modelos lineales mixtos pueden usarse para analizar datos con medidas repetidas; cuando interesa considerar una población o quizá varias subpoblaciones (incluso saber cuántas son, si es que las hay). Los modelos mixtos de clases latentes (LCLMM) combinan las características de un modelo lineal mixto con una variable latente adicional que divide a la población en subpoblaciones que son más homogéneas.

En un modelo lineal mixto, para la observación i :

$$y_i = X_i\beta + Z_i b_i + e_i,$$

donde y_i es el vector de observaciones, X_i es la matriz de diseño para los efectos fijos, Z_i es la matriz de diseño para los efectos aleatorios y e_i es el vector de errores aleatorios. Adicionalmente b_i y e_i se asumen aleatorios.

El modelo de clases latentes considera - para un número de subpoblaciones K dado - una variable latente para cada observación:

$$c_{ik} = \begin{cases} 1 & \text{si la observación } i \text{ pertenece a la clase } k, \\ 0 & \text{si la observación } i \text{ no pertenece a la clase } k, \end{cases}$$
$$(c_{i1}, \dots, c_{iK}) \sim \text{Multinomial}(1, (\pi_{i1}, \dots, \pi_{iK})),$$

donde $c_i = (c_{i1}, \dots, c_{iK})'$. Las probabilidades π_{ik} , correspondientes a la clase latente en la multinomial están dadas por:

$$\pi_{ik} = \mathbb{P}(c_{ik} = 1 | v_i) = \frac{\exp(v_i' \alpha_k)}{\sum_{j=1}^K \exp(v_i' \alpha_j)},$$

donde v_i es el vector de variables relacionadas a la asignación de la clase para la observación i , y α_k es el vector de parámetros asociado a esta asignación k donde para la primera clase latente $\alpha_1 = 0$. El número de clases se tendrá que determinar como parte de la selección del modelo.

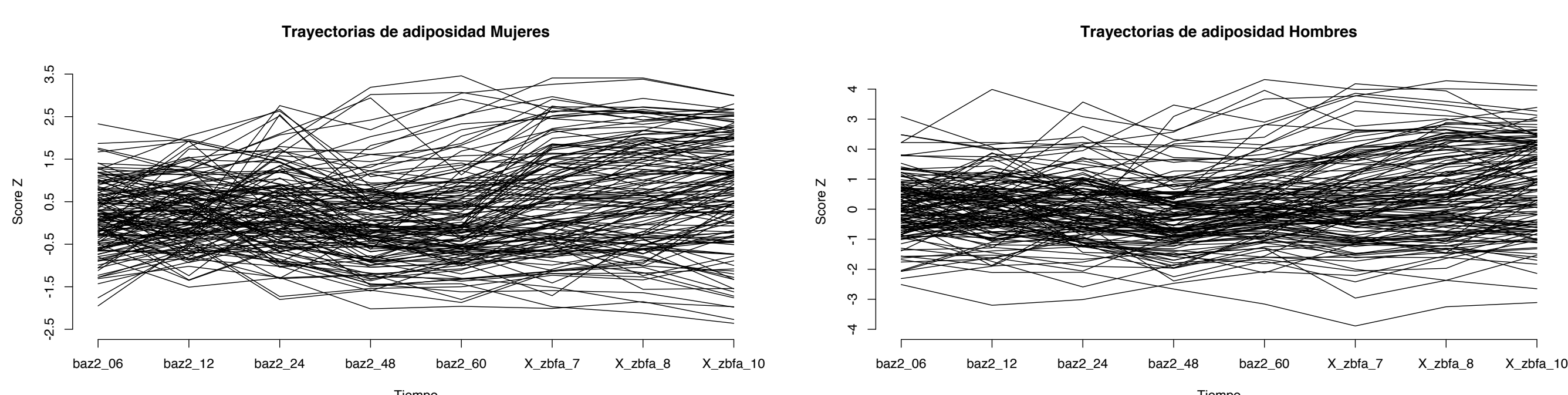
La log-verosimilitud asociada

$$\log L(\theta) = \sum_{i=1}^n \log \sum_{k=1}^K \pi_{ik} f(y_{ik} | c_{ik} = 1)$$

La estimación de los parámetros del modelo se puede hacer con generalizaciones del algoritmo EM u otros algoritmos de optimización. Existen rutinas disponibles en SAS, Stata, Mplus y otros softwares que trabajan con licencias. En los ejemplos descritos se estimaron los modelos utilizando la librería `lcmm` [3] de R, que es de uso libre.

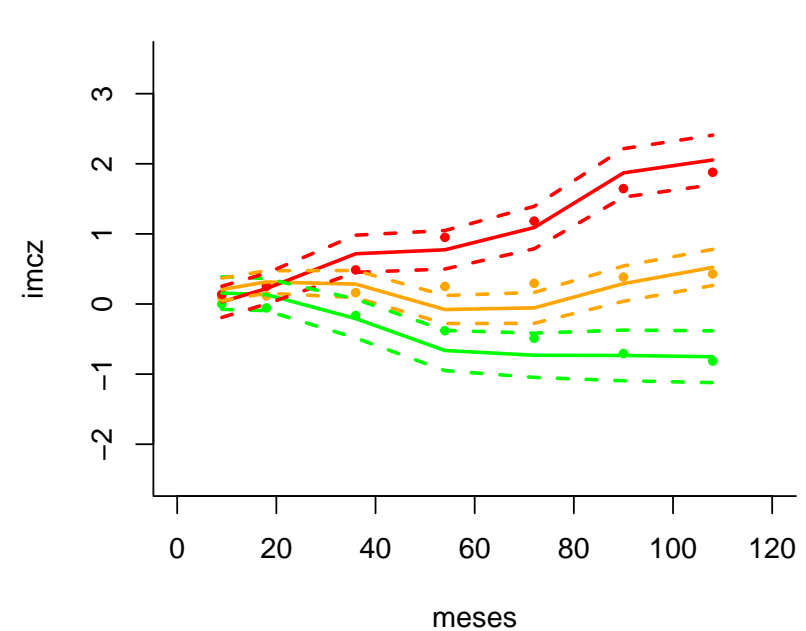
Ejemplo I

Este ejemplo considera 286 niños de una cohorte que cuenta con 8 mediciones del score Z del índice de masa corporal. Este ajuste permitió identificar diferentes trayectorias de adiposidad infantil.

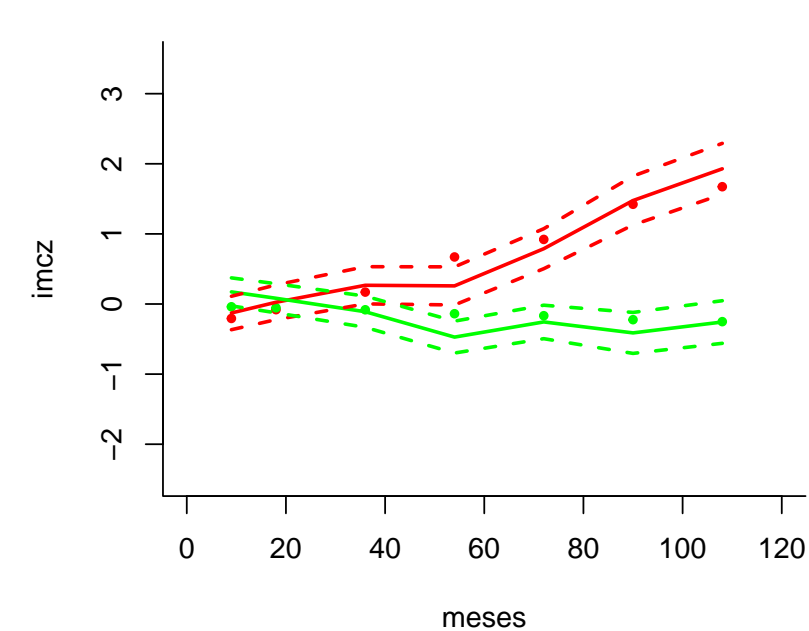


Trayectorias del imcz niñas.

Trayectorias del imcz niños.



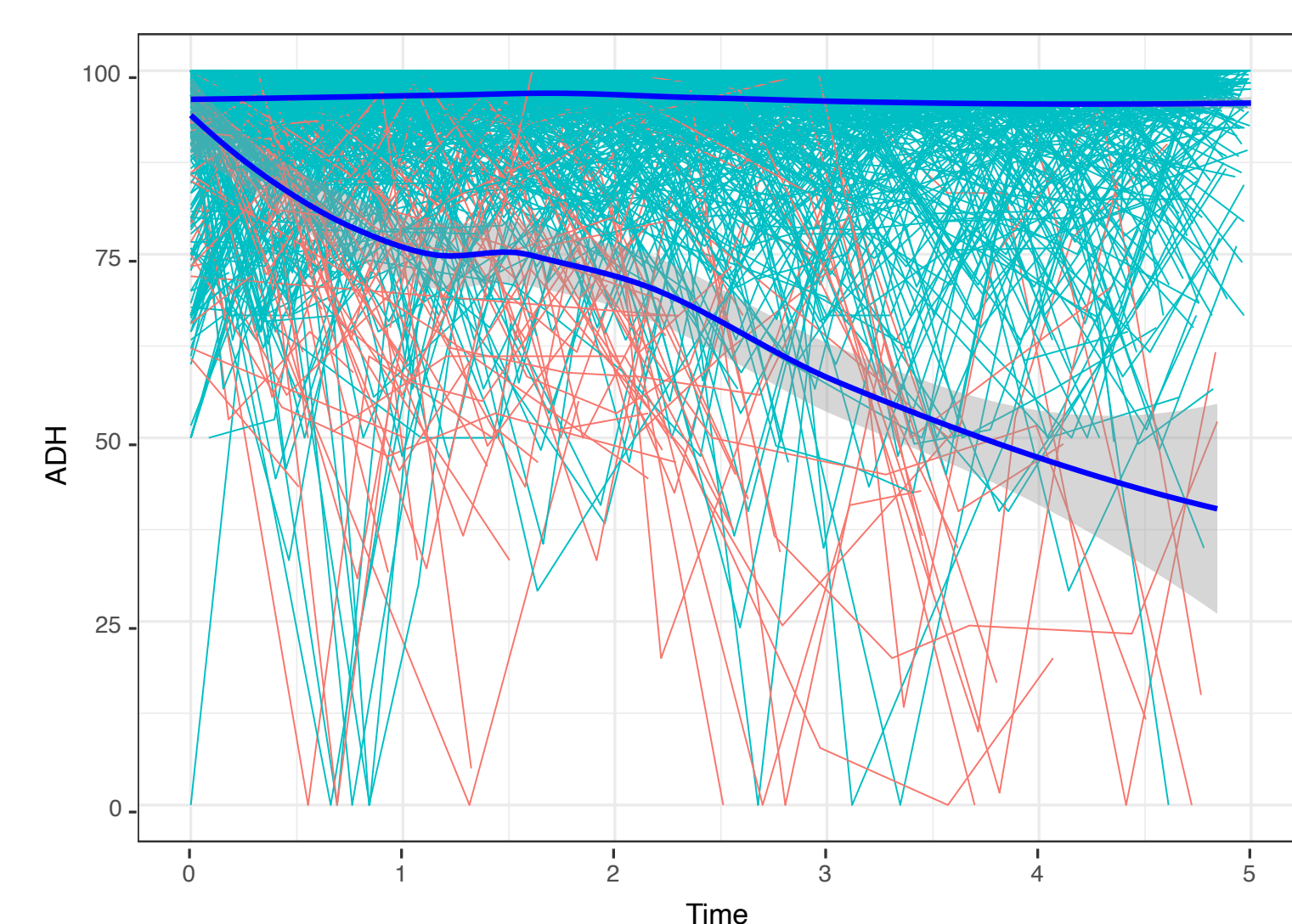
Trayectorias ajustadas.



Trayectorias ajustadas.

Ejemplo II

En este ejemplo se considera la adherencia a tratamiento de pacientes, esta adherencia se estima como la proporción de píldoras solicitadas en su clínica por mes. El análisis permite identificar grupos de comportamiento que pueden repercutir en la calidad de vida de los pacientes.



Bibliografía

- ▶ Muthen, B. (2004) Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In DKaplan (ed) Handbook of Quantitative Methodology for the Social Sciences, pp 345-368. Newbury Parc, CA: Sage.
- ▶ Proust-Lima, C., Pillips, V, Liqueet, B. (2017) Estimation of Extended Mixed Models Using Latent Classes and Latent Processes: The R Package `lcmm`. Journal of Statistical Software, 78. 2.
- ▶ Castrejón-Salgado, R., Ramírez-Silva, I., Avila-Jiménez, L., Sánchez-Zamorano, L.M., Fuentes- García, R., Ramakrishnan, U. , Barraza-Villarreal, A., Romieu, A., Rivera-Dommarco., J.A. (2020) Adiposity trajectories in mexican children and their early determinants. A prospective birth cohort study. Ann Nutr Metab, 76, (suppl 4):1, 232