PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

ESCUELA DE GRADUADOS

Modelos Longitudinales de Diagnóstico Cognitivo

Tesis para optar el grado de Magister en Estadística

Presentado por:

Ronald Minchola

Asesor:

Dr. Luis Valdivieso

Miembros del jurado:

Dr. Nombre completo jurado 1
Dr. Nombre completo jurado 2
Dr. Nombre completo jurado 3

Lima, Diciembre 2019

Dedicatoria

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris. Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa. Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula. Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

Con mucho cariño para mis padres: José Ronier Minchola Zavaleta e Hilda Mary Alza Silva por haberme guiado desde pequeño en el mundo del estudio por haber dedicado mucho tiempo a mi formación académica y también como ser humano. Para Sinthia y mis pequeños hijos Sergio, Emilio y Valery que son fuente de inspiración siempre. Estoy seguro que algun dia leerán mi trabajo por su curiosidad hacia las ciencias. A mis hermanos Bruno y Aldo por su constante motivación para terminar con éxito esta gran proeza.

Agradecimentos

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris. Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa. Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula. Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus. A mis padres Hilda Mary y José Ronier con ese amor infinito que se tiene por los creadores, por haber inculcado en mi esas ganas de ser un profesional y mejor persona. A Sinthia por ser esposa y compañera por su comprensión y paciencia, a mis hijos Sergio, Emilio y Valery por ser fuente de inspiración siempre en cada paso que doy. Algún dia leeran este trabajo espero que si. Quiero agradecer a todos mis profesores desde el buhonero que me venia a enseñar fracciones, MCM y MCD cuando todavia era un niño alla en la lejana década de los ochentas en la desaparecida tienda de Comercial Rominza, pasando por todos mis profesores de las academias y del CEP Claretiano recordando al gran Osiel Linares con sus tizas de colores y sus Diagramas de Venn, hasta todos los profesores de la Maestria en Estadistica muy en especial al Dr. Luis Valdivieso Serrano por su calidad de persona y ser humano, además de su sabiduria y paciencia hacia mi persona Agradecer también con mucho corazón a mis tios Efrain y mi dulce tia Josefina Çhepita"por haberme apoyado cobijandome del frio limeño mil gracias me han dado mucho cariño y enseñanzas diariamente.

Resumen

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum. Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum. Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Palabras clave: palabra-clave1, palabra-clave2, palabra-clave3.

Abstract

.Abstract ..Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum. Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum. Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Keywords: keyword1, keyword2, keyword3.

Índice general

Capítulo 1

Introducción

1.1. Consideraciones Preliminares

Los modelos de diagnóstico cognitivo (CDM) son modelos de clases latentes que se utilizan para clasificar a los encuestados en grupos homogéneos basados en múltiples variables latentes categóricas que representan a los atributos cognitivos medidos en una prueba. Uno de los modelos más populares de esta gran familia es el llamado modelo DINA, el cual tuvo su primera aparición con los trabajos de Haertel (1989) enfocados principalmente en el campo educacional. Este modelo considera respuestas observadas dicotómicas de parte de los examinados, variables predictoras latentes dicotómicas y tiene como restricción que los examinados deben de dominar obligatoriamente todas las habilidades requeridas para correctamente responder cada item, aquellas que se resumen en una matriz denominada Q. Asimismo, este modelo estima parámetros para los items, los cuales son denominados parámetros de ruido (estimación de dos probabilidades de error): Adivinación y Desliz, el primero es la probabilidad de responder correctamente a un item j a pesar de no dominar las habilidades requeridas para hacerlo, mientras que el segundo es la probabilidad de fallar a un item a pesar de dominar las habilidades requeridas para hacerlo; además el modelo DINA se encuentra clasificado como un modelo no compensatorio conjuntivo. Es no compensatorio porque requiere que cada una de las habilidades esté presente para producir una respuesta correcta al ítem (dicotómico o politómico) y es conjuntivo cuando todas las habilidades requeridas para responder un ítem, necesariamente tienen que ser dominadas por el individuo para obtener la respuesta correcta.

En este proyecto se busca estudiar cómo esta clasificación pudiera verse afectada en el tiempo, cuando es factible el aplicar la prueba en repetidas ocasiones.

Para ello, algunos autores como Kaya and Leite [2017] han propuesto un modelo que combina el modelo de análisis de transición de clases latentes (LTA) y los CDMs. De manera similar Li, F. et.al. [2016] han utilizado el LTA para proponer una metodología de transición de clases con el modelo DINA, donde la estimación de su modelo lo realizan usando métodos bayesianos, a diferencia de los primeros autores que utilizaron el paquete Mplus implementado para analizar CDMs longitudinales junto con estudios de simulación de Montecarlo. El objetivo de esta investigación es analizar estas propuestas u otras relacionadas al problema e implementar la estimación de sus parámetros. Se ilustrará también todo ello con una aplicación real y se espera que puedan realizarse, de ser posible, algunas variantes o extensiones del modelo para lidiar con las restricciones que tienen estos modelos, sobre todo en lo concerniente a la replicabilidad de la prueba original.

Usualmente cuando queremos analizar datos longitudinales las preguntas de investigación están en relación al cambio a través del tiempo, tal cambio, en nuestro modelo estará referido a cambios en las clases cognitivas latentes de los individuos.

El análisis de transición de clases latentes (LTA) es una aplicación longitudinal del modelo de análisis de clases latentes, que tiene por objetivo identificar si existe un cambio entre las clases latentes a través del tiempo (Collins y Lanza, 2010). Mientras que el análisis de clases latentes es aplicado a un conjunto de variables recolectadas en un tiempo específico para identificar las clases latentes que presentan los patrones de respuesta en los datos, el análisis de transición de clases latentes es aplicado a mediciones repetidas en el tiempo para estudiar el movimiento entre las clases latentes identificadas (Collins y Lanza, 2010).

El LTA fue desarrollado inicialmente para estudiar el cambio secuencial por etapas de un tipo de variables latentes llamadas variables latentes dinámicas (Collins and Wugalter, 1992). Las variables latentes dinámicas incluyen características tales como actitudes y patrones de personalidad que cambian a través del tiempo. La técnica ha sido, posteriormente utilizada para el estudio sobre las intervenciones como por ejemplo el abandono del consumo de tabaco y la disminución de malas conductas (comportamiento inapropiado). El análisis de transición latente es un método muy usado para investigar el crecimiento académico cuando las variables latentes son categóricas (Boscardin et al., 2008; Compton et al., 2008; Trentacosta et al., 2011)

En esta investigación, describiremos el uso del LTA en combinación con un modelo de diagnóstico cognitivo para analizar pruebas educativas largas que miden habilidades cognitivas múltiples. Esta nueva clase de modelos que los llamamos de transición de clases latentes cognitivas (LTCA) permiten investigar hipótesis acerca del efecto de una intervención (por ejemplo, un programa instruccional o un cambio en la política educativa) mediante el cambio en las probabilidades de transición antes y después de una intervención con relación al estado cognitivo.

En este nuevo modelo se usan datos longitudinales para investigar si algún cambio ha ocurrido entre los estados latentes a través del tiempo. Las transiciones son expresadas junto con sus probabilidades de cambio de un estado latente a otro.

El LTCA presenta varias razones para su uso: primero, que estos modelos pueden representar variables latentes multidimensionales, segundo pueden modelar y predecir el cambio a través del tiempo, el cual es en cierto sentido discreto y tercero nos permiten conocer si algunos estados latentes pueden tener prevalencias muy bajas en etapas iniciales, pero a medida que los individuos hacen la transición en el tiempo, su prevalencia aumenta (Collins y Lanza, 2010).

Este nuevo modelo, entonces permitirá a un investigador responder directamente a un conjunto de preguntas tales como: ¿Existe algún cambio entre las clases latentes a través del tiempo? y si la respuesta es positiva explicar ¿cómo caracterizar dicho cambio a través de las probabilidades?

Así, como en el análisis de clases latentes, en este nuevo modelo se estiman también las probabilidades de respuesta al item, pero además la prevalencia en las clases latentes que se entiende como el número de casos de un evento en una población en un momento dado y la incidencia de las transiciones entre las clases latentes junto con una medida del error.

Tres grupos de parámetros son estimados en un modelo LTCA. El primer grupo de parámetros está compuesto por las probabilidades de transición entre las clases latentes. Estas son particularmente importantes porque ofrecen la solución a la mayoría de preguntas del tipo: ¿Cómo cambian los estados de dominio de las habilidades cognitivas latentes de etapa en etapa? El segundo grupo de parámetros estima el dominio o no de las digamos C distintas habilidades cognitivas por parte de cada estudiante en su primera medición o etapa. El tercer grupo de parámetros estima la relación entre el estado latente y la pregunta. Esto es, produce una probabilidad de respuesta correcta o incorrecta para cada item dado el dominio de los diferentes estados latentes.

El segundo y el tercer grupo de parámetros proporcionan también la solución a otra pregunta de investigación: ¿Cuál es el estado de dominio individual en cada habilidad cognitiva en cada etapa u ocasión? La respuesta a esta pregunta proporciona información útil para el estado del estudiante o paciente en cada punto en el tiempo.

1.2. Objetivos

El objetivo general de la tesis es estudiar un nuevo modelo de transición de clases latentes de habilidades cognitivas para datos longitudinales, así como observar su aplicación a un conjunto de datos reales. De manera específica:

- Revisar la literatura acerca del análisis de transición latente
- Presentar los fundamentos y propiedades de los modelos de transición de clases latentes (LTA), los modelos de análisis de clases latentes (LCA) y los modelos de diagnóstico cognitivo (MDCs)
- Estudiar puntualmente el nuevo modelo de transición latente.

- Estudiar el proceso de estimación de parámetros en los modelos de transición latentes bajo el enfoque clásico o frecuentista
- Realizar un estudio de simulación a efectos de comparar la estimación obtenida
- Aplicar el modelo a un conjunto de datos longitudinales en el área educativa que involucren algún o algunos tipos de tareas cognitivas.

1.3. Organización del Trabajo

Para alcanzar los objetivos de la investigación se considera la organización siguiente:

En el Capítulo 1 se presenta consideraciones generales acerca de los principales modelos de diagnóstico cognitivo entre los que destacan el modelo DINA, el modelo G-DINA y a su vez se estudia la literatura acerca del análisis de transición latente (LTA) y las cadenas de Márkov. En el Capítulo 3 se estudia el modelo de transición de clases latentes cognitivas, la estimación de sus parámetros y su implementación computacional. En el capitulo 4 se presentan los resultados de la aplicación del modelo a un conjunto de datos longitudinales en el área educativa orientados a conocer algún o algunos tipos de tareas cognitivas. En el capitulo **??** se presentan algunas conclusiones, recomendaciones y sugerencias para futuras investigaciones que se podrian derivar de este trabajo. Se incluirá finalmente en anexos código en R para la estimación de los parámetros.

Capítulo 2

Modelos de clases latentes y diagnóstico cognitivo

En muchas áreas de estudio, como en las ciencias sociales, ciencias de la salud y particularmente en las ciencias del comportamiento como la psicología, es de sumo interés aproximarnos a constructos teóricos, los cuales si bien no pueden ser observados directamente, asumimos que tienen un efecto sobre variables que si pueden ser medidas, permitiéndonos un acercamiento a estos constructos y un mejor entendimiento acerca de las variables medidas. Estos constructos son comúnmente llamados variables latentes en la literatura estadística y su modelo estadístico especifica la distribución conjunta de las variables observadas o llamadas también manifiestas (indicadores) y las variables latentes. Por otro lado, los modelos de diagnóstico cognitivo son modelos psicométricos usados para evaluación de estudiantes en cuanto a sus perfiles o clases y que nos permiten una medición efectiva del aprendizaje así como también una mejor instrucción y posiblemente políticas de intervención educativa para hacer frente a las necesidades individuales y grupales de los estudiantes. Es por eso, que se pasará a estudiar todo lo concerniente a la integración de estos conceptos

2.1. Modelos de clases latentes

El modelo de clases latentes es un caso particular de los modelos de variables latentes en el cual tanto las variables observables o manifiestas como las no observables o latentes son categóricas. La idea es poder agrupar a los individuos en clases según las variables observables. Se considera que la variable latente es categórica como producto de una evidencia a priori o teórica, o simplemente tomada así por una cuestión práctica. Para mayor detalle ver Lazarsfeld and Henry (1968). Las ventajas del análisis de clases latentes son varios: reducen la complejidad de un conjunto de datos agrupándolos en clases, permite estimar varias probabilidades, y permite analizar datos categóricos sin la necesidad de aplicar transformaciones. Se comienza a partir de dos supuestos para la versión estándar del modelo. Dentro de cada clase, todos los individuos tienen las mismas probabilidades de respuesta a las variables manifiestas. Se cumple la independencia condicional; es decir, las respuestas entre los individuos son independientes dado que estos pertenecen a una misma clase. Para tener una idea analicemos primero un modelo de clases latentes con variables manifiestas binarias, en el que la variable latente L es unidimensional y presenta C categorías o clases que los denotaremos con el índice $c = 1, 2, \ldots C$. Sea π_{ic} la probabilidad de obtener una respuesta positiva a la variable manifiesta X_i, dado que el individuo i esté en la clase c; es decir, $\pi_{ic} = P\left(X_i = 1|L = c\right)$ y sea η_c la probabilidad a priori de que un individuo pertenezca a la clase c. Entonces la función de probabilidad conjunta del vector de respuestas manifiestas o patrones de respuesta $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)$ viene dada por:

$$f(\mathbf{x}) = P(\mathbf{X} = \mathbf{x}) = \sum_{c=1}^{C} P(\mathbf{X} = \mathbf{x}|L = c)P(Y = c) = \sum_{c=1}^{C} g(\mathbf{x}|c)\eta_c, \tag{2.1}$$

donde

$$g(\mathbf{x}|c) = \prod_{i=1}^{p} \pi_{ic}^{x_i} \left(1 - \pi_{ic}\right)^{1-x_i} \tag{2.2}$$

y x_i es la respuesta 1 o 0 a X_i.

De esta manera, cada individuo es descrito por dos dimensiones: el vector de variables manifiestas $\boldsymbol{X}$ y el indicador de pertenencia a la clase. La probabilidad a posteriori de que un individuo con un patrón de respuestas $\boldsymbol{x} = (x_1, x_2, \ldots, x_p)'$ pertenezca a la clase j es por tanto:

$$h(c|\mathbf{x}) = P(L = c|\mathbf{X} = \mathbf{x}) = \frac{\eta_c \prod\limits_{i=1}^{p} \pi_{ic}^{x_i} \left(1 - \pi_{ic}\right)^{1-x_i}}{\sum_{c=1}^{K} \eta_c \prod_{i=1}^{p} \pi_{ic}^{x_i} \left(1 - \pi_{ic}\right)^{1-x_i}}. \tag{2.3}$$

Esta función es utilizada luego como regla de clasificación para asignar a un individuo a la clase que tiene mayor probabilidad de pertenencia, dado el patrón de respuestas que tenga. El modelo requiere de la estimación de los parámetros η_c y π_{ic}, y el cálculo del ajuste del modelo. Otro aspecto es también identificar las clases latentes subyacentes e interpretarlas de manera que tengan un sentido con los datos.

2.2. Cadenas de Markov

Extenderemos ahora el modelo de clases latentes a datos de caracter longitudinal para averiguar entre otras cosas como cambian, según los datos, la pertenencia de un individuo a una clase. Para ello requeriremos del concepto de cadenas de Markov que pasamos brevemente a revisar.

Consideremos un conjunto E finito o numerable. Sea $X_0, X_1, X_2, \ldots$ una sucesión de variables aleatorias discretas las cuales toman valores en E y se asumen están definidas en un espacio probabilistico (Ω, A, P). El conjunto E se denominará espacio de estados, y sus elementos estados. Las variables aleatorias $X_0, X_1, \ldots, X_n$ se dicen independientes si y sólo si, se cumple que:

$$P(X_0 = X_0, X_1 = X_1, \ldots, X_n = X_n) = \prod_{k=1}^{n} P(X_k = X_k), \tag{2.4}$$

para todo estado $x_0, X_1, \ldots, X_n$. En este caso los eventos $A = \{X_0 = X_0, \ldots, X_n = X_n\}$ y $B = \{X_{n+1} = X_{n+1}, \ldots, X_{n+m} = X_{n+m}\}$ son independientes para cualesquier $n = 0, 1, \ldots$ y $m = 1, 2, \ldots$ y cualquiera sea la sucesión de estados $X_0, \ldots, X_{n+m}$. La dependencia markoviana consiste en que la probabilidad del suceso B depende sólo del valor que toma la variable aleatoria X_n y no de los valores que toman las variables aleatorias $X_0, \ldots, X_{n-1}$. Si el índice de la sucesión representa el tiempo y n es el instante presente en una cadena de Markov, podemos decir que la probabilidad de que un suceso ocurra en los instantes futuros $n+1$,. . . ,$n+m$, depende solamente del estado en que se encuentra la sucesión en el instante presente n y no de los estados en que se encontró en los instantes pasados $0, 1, \ldots, n-1$.
Formalmente diremos que la sucesión $\{X_n\}$ es una cadena de Markov con espacio de estados discreto si:

$$\begin{aligned} P(X_{n+1} = X_{n+1} | X_n = X_n, \ldots, X_0 = X_0) &= \ldots \\ &= P(X_{n+1} = X_{n+1} | X_n = X_n), \end{aligned} \tag{2.5}$$

para todo $n = 1, 2, \ldots$ y cualquier sucesión de estados $X_0, \ldots, X_{n+1}$ en E, siempre que $P(X_n = X_n, \ldots, X_0 = X_0) > 0$. A la identidad (2.5) se le llama la propiedad de Márkov. En adelante para no recargar notaciones asumiremos, sin pérdida de generalidad, que el espacio de estados E es un subconjunto de $\aleph$. En tal sentido, diremos que una cadena de Markov es homogénea en el tiempo si para todo par de estados i, j, la probabilidad condicionada $P(X_{n+1} = j | X_n = i)$ no depende de n, es decir,

$$P(X_1 = j|X_0 = i) = P(X_2 = j|X_1 = i) = \ldots = P(X_{n+1} = j|X_n = i). \tag{2.6}$$

En general, utilizamos la expresión cadena de Márkov para referirnos a una cadena de Markov homogénea en el tiempo. La caracterización de la cadena en este caso se definirá a través de las probabilidades:

$$p_{ij} = P(X_1 = j|X_0 = i) \tag{2.7}$$

y

$$\pi_i^0 = P(X_0 = i)$$

A la matriz $P = [p_{ij}], i \in E, j \in E$ (posiblemente infinita) se le denomina la matriz de transición, y al vector $\pi^0 = (\pi_i^0)$, la distribución inicial de la cadena de Markov. Es sencillo, ver que la matriz de transición satisface las siguientes propiedades:

i) $p_{ij} \geq 0$ para todo par de estados i, j en E

ii) $\sum_j p_{ij} = 1$ para todo estado i en E

Por su parte, la distribución inicial satisface las propiedades:

i) $\pi_i^0 \geq 0$ para todo estado i en E,

ii) $\sum_i \pi_i^0 = 1$

La definición de cadena de Markov incluye como caso particular, las sucesiones de variables aleatorias independientes e idénticamente distribuidas y también las sucesiones formadas por sumas parciales de variables aleatorias independientes e idénticamente distribuidas, cuando dichas variables aleatorias toman valores enteros. Consideremos ahora las probabilidades de transición de orden n de una cadena Markov $\{X_k\}$ con espacios de estados E, matriz de transición P y distribución inicial π^0. Dados dos estados i, j sean:

$$p_{ij}^{(n)} = P(X_n = j|X_0 = i) \quad \text{y} \qquad \pi_i^{(n)} = P(X_n = i), \tag{2.8}$$

a la matriz $p^{(n)} = [p_{ij}^{(n)}]$, se le llama la matriz de transición de orden n y al vector $\pi^n = (\pi_i^n)$ la distribución de probabilidad en el periodo n de la cadena de Markov. Observemos que π^0 es la distribución inicial de la cadena de Markov y $P = p^{(1)}$ es su matriz de transición. Las probabilidades de transición de orden n satisfacen la ecuación de Chapman-Kolgomorov:

$$p_{ij}^{(m+n)} = \sum_k p_{ik}^{(m)} p_{kj}^{(n)}, \tag{2.9}$$

para todo par de indices m, n y todo par de estados i, j. Esto es, en notación matricial, (2.9) tiene la forma

$$p^{(m+n)} = p^{(m)} \times p^{(n)}, \tag{2.10}$$

donde x denota el producto de matrices. En efecto, aplicando la fórmula de la probabilidad total y la propiedad de Markov (2.5) se obtiene:

$$\begin{aligned} p_{ij}^{(m+n)} &= P(X_{m+n} = j|X_0 = i) = \sum_k P(X_{m+n} = j, X_m = k|X_0 = i) = \\ &= \sum_k P(X_{m+n} = j|X_m = k)P(X_m = k|X_0 = i) = \sum_{k\epsilon E} p_{ik}^{(m)} p_{kj}^{(n)} \end{aligned} \tag{2.11}$$

Análogamente se prueba que para la distribución de probabilidad en el instante n tiene lugar la identidad

$$\pi_j^{(m+n)} = \sum_k \pi_k^{(m)} p_{kj}^{(n)}, \tag{2.12}$$

para todo par de indices m, n y todo estado i. En notación matricial esto se escribe como:

$$\pi^{(m+n)} = \pi^{(m)} \times p^{(n)}$$

En particular, de (2.10) resulta que $p^{(n)} = P \times p^{(n-1)}$ y $p^{(n)} = P^n$. En conclusión, la matriz de transición de orden n es la potencia n-ésima de la matriz de transición P, y es correcto interpretar el superíndice n en la notación p^n como la potencia n-ésima de la matriz P. La distribución de probabilidad en el instante n se obtiene mediante la fórmula

$$\pi^n = \pi^0 \times p^n, \tag{2.13}$$

que también se escribe como:

$$\pi_j^n = \sum_k \pi_k^0 p_{kj}^{n-1}, \tag{2.14}$$

Calculemos ahora, para una elección de índices $n_1, \ldots, n_k$ arbitraria, la distribución del vector aleatorio $(X_{n_1}, \ldots X_{n_k})$, que llamamos la distribución finito-dimensional de la cadena de Markov. Es claro que para este cálculo es suficiente conocer las probabilidades de la forma

$$P(X_0 \epsilon A_0, \ldots, X_n \epsilon A_n) \qquad (n = 0, 1, \ldots),$$

donde $A_0, \ldots, A_n$ son subconjuntos arbitrarios de E. A su vez, para calcular estas últimas probabilidades, es suficiente, dada una sucesión de estados $i_0, \ldots, i_n$, conocer las probabilidades

$$\begin{aligned} P(X_n = i_n, \ldots, X_0 = i_0) &= \\ &= P(X_n = i_n | X_{n-1} = i_{n-1}) \ldots P(X_1 = i_1 | X_0 = i_0) P(X_0 = i_0) \\ &= p_{i_{n-1} i_n} \ldots p_{i_0 i_1} \pi_{i_0}. \end{aligned} \tag{2.15}$$

Este último cálculo muestra que las distribuciones finito dimensionales de una cadena de Markov pueden determinarse si se conocen la matriz de transición P y su distribución inicial π^0. Más aún, es posible demostrar que la probabilidad de un suceso que depende de una cantidad arbitraria, no necesariamente finita, de variables aleatorias de la cadena de Markov, se puede hallar a partir de P y de π^0 y que reciprocamente, dados un conjunto E, finito o numerable, una matriz $P = [p_{ij}], i \epsilon E, j \epsilon E$ que cumple las propiedades i) y ii) y un vector $\pi = (\pi_i)_{i \epsilon E}$ que satisface las propiedades i) y ii) existe un espacio de probabilidades y una cadena de Markov homogénea con espacio con espacio de estados E y variables aleatorias definidas en este espacio de probabilidades, que tiene a P como matriz de transición y a π^0 como distribución inicial. De aquí se obtiene la siguiente conclusión: Las propiedades probabilísticas de una cadena de Markov, dependen únicamente de la matriz transición y de la distribución inicial de la cadena de Markov. Cuando es necesario escribir de forma explícita la distribución inicial π^0 de una cadena de Markov, escribimos p_{π^0} en lugar de P. En particular, si esta distribución inicial está concentrada en un estado i escribimos p_i en vez de p_{π^0} y diremos que la cadena de Markov parte de i ya que $p_0(X_0 = i) = 1$. Consideremos un suceso arbitario de la forma

$$A = \{X_{n_1} = i_{n_1}, \ldots, X_{n_k} = i_{n_k}\}$$

Observamos que las probabilidades p_{π^0} y p_i correspondientes a la cadena de Markov con la misma matriz de transición, satisfacen la igualdad:

$$p_{\pi^0}(A|X_0 = i) = p_i(A),$$

como resulta de hacer ${\pi^0}_{i_0} = 1$ en la fórmula (2.5). Además, también se cumple:

$$p_\pi(A) = \sum_{i \epsilon E} p_{\pi^0}(A|X_0 = i)\pi_i = \sum_{i \epsilon E} p_i(A)\pi_i, \tag{2.16}$$

como resulta de aplicar el teorema de probabilidad total.

2.3. Modelos de transición de clases latentes

2.3.1. Introducción

El modelo de transición de clases latentes llamado también de mixtura de clases latentes de Markov es un modelo de variable latente similar al de clases latentes (LCA). Sin embargo, mientras que el LCA usa datos de corte transversal, el LTA es usado con data tipo panel o longitudinal para investigar si algún cambio ha ocurrido dentro de las clases latentes a través del tiempo. Cuando se dispone de datos longitudinales el modelo de transición de clases latentes permite direccionar un conjunto de preguntas: ¿Existe algún cambio entre las clases latentes a través del tiempo? Si es asi como se puede caracterizar este cambio? Si un individuo se encuentra en una clase latente particular en el tiempo t, cuál es la probabilidad de que este individuo permanezca en la misma clase latente en el tiempo $t + 1$, y cual la probabilidad de que el individuo este en una clase latente diferente. El LTA es una forma de ajustar modelos que abordan este conjunto de preguntas adicionalmente al conjunto de preguntas abordadas. Es por eso que esta investigación propone, no solo estudiar las clases latentes de algunos fenómenos educativos, sino también observar como los individuos transitan por las diferentes clases latentes a través del tiempo. El LTA al igual que el LCA estima probabilidades de respuesta al item. Por lo tanto, se estiman también la prevalencia de las clases latentes y la incidencia de las transiciones entre las clases latentes mientras se ajusta un error de medición. El modelo de transición de clase latente con solo dos puntos en el tiempo se plantea como:

$$P(Y = y) = \sum_{c_1=1}^{C} \sum_{c_2=1}^{C} \delta_{c_1} \tau_{c_2|c_1} \prod_{t=1}^{2} \prod_{j=1}^{J} \prod_{r_{j,t}=1}^{R_j} \rho_{j,r_{j,t}|c_t}^{I(y_{j,t}=r_{j,t})}, \tag{2.17}$$

donde

$$\sum_{c_1=1}^{C} \delta_{c_1} = 1, \sum_{c_2=1}^{C} \tau_{c2|c1} = 1, \sum_{r_{j,t}=1}^{R_j} \rho_{j,r_{j,t}|c_t} = 1, \tag{2.18}$$

siendo $Y = [Y_{j,t}]$ la matriz de respuestas de cualquier individuo, $Y_{j,t}$, su respuesta al item j en el tiempo t, C el número de clases latentes, J el número de items, δ_{c1} la probabilidad de pertenencia del individuo a la clase latente c_1 en el tiempo 1, $\tau_{c2|c1}$ la probabilidad de transición de que el individuo pase de la clase latente c_1 en el tiempo 1 a una clase c_2 en el tiempo 2, $r_{j,t}$ la categoría de respuesta al item j en el tiempo t, $I(y_{j,t} = r_{j,t})$ una variable indicadora que es igual a 1 si la respuesta al item j en el tiempo t es $r_{j,t}$ y 0 en caso contrario, y $\rho_{j,r_{j,t}|c_t}$, es la probabilidad de que su respuesta al item j en el tiempo t sea $r_{j,t}$ condicionado a la membresia a la clase latente c_t. Se asume que cada item j tiene R_j posibles categorías de respuestas. La consideración de un modelo con más de dos puntos en el tiempo es inmediata pero esta implica el incremento en el número de parámetros, no sólo en términos de las probabilidades de transición sino también de los parámetros dentro de las probabilidades $\rho_{j,r_{j,t}|c_t}$ de respuesta al item del modelo.

Ejemplo

Supongamos como ilustración que se tienen $J = 6$ variables observables binarias, que han sido medidas en los tiempos $t = 1, 2$. Por ejemplo, si estamos estudiando la delincuencia en adolescentes, las variables observadas podrian ser 6 items del cuestionario de delincuencia, donde cada variable observada j asumimos tiene $r_{j,t} = 1, 2$ categorías de respuesta (si o no) en cada tiempo. La tabla de contingencia cruzando las 6 variables con los dos tiempos tiene $2^{2\times6} = 4,096$ celdas o patrones de respuesta. Los patrones de respuesta en estas, son representados como $(r_{1,1}, \ldots, r_{6,2})$. Por ejemplo, un patrón de respuesta podria ser (si,no,no,no,no,no,si,si,no,no,no,no) que representa la respuesta de "*si*" para el primer item del cuestionario en el tiempo 1 hasta una respuesta de no para el sexto item en el tiempo 2. Sea L como antes, la variable latente categórica subyacente que tiene C estados latentes. Sea L_1 que representa la variable latente categórica en el tiempo 1, donde $c_1 = 1, \ldots, C$, L_2 representa la variable latente categórica en el tiempo 2, donde $c_2 = 1, \ldots, C$ en el tiempo 2 y así sucesivamente hasta L_T que representa a la variable latente categórica en el tiempo T, con $c_T = 1, \ldots, C$.

Las prevalencias en las clases latentes son estimadas para cada punto en el tiempo. Así, por ejemplo para un modelo con dos puntos en el tiempo como el anterior y con cuatro clases latentes, ocho prevalencias requerirán ser estimadas en total, mostrando ellas la proporción de personas en cada uno de los estados latentes. Por lo tanto, observando la prevalencia en estos estados podemos ver el aumento o disminución en las prevalencias entre los puntos en el tiempo considerados en el estudio, asimismo observar patrones de cambio y como las personas se mueven entre los estados.

Las probabilidades de respuesta al item representan las probabilidades de que las personas den respuestas correctas a cada item en particular sujeto a su clase de pertenencia. El modelo de transición de clase latente es particularmente usado más como un método exploratorio, en donde el etiquetado o codificación de los estados latentes se realiza evaluando las probabilidades de respuesta al item. Existe una probabilidad de respuesta al item para cada combinación item-estado. Por lo tanto, el número de celdas en una tabla de probabilidades de respuesta al item es el número de items multiplicado por el número de estados latentes.

Las probabilidades de transición son de relevante importancia porque permiten identificar si ocurrió algún cambio entre los estados latentes a través del tiempo. Para un modelo con dos puntos en el tiempo ($T = 2$), las probabilidades de transición permiten la creación de una tabla de clasificación cruzada con el número de estados latentes en el tiempo o etapa 1 y el número de estados latentes en el tiempo o etapa 2. En general, dado cierto número de puntos en el tiempo T, el modelo de transición de clase latente estima $T - 1$ matrices de probabilidades de transición. Cada examinado, se asume que puede ser miembro de uno y solo un estado latente en cada punto en el tiempo.

El modelo de transición de clases latentes asume la no existencia de datos faltantes en las variables indicadoras observadas. Sin embargo, los datos faltantes en el modelo LTA es sobrellevada de la misma forma que en el modelo de clases latentes LCA. Los datos pueden usarse a partir de las respuestas que dan los individuos a algunas de las preguntas en un tiempo en particular y para aquellos que solo están presentes en un subconjunto en el tiempo de medición. Se hace entonces la clásica suposición datos faltantes (missing at random).

En el presente estudio, a las clases latentes del modelo de transición las llamaremos estados latentes teniendo en cuenta que las clases latentes son estados temporales, y que los individuos pueden moverse hacía dentro y fuera de estos estados.

La combinación de grandes grados de libertad y la extrema dispersión no permiten probar hipótesis tradicionales acerca del ajuste absoluto del modelo de transición de clases latentes. Esto se debe a que la distribución de la estadística G^2(radio de verosimilitud) no es muy bien aproximada por una distribución Ji-cuadrado y por consiguiente los valores p, son bastante inexactos. Es por eso que se prefiere enmarcar la selección del modelo en términos relativos donde sea posible, es decir, ajustar una serie de modelos y confiar en los AIC y BIC para tomar decisiones acerca del modelo que represente mejor a nuestros datos. Las pruebas de hipótesis son usadas cuando se desea comparar modelos de transición anidados. También la parsimonia y la conceptualización teórica son importantes criterios de selección en un modelo LTA, tal como lo es en un modelo de clase latente.

Como dijimos anteriormente en el modelo LTA interesan estimar 3 conjuntos diferentes de parámetro.

2.3.2. Las prevalencias de los estados latentes

La prevalencia del estado latente c en el tiempo t se denota por δ_{ct} , y viene dada por la probabilidad de membresía al estado latente c en el tiempo t. En el ejemplo anterior hay 2 estados latentes en cada uno de los dos tiempos, y por tanto 4 prevalencias. Se tendrá una probabilidad de la membresía al estado latente 1 en el tiempo 1. Debido a que los estados latentes son mutuamente excluyentes y exhaustivos en cada periodo de tiempo, es decir, que cada individuo es miembro de uno y solo un estado latente en el tiempo t, se debe cumplir que:

$$\sum_{c_t=1}^{C} \delta_{c_t} = 1. \tag{2.19}$$

En otras palabras, dentro de un tiempo en particular t, la prevalencia en los estados latentes debe sumar 1.

2.3.3. Las probabilidades de respuesta al item

La probabilidad de obtener una respuesta $r_{j,t}$ a la variable observada j, condicionada a la membresía en el estado latente c_t en el tiempo t se denota mediante $\rho_{j,r_{j,t}|c_t}$. Para cada combinación del estado latente c_t de la variable observada j en el tiempo t, hay R_j probabilidades de respuesta al item. Nuevamente cada individuo proporciona una y sola una respuesta alternativa a la variable j en un tiempo específico t y por tanto se debe cumplir que:

$$\sum_{r_{j,t}=1}^{R_j} \rho_{j,r_{j,t}|s_t} = 1, \tag{2.20}$$

para todo j, t. En otras palabras, para los individuos en el estado latente c_t en el tiempo

t, las probabilidades de cada respuesta alternativa a la variable j deben sumar 1.

2.3.4. Las probabilidades de transición

La probabilidad de transición a un estado latente i en el tiempo T, condicionada a su membresia al estado latente j en el tiempo $T-1$ se denota mediante $\tau_{i|j}$. Las τ s por lo general son acomodadas en una matriz de probabilidades de transición de la siguiente manera:

$$\begin{bmatrix} \tau_{1|1} & \tau_{2|1} & \cdots & \tau_{c|1} \\ \tau_{1|2} & \tau_{2|2} & \cdots & \tau_{c|2} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{1|c} & \tau_{2|c} & \cdots & \tau_{c|c} \end{bmatrix}, \tag{2.21}$$

donde $\tau_{i|j} = P(L_T = i|L_{T-1} = j)$.

Los individuos que se encuentran en un estado latente c_T en el tiempo T, estarán en solo un estado latente en el tiempo $T+1$, c_{T+1} el cual puede ser el mismo estado latente c_t o puede ser uno diferente. En cada uno de los estados latentes de la membresía estos son mutuamente excluyentes y exhaustivos, es decir, los individuos pertenecen solo a un estado latente en cada tiempo. Por consiguiente se cumple:

$$\sum_{c_{T+1}=1}^{C} \tau_{c_{t+1}|c_t} = 1, \tag{2.22}$$

En otras palabras, cada fila de la matriz de probabilidades de transición debe sumar 1

2.3.5. Restricción de parámetros en el análisis de transición latente

Algunas restricciones en los parámetros pueden ser usadas en el modelo LTA. Estas se fijan en la prevalencia de los estados latentes, en las probabilidades de respuesta al item, o en las probabilidades de transición en el modelo de transición latente. Dos tipos diferentes de restricción de parámetros son comunmente usados: los parámetros pueden ser fijos o restringidos. Un parámetro que es fijo en cierto valor particular no es estimado. Antes de comenzar la estimación, su valor debe ser especificado en el rango de 0 a 1. Este valor fijo está fuera de los limites del procedimiento de estimación. Cuando los parámetros están restringidos, estos son ubicados en un conjunto equivalente junto con otros parámetros. La estimación de todos los parámetros en un conjunto equivalente está restringida a ser igual al mismo valor, el cual puede ser cualquier valor en el rango de 0 a 1. Un modelo simple de transición latente puede contener cualquier combinación de parámetros fijos y restringidos. Debido a que los parámetros fijos no se estiman, estos no contribuyen al número total de parámetros a estimar. El conjunto equivalente cuenta como un único parametro de estimación, independientemente de cuantos parámetros conforman el conjunto.

2.3.6. Restricciones en las probabilidades de respuesta al ítem

Es una buena idea restringir las probabilidades de respuesta al item en el modelo LTA para que estas sean iguales a través del tiempo donde sea posible y razonable hacerlo, es decir, donde quiera una medida de invarianza a través del tiempo puede ser supuesta. Existen razones conceptuales y practicas para hacer esto. La razón conceptual es que estos modelos de transición latente son mucho más faciles de interpretar si las probabilidades de respuesta al item son idénticas a través del tiempo. Se ha discutido mucho acerca de la medida de invarianza a través de los grupos y también como probar hipótesis acerca de la medida de invarianza. Si la probabilidades de respuesta al item son iguales a lo largo de los grupos entonces la interpretación de las clases latentes es identica a lo largo de los grupos también. Esto significa que cualquier diferencia en los grupos observados en la prevalencia de las clases latentes pueden ser interpretadas simplemente como diferencias cuantitativas; ciertas clases latentes serán más grandes en algunos grupos que en otros. Por otro lado, si las probabilidades de respuesta al item no son iguales a lo largo de los grupos, entonces el significado de las clases latentes varia a lo largo de los grupos. Si esto sucede las comparaciones en los grupos de las prevalencias en las clases latentes llegar a ser menos sencillas. Cuando se interpreta las diferencias en la prevalencia de clases latentes se hace necesario tener en cuenta cualquier diferencia en el significado de las clases latentes en el mismo tiempo. Desde luego, dependiendo de las preguntas de investigación, algunas veces las diferencias cualitativas pueden ser interesantes en si mismas. En el mismo sentido, la matriz de probabilidades de transición en el LTA es fácil de interpretar si hay medidas equivalentes a través de los tiempos. Si las probabilidades de respuesta al item son idénticas a través del tiempo, el significado de los estados latentes permanece constante a través del tiempo. Esto significa por ejemplo, que un elemento en la diagonal de la matriz de probabilidades de transición refleja la probabilidad de la membresia en el estado latente s en el tiempo $t+1$ condicionada a la membresia en el mismo estado latente s en el tiempo t. Para extender de que las probabilidades de respuesta al item correspondientes al estado latente s cambie a través del tiempo, el significado del estado latente s cambiará. Luego, ya no es tan claro como interpretar esta probabilidad de transición, porque junto con la interpretación cuantitativa también cambia a través del tiempo la membresia del estado latente. Es necesario entonces interpretar el cambio a través del tiempo en el sentido de las clases latentes. Sin embargo, dependiendo de las preguntas de investigación este cambio cualitativo puede ser muy interesante, particularmente si es significativo para su desarrollo.

La razón principal para restringir las probabilidades de respuesta al item a que sean iguales a través del tiempo es para ayudar a estabilizar la estimación y mejorar el problema de identificabilidad. En los modelos de transición latente pueden haber un gran número de probabilidades de respuesta al item, particularmente en modelos con más de dos mediciones en el tiempo. Imponer restricciones a los parámetros en el tiempo puede reducir considerablemente el número de parámetros a estimar.

2.4. Modelos de diagnóstico cognitivo

Los modelos de diagnóstico cognitivo han sido desarrollados para estudiar los procesos cognitivos subyacentes al aprendizaje y para proporcionar información formativa acerca de cada estudiante. En términos más técnicos, el diagnóstico cognitivo educacional es una evaluación usada para identificar los atributos cognitivos que los estudiantes necesitan tener para lograr un dominio en particular y para clasificarlos dentro de grupos de diagnóstico basados en los atributos que ellos poseen (Rupp et al.,2010). Un diagnóstico cognitivo permite conocer un perfil detallado de habilidades dominadas o no dominadas para cada examinado más alla de darnos un puntaje general como lo hace la teoría de respuesta al item. Además cada participante es clasificado dentro de una clase cognitiva en la cual todos los individuos tienen el mismo patrón en el dominio de los atributos.

Antes de usar la evaluación de diagnóstico cognitivo, es necesario primero identificar las variables latentes que serán medidas en una evaluación en particular, asi como los items que buscarán medir estas variables. Estas variables latentes han recibido diversos nombres en la literatura como por ejemplo: habilidades, atributos, rasgos latentes o perfiles latentes. Para especificar que variables son medidas con que items, la evaluación cognitiva parte de la llamada matriz Q:

$$Q = \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1K} \\ q_{21} & q_{22} & \cdots & q_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ q_{J1} & q_{J2} & \cdots & q_{JK} \end{bmatrix} \tag{2.23}$$

Esta matriz define la relación entre todos los J items y K atributos, nombre que en adelante usaremos para representar a las variables latentes. Especificamente, la matriz Q es una tabla de items por atributo que especifica si cada atributo es medido o no por un item a través de valores binarios 1s y 0s.

Se realiza un test construído por expertos en el área de interés quienes definen una matriz Q que indica los atributos requeridos para responder cada uno de los J items del test. Así, q_{jk} toma el valor de 1 si el atributo k es relevante para responder correctamente el item j y 0 en caso contrario. Por ejemplo, para un test con 10 items que evalua la presencia de 4 atributos, una matriz Q podría estar dada por:

$$Q = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix} \tag{2.24}$$

La fila 1 indica que para responder correctamente el item 1, es necesario que el individuo posea el segundo y cuarto atributo, más no el primero y tercero.

La creación de esta matriz es realizada comunmente por expertos en la materia, basándose en juicios teóricos. Cualquier error de especificación en la matriz Q puede causar errores en la clasificación de los examinados al considerarlos en los grupos de diagnóstico erróneos. Para minimizar este error de clasificación debido a la falsa especificación de la matriz Q y para mejorar la evaluación varios métodos de especificación han sido propuestos(Chiu, 2013, Close ,2012; J. Liu, Xu and Ying, 2012; Xu, 2013).
Diferentes modelos de diagnóstico cognitivo aparecen en la literatura, los cuales pueden clasificarse como compensatorios y no compensatorios. Los CDMs compensatorios permiten que una habilidad requerida para responder un item pueda ser compensada con otra habilidad adicional. Mientras, que los modelos no compensatorios requieren que cada una de las habilidades estén presentes para producir una respuesta correcta al item. Como un ejemplo de modelo no compensatorio tenemos al modelo DINA. De otro lado, como ejemplo de un modelo compensatorio podemos considerar el modelo DINO. Seguidamente, describiremos brevemente estos modelos y su generalización.

2.4.1. El modelo DINA

El modelo DINA(de la Torre y Douglas, 2004; Haertel, 1989; Junker and Sijtsma, 2001; Macready and Dayton, 1977) es un CDM no compensatorio que asume que la falta de un atributo no puede ser compensada por la existencia de otro atributo. Esta restricción significa que para responder correctamente a un item, un examinado necesita en teoría dominar todos los atributos requeridos para ese item. La principal restricción en el modelo DINA es que este no hace distinciones entre los evaluados que no dominaron uno o más atributos de los requeridos.

El modelo DINA estima la relación entre N individuos y las K habilidades basándose en el patrón de respuestas de los individuos a los items, el cual es expresado como una matriz de unos y ceros, donde 1 se asigna cuando el individuo responde correctamente al item j y 0 en caso contrario. Esta relación se resume en una matriz $N \times K$, llamada A y que se define por:

$$A = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1k} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{N1} & \alpha_{N2} & \dots & \alpha_{NK} \end{bmatrix}, \tag{2.25}$$

la cual posee al igual que Q una estructura binaria, es decir, con elementos 0 y 1, donde cada α_{ik} se define como:

$$\alpha_{ik} = \begin{cases} 1 & \text{, si el individuo } i \text{ domina la habilidad } k \\ 0 & \text{, en caso contrario} \end{cases} \tag{2.26}$$

A diferencia de la matriz Q, las entradas en la matriz A son latentes; es decir, no observables.
El modelo DINA del acrónimo en inglés (Deterministic Input Noisy AND gate) es de naturaleza conjuntiva, es decir, que para que un individuo i responda correctamente al item j necesitará de todas las habilidades que conciernen a tal item, el proceso tendrá como salida 1 si el individuo posee todas las habilidades requeridas para responder correctamente al item o 0 en caso contrario. Ello hace que sea necesario definir las siguientes variables dicotómicas:

$$\eta_{ij} = \begin{cases} 1 & \text{, si el individuo } i \text{ domina todas las habilidades que} \\ & \text{se necesitan para responder correctamente el item } j \\ 0 & \text{, en caso contrario} \end{cases} \tag{2.27}$$

y se cumple que $\eta_{ij} = \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}}$.

Además, la denominación de ruido atribuída al modelo se da pues, pudiera darse el caso que los individuos que dominan todas las habilidades necesarias para un item fallen y aquellos que no las dominan todas acierten. Esto introduce dos probabilidades de error(parámetros), los cuales al definirse la respuesta binaria del individuo i a un item j por Y_{ij} vienen dadas por:

- El parámetro de desliz del item j (s_j) es definido como la probabilidad de responder incorrectamente a un item j a pesar de que el examinado domine todos los atributos requeridos para hacerlo, es decir:

$$s_j = P(Y_{ij} = 0|\eta_{ij} = 1). \tag{2.28}$$

- El parámetro de adivinación (g_j) es definido como la probabilidad de responder correctamente a un item j a pesar de no dominar todas las habilidades requeridas para hacerlo:

$$g_j = P(Y_{ij} = 1|\eta_{ij} = 0). \tag{2.29}$$

Además de la estimación de estos parámetros el modelo DINA requiere estimar las probabilidades $\pi_c = P(\boldsymbol{\alpha_i} = \boldsymbol{\alpha_c})$ de pertenencia de cada individuo i a la clase o perfil latente c. En adelante denotaremos por L a la variable aleatoria que nos indica la clase de pertenencia de un individuo.

El modelo DINA define la probabilidad de que un individuo i, que pertenece a la clase c, responda correctamente a un item j mediante:

$$P_{cj} = P\left(Y_{ij} = 1|\alpha_i = \alpha_c, \theta_j\right) = \left(1 - s_j\right)^{\eta_{ij}} g_j^{1-\eta_{ij}}, \tag{2.30}$$

donde $\theta_j = (g_j, s_j)$ es un vector fila de las probabilidades de error del item ya definidos anteriormente. En general, asumiendo independencia condicional y usando la ecuación anterior, la probabilidad de observar la respuesta $x_{ij} \in \{0, 1\}$ en el individuo i, que pertenece a la clase c, viene dada por:

$$P(Y_{ij} = y_{ij} \quad |\alpha_i = \alpha_c, \theta_j) = p_{cj}^{y_{ij}} (1 - p_{cj})^{1-y_{ij}}, \tag{2.31}$$

donde P_{cj} viene dada por la expresión (2.30). Utilizando el teorema de probabilidad total y la ecuación (2.31), se deduce que la probabilidad de obtener por parte del individuo i un patrón de respuestas $\boldsymbol{y} = [y_1, y_2, \ldots, y_J]$ dado los parámetros de todos los items, que resumimos en un vector θ, y las probabilidades de pertenencia a las clases latentes $\pi = (\pi_1, \pi_2, \ldots, \pi_C)$, está dada por:

$$P\left(\boldsymbol{Y}_i = \boldsymbol{y}|\boldsymbol{\alpha}_i, \boldsymbol{\pi}, \boldsymbol{\theta}\right) = \sum_{c=1}^{C} \pi_c \left[\prod_{j=1}^{J} P\left(Y_{ij} = y_j|\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_c, \boldsymbol{\theta}_j\right)\right] \tag{2.32}$$

Con esto se obtiene finalmente la siguiente función de verosimilitud de observar una muestra de respuestas de N individuos a los J items:

$$L(\boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{\theta}) = \prod_{i=1}^{N} P\left(\boldsymbol{Y_i} = \boldsymbol{y}_i | \boldsymbol{\alpha}_i, \boldsymbol{\pi}, \boldsymbol{\theta}\right) = \prod_{i=1}^{N} \left\{ \sum_{c=1}^{C} \pi_c \left[\prod_{j=1}^{J} P_{cj}^{y_{ij}} \left(1 - P_{cj}\right)^{1-y_{ij}} \right] \right\}, \tag{2.33}$$

donde $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_C)$ y $\boldsymbol{y_i} = (y_{11}, \ldots, y_{i_J})$ denota al patrón de respuestas observado para el examinado i.

2.4.2. El modelo DINO

El modelo DINO (del acrónimo en inglés deterministic input, noisy or gate) es un modelo de diagnostico cognitivo compensatorio(Templin,2004; Templin, et al.,2006), porque asume que la falta de un atributo puede ser compensada por otro. En otras palabras, el dominar al menos un atributo compensa el deficit de no dominar todos los otros atributos medidos. De la misma manera que en el modelo DINA, los parámetros de adivinación y desliz son estimados en los niveles de los items. El modelo DINO se caracteriza por emplear una regla de compensación disyuntiva en la cual la presencia de al menos una medida del atributo garantiza una probabilidad alta de aprobación del item. El modelo DINO estima la probabilidad de respuesta correcta al item j por parte de un individuo i que pertenece a la clase latente c de la siguiente manera:

$$P_{cj} = P\left(Y_{ij} = 1 | \alpha_i = \alpha_c, \theta_j\right) = \left(1 - s_j\right)^{\omega_{ij}} g_j^{(1-\omega_{ij})}, \tag{2.34}$$

donde:

$$\omega_{ij} = 1 - \prod_{k=1}^{K} \left(1 - \alpha_{ik}\right)^{q_{jk}}. \tag{2.35}$$

En caso de que el atributo k no sea medido por el item j, q_{jk} tomará el valor de 0 y en consecuencia el valor de $1 - \alpha_{ik}$ no importaría. Por otra parte, si el atributo k es medido por el item j, q_{jk} tomará el valor de 1, y en consecuencia $1 - \alpha_{ik}$ se toma en cuenta para el valor final que ω_{ij} tome. Si el respondiente en la clase latente i domina el atributo k, α_{ik} toma el valor de 1, y de este modo $1 - \alpha_{ik}$ será igual a 0. Sin embargo, si el respondiente en la clase latente i no domina el atributo k, entonces $1 - \alpha_{ik}$ es 1. Debido a que la ocurrencia de $\omega_{ij} = 1$

depende de la existencia de al menos un 0 en el término de multiplicación, el dominio de al menos un atributo aumenta significativamente la probabilidad de acertar el item. El modelo DINO es usado cuando solo un atributo se necesita para dominar la habilidad a diferencia de otros que necesitan más de un atributo(Rupp et al., 2010). Los parámetros de adivinación g_j y desliz s_j en el modelo DINO se definen de la misma manera que en el modelo DINA.

2.4.3. El modelo G-DINA

Como se comentó anteriormente los modelos DINA y DINO dividen para cada item j a las personas en dos grupos. En el modelo DINA por ejemplo, un primer grupo está compuesto por los individuos que presentan todos los atributos requeridos para dar una respuesta correcta a este item, y el segundo grupo formado por el resto de individuos. El modelo G-DINA (del acrónimo en inglés Generalized Deterministic Input Noisy and gate) cuestiona la suposición de igual probabilidad de responder correctamente para los individuos del segundo grupo y propone una generalización. Así, en lugar de formar solo dos grupos, el modelo G-DINA forma $2^{K_j^*}$ grupos, donde K_j^* es el número de atributos requeridos para el item j. Por ejemplo cuando $K_j^* = 2$, en lugar de 2, se crean 4 grupos latentes, donde cada uno de ellos pueden tener diferentes probabilidades de éxito.

El modelo generalizado de entrada determinística con ruido y salida (G-DINA) parte de un contexto en el que N individuos denotados por $i = 1, 2, \ldots, N$, son examinados mediante un test con J items denotados por $j = 1, 2, \ldots, J$, cada uno de los cuales requieren la presencia de K_j^* atributos para ser respondidos correctamente, donde $K_j^* \leq K$, siendo K la totalidad de atributos evaluados por el test.

De los resultados del test obtenemos una matriz observable binaria $y = [y_{ij}]$ de orden $N \times J$ que representa las respuestas de los N entrevistados a los J items. De tal manera, que x_{ij} toma el valor 1 si el individuo respondió correctamente al item j y 0 en caso contrario.

$$y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1J} \\ y_{21} & y_{22} & \cdots & y_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ y_{N1} & y_{N2} & \cdots & y_{NJ} \end{bmatrix}. \tag{2.36}$$

Para cada individuo i se asume que existe un vector latente $\alpha_i = [\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{ik}]$ tal que $\alpha_{ik} = 1$ si el individuo i posee el atributo k, y $\alpha_{ik} = 0$ en caso contrario. Estos, que llamaremos perfiles latentes, son desconocidos y deben ser estimados. Para cada item j, se separan a los individuos en $2^{k_j^*}$ grupos latentes (l), donde $K_j^* = \sum_{k=1}^{K} q_{jk}$ representa el número de atributos requeridos para el item j. Así, cada individuo pertenecerá a solo una de estas clases. La presencia o ausencia de los atributos que no se requieren para responder correctamente este item, no afecta la pertenencia de un individuo a una u otra clase.

Retornando al ejemplo en la matriz Q dada en (2.22), vemos que el item 3 solo requiere de la presencia de 3 atributos por lo que para este item se forman $2^3 = 8$ grupos latentes.

Con el fin de simplificar la notación y sin perder generalidad, consideraremos siempre que son los primeros K_j^* atributos los requeridos para responder correctamente el item j, y que correspondientemente a_{lj}^* es el vector binario de dimensión K_j^* que contiene 1s solo si un individuo cualquiera de la clase l posee tales componentes para responder correctamente el item j. Por ejemplo, el item 10 en la matriz solo requiere dos atributos para ser respondido correctamente, por lo que el vector a_{lj} toma el valor de $a_{lj}^* = (a_{lj1}, a_{lj2})$, en lugar del vector completo $(a_{lj1}, a_{lj2}, a_{lj3}, a_{lj4})$

En modelos más sencillos como el DINA, carecer de un atributo requerido para determinado item, es lo mismo que carecer de todos los atributos requeridos. Sin embargo, esto podría no ser siempre cierto, ya que un individuo que posee alguno de los K_j^* atributos requeridos para el item j, podría tener mayor probabilidad de responder correctamente que aquel que no tiene ninguno. El modelo G-DINA relaja esta hipótesis de igual probabilidad de éxito.

Por esta razón es importante establecer una relación entre los vectores α_{lj}^* y $\alpha_{l'j}^*$ que denotan a los estados de conocimiento o perfiles de atributos de 2 sujetos en las clases l y l'. Así, para esta investigación, diremos que $\alpha_{lj}^* < \alpha_{l'j}^*$, si α_{lj}^* posee menos atributos de los requeridos para el item j que $\alpha_{l'j}^*$, es decir, $\sum_{k=1}^{K_j^*} \alpha_{ljk}^* < \sum_{k=1}^{K_j^*} \alpha_{l'jk}^*$
En el ejemplo de la matriz Q, para el item 3, que requiere la presencia de tres atributos, podemos afirmar que el vector de estado $a_{l3}^* = (0, 0, 1)$ es menor que el vector $a_{l'3}^* = (1, 1, 0)$, ya que posee una menor cantidad de atributos requeridos.

En adelante denotaremos a la probabilidad de que los entrevistados con perfil de atributos α_{lj}^* respondan el item j correctamente por: $P\left(X_j = 1|\alpha_{lj}^*\right) = P\left(\alpha_{lj}^*\right)$, siendo X_j la variable aleatoria que denota la respuesta de un entrevistado al item j. Como es natural estas probabilidades deberán de satisfacer que $P\left(\alpha_{lj}^*\right) \leq P\left(\alpha_{l'j}^*\right)$ cuando $\alpha_{lj}^* < \alpha_{l'j}^*$. El modelo G-DINA plantea para estas probabilidades la siguiente ecuación:

$$P\left(\alpha_{lj}^*\right) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk}\alpha_{lk} + \sum_{k=1}^{K_j^*-1} \sum_{k'=k+1}^{K_j^*} \delta_{jkk'\alpha_{lk}\alpha_{lk'}} + \ldots + \delta_{j12\ldots K_j^*} \prod_{k=1}^{K_j^*} \alpha_{lk}, \tag{2.37}$$

donde $\delta_j = \left(\delta_{j0}, \delta_{j1}, \ldots, \delta_{jK_j^*}, \delta_{j12}, \ldots, \delta_{j12\ldots K_j^*}\right)$, representa un vector de parámetros a estimarse para el item j, δ_{j0} es un parámetro de intercepto (concretamente la probabilidad de que los encuestados respondan correctamente al item j cuando no posean ninguno de los atributos requeridos para responder satisfactoriamente este item), δ_{jk} representa el efecto principal debido al atributo k (el incremento marginal en la probabilidad de responder correctamente al item j como resultado de la presencia del atributo k); δ_{jkkj} representa el efecto de segundo orden, (el incremento marginal en la probabilidad de obtener una respuesta correcta debido a la presencia de los atributos k y k_j); y así sucesivamente hasta $\delta_{j12\ldots k_j^*}$ que representa el cambio en la probabilidad de obtener una respuesta correcta debido a la presencia de todos los atributos requeridos. Uno de los aspectos que resaltan la importancia del modelo G-DINA, es que además de medir la contribución marginal de poseer un atributo particular en las probabilidades de responder correctamente un item, puede medir también (si existiera) el efecto conjunto de dos o más atributos en ella.

Capítulo 3

Modelos de transición de clases latentes cognitivas

Según **?** el modelo de transición de clases latentes cognitivas permite investigar hipótesis acerca del efecto de una intervención mediante el cambio en las probabilidades de transición antes y después de esta con relación al estado cognitivo. Esta clase de modelos es una excelente manera de estudiar cambios en el desarrollo y crecimiento, tales como las transiciones entre los estados descritos en la teoria de Piaget, según lo citado por **?**. Suponga que observamos J variables categóricas manifiestas politómicas $Y_1, Y_2, ..., Y_j$, donde Y_j tiene R_j posibles resultados. Se asume que estas variables deben ser la respuesta de un individuo a una evaluación con J items. El modelo de transición de clases latentes con solo dos puntos en el tiempo ($T = 2$), expresa que la distribución de probabilidad conjunta de las respuestas $Y_1 = [Y_{1,1}, Y_{2,1}, ...Y_{J,1}]$, $Y_2 = [Y_{1,2}, Y_{2,2}, ...Y_{J,2}]$ de cualquier individuo, donde $Y_{j,t}$ denota a su respuesta al item j en el tiempo t, viene dada por:

$$
\begin{aligned}
&P\left[\mathbf{Y}_1 = \mathbf{y}_1, \mathbf{Y}_2 = \mathbf{y}_2\right] = \\
&\quad = \sum_{c_1=1}^{C} \sum_{c_2=1}^{C} P\left(Y_1 = \mathbf{y}_1, Y_2 = \mathbf{y}_2 | C_1 = c_1, C_2 = c_2\right) P\left(C_2 = c_2 | C_1 = c_1\right) P\left(C_1 = c_1\right) &(3.1)\\
&\quad = \sum_{c_1=1}^{C} \sum_{c_2=1}^{C} \delta_{c_1} \tau_{c_2|c_1} P\left(Y_1 = \mathbf{y}_1 | C_1 = c_1\right) P\left(Y_2 = \mathbf{y}_2 | C_2 = c_2\right) &(3.2)\\
& &(3.3)
\end{aligned}
$$

$$= \sum_{c_1=1}^{C}\sum_{c_2=1}^{C}\delta_{c_1}\tau_{c_2|c_1}\prod_{j=1}^{J}\prod_{r_{j,1}=1}^{R_j}\rho_{j,r_{j,1}|c_1}^{I(y_{j,1}=r_{j,1})}\prod_{j=1}^{J}\prod_{r_{j,2}=1}^{R_j}\rho_{j,r_{j,2}|c_2}^{I(y_{j,2}=r_{j,2})}. \tag{3.4}$$

$$= \sum_{c_1=1}^{C}\sum_{c_2=1}^{C}\delta_{c_1}\tau_{c_2|c_1}\prod_{t=1}^{2}\prod_{j=1}^{J}\prod_{r_{j,t}=1}^{R_j}\rho_{j,r_{j,t}|c_t}^{I(y_{j,t}=r_{j,t})}, \tag{3.5}$$

donde

$$\sum_{c_1=1}^{C}\delta_{c_1}=1, \sum_{c_2=1}^{C_L}\tau_{c_2|c_1}=1, \sum_{r_{j,t}=1}^{R_j}\rho_{j,r_{j,t}|c_t}=1, \tag{3.6}$$

δ_{c1} la probabilidad de pertenencia del individuo a la clase c_1 en el tiempo 1, $\tau_{c2|c1}$ es la probabilidad de transición de que el individuo cambie a una clase c_2 en el tiempo 2 dado que él ha pertenecido a una clase c_1 en el tiempo 1, $I(y_{j,t}=r_{j,t})$ es una variable indicadora que es igual a 1 si la respuesta al item j en el tiempo t es $r_{j,t}$ y 0 en caso contrario, y $\rho_{j,r_{j,t}|c_t}$, es la probabilidad de que la respuesta del individuo al item j en el tiempo t sea $r_{j,t}$ condicionado a su membresia a la clase latente c_t.
La consideración de un modelo con más de dos puntos en el tiempo es inmediata pero implica el incremento del número de parámetros, no sólo en términos de las probabilidades de transición, sino también en los parámetros dentro de las probabilidades de respuesta al item del modelo para $\rho_{j,r_{j,t}|c_t}$. Este modelo se presenta a continuación:

$$P\left[\mathbf{Y}=\mathbf{y}\right] = \sum_{c_1=1}^{C}\cdots\sum_{c_T=1}^{C}\delta_{c_1}\tau_{c_2|c_1}^{(1)}\cdots\tau_{c_T|c_{T-1}}^{(T)}\prod_{t=1}^{T}\prod_{j=1}^{J}\prod_{r_{j,t}=1}^{R_j}\rho_{j,r_{j,t}|c_t}^{I(y_{j,t}=r_{j,t})}, \tag{3.7}$$

donde $\tau_{a|b}^{T}$ representa la probabilidad de transición de que un individuo en la clase b en el tiempo $t-1$ pase a la clase a en el tiempo t
En este capítulo detallaremos los modelos longitudinales DINA y DINO para clasificar a los examinados con respecto al dominio de atributos en múltiples períodos de medición. Estos modelos clasifican a los examinados en los estados latentes a través de mediciones consecutivas y estiman las probabilidades de transición entre clases de los examinados desde el tiempo 1 hasta el tiempo T.
El modelo de transición de clases latentes cognitivas DINA que denotaremos por LTA-DINA, estima la probabilidad de que un individuo presente un patrón particular de respuestas como una función de sus probabilidades de membresía a cada clase cognitiva en el tiempo 1, las probabilidades de transición entre las clases cognitivas y las probabilidades de observar una respuesta en cada punto en el tiempo condicionado a su clase cognitiva de membresía. El modelo LTA-DINA integra el modelo LTA con el DINA a través de:

$$
\begin{aligned}
&P\left[\mathbf{Y}_1 = \mathbf{y}_1, \mathbf{Y}_2 = \mathbf{y}_2, ..., \mathbf{Y}_T = \mathbf{y}_T\right] = \\
&= \sum_{c_1=1}^{C} \cdots \sum_{c_T=1}^{C} \delta_{c_1} \tau_{c_2|c_1} \ldots \tau_{c_T|c_{T-1}} \prod_{t=1}^{T} \prod_{j=1}^{J} \left(1 - P_{c_{t_j}}\right)^{I(y_{j,t}=0)} P_{c_t j}^{I(y_{j,t}=1)},
\end{aligned} \tag{3.8}
$$

donde $P_{c_t j} = (1 - s_j)^{\eta_{c_t j}} g_j^{(1-\eta_{c_t j})}$ es la probabilidad dada en (2.30) y $\eta_{c_t j}$ es el estado del conocimiento de los individuos en la clase c_t para el item j

De la misma manera, podemos definir el modelo de transición DINO, que lo denotaremos por LTA-DINO mediante:

$$
P\left[\mathbf{Y} = \mathbf{y}\right] = \sum_{c_1=1}^{C} \cdots \sum_{c_t=1}^{C} \delta_{c_1} \tau_{c_2|c_1} \ldots \tau_{c_t|c_{t-1}} \prod_{t=1}^{T} \prod_{j=1}^{J} \left(1 - P_{c_{t_j}}\right)^{I(y_{j,t}=0)} P_{c_t j}^{I(y_{j,t}=1)}, \tag{3.9}
$$

donde $P_{c_t j} = (1 - s_j)^{\omega_{c_t j}} g_j^{(1-\omega_{c_t j})}$ es la probabilidad dada en (2.34) y $\omega_{c_t j}$ es el estado de conocimiento para el modelo DINO, la cual lo diferencia del modelo DINA cuyo estado de conocimiento es $\eta_{c_t j}$.

Adicionalmente, a la estimación de estos parámetros, será de interés estimar las probabilidades de pertenencia aposteriori para determinar las clases para cada individuo en cada punto en el tiempo. Para el uso práctico de los modelos longitudinales DINA y DINO, los parámetros de interés son probabilidades aposteriori y las probabilidades de transición. Estas brindan información individual y son usadas para evaluar el rendimiento de cada estudiante basados en su clase grupal estimada. Por otro lado, las probabilidades de transición consideran las probabilidades de permanencia en la misma clase o el movimiento entre las clases de un tiempo t a un tiempo $t + 1$ para los examinados que están en la misma clase cognitiva, quienes nos brindan una idea general del éxito del tratamiento aplicado entre las 2 mediciones.

3.1. Estimación del modelo

El modelo de de transición latente en combinación con el modelo de diagnóstico cognitivo DINA(LTA-DINA)nos pemite analizar el comportamiento de los evaluados en las diferentes clases cognitivas a través de las probabilidades de transición. Para la estimación de este modelo desarrollaremos el método de máxima verosimilitud via el algoritmo de Esperanza-Maximización (EM). La implementación de este se encuentra en el paquete Mplus (Version 7; Muthén y Muthén, 2017).

Los parámetros en un modelo de transición latente incluyen la probabilidad de la membresia en el tiempo 1, las probabilidades de transición y los parámetros involucrados en la probabilidad de respuesta al item condicionadas a la clase latente.

La probabilidad conjunta de un individuo, que pertenece al conjunto de clases $\mathbf{c}$ =

$(c_1, \ldots, c_T)$ en el tiempo, proporcione un cierto patrón de respuestas es:

$$P\left[\mathbf{Y}_1 = \mathbf{y}_1, ..., \mathbf{Y}_T = \mathbf{y}_T | C{=}c\right] = \left[\delta_{c1} \prod_{t=2}^{T} \tau^{(t)}_{c_t|c_{t-1}}\right] \times \left[\prod_{t=1}^{T} \prod_{j=1}^{J} \prod_{r_j=1}^{R_j} \rho_{j,r_j,t|c_t}^{I(y_{j,t}=r_{j,t})}\right], \qquad (3.10)$$

donde $\delta_{c_1} = P\left[C_1 = c_1\right]$,$\tau^{(t)}_{c_t|c_{t-1}} = P\left[C_t = c_t | C_{t-1} = c_{t-1}\right]$ y $\rho_{j,r,j|c_t} = P\left[Y_{jt} = y_{j,t} | C_t = c_t\right]$. Se asume que $Y_{1t}, ..., Y_{Jt}$ son condicionalmente independientes dentro de cada clase c_t para $t = 1, ..., T$. Este supuesto, llamado de independencia local nos permite realizar inferencias acerca de la variable de clase latente.(Lazarsfeld PF et al.,1968). También se asume que la secuencia $\{C_t\}$ donde C_t denota a la clase de pertenencia de un individuo en el tiempo t, constituye una cadena de Markov de primer orden para $t = 2, ..., T$. En la ecuación (3.10) solo la probabilidad marginal de la clase de la membresia en el tiempo de inicio $t = 1$, δ_{c_1} es estimada; las probabilidades marginales de la clase de membresía en el tiempo $t(\geqslant 2)$ no son estimadas directamente sin embargo, están en función de otros parámetros. La prevalencia marginal de cada clase en el tiempo $t(\geqslant 2)$ es calculado mediante:

$$\delta^{(t)}_{c_t} = P\left[C_t = c_t\right] = \sum_{c_1=1}^{C} \cdots \sum_{c_{t-1}=1}^{C} \delta_{c_1} \prod_{j=2}^{J} \tau^{(t)}_{c_t|c_{t-1}}. \qquad (3.11)$$

Entonces la función de verosimilitud para un individuo o distribución conjunta incondicional de su respuesta es:

$$P\left[\mathbf{Y}_1 = \mathbf{y}_1, \ldots, \mathbf{Y}_T = \mathbf{y}_T\right] = \sum_{c_1=1}^{C} \cdots \sum_{c_T=1}^{C} \left[\delta_{c1} \prod_{t=2}^{T} \tau^{(t)}_{c_t|c_{t-1}}\right] \times \left[\prod_{t=1}^{T} \prod_{j=1}^{J} \prod_{r_j=1}^{R_j} \rho_{j,r_j,t|c_t}^{I(y_{j,t}=r_{j,t})}\right] \quad (3.12)$$

Por simplicidad consideremos en este trabajo una muestra de n individuos quienes responden a J items binarios medidos en dos períodos de tiempo. Consideremos aqui el modelo restringido de transición de clases latentes donde las probabilidades de respuesta al item son ajustadas para ser iguales durante el tiempo, aunque una extensión del modelo LTA sin restricciones es sencilla. En nuestro caso la función de verosimilitud en (3.12) se reduce a:

$$P\left[\mathbf{Y}_1 = \mathbf{y}_1, \mathbf{Y}_2 = \mathbf{y}_2\right] = \sum_{c_1=1}^{C} \sum_{c_2=1}^{C} \delta_{c_1} \tau_{c_2|c_1} \prod_{t=1}^{2} \prod_{j=1}^{J} \prod_{r_j=1}^{2} \rho_{j,r_j|c_t}^{I(y_{j,t}=r_{j,t})}, \qquad (3.13)$$

donde $\tau_{c_2|c_1} = P\left[C_2 = c_2 | C_1 = c_1\right]$. En (3.13), los parámetros libres son $\theta = (\delta, \tau_1, \ldots, \tau_C, \rho_1, \ldots, \rho_C)$, donde $\delta = (\delta_1, \ldots, \delta_{C-1})$, $\tau_c = \left(\tau_{1|c}, \ldots, \tau_{C-1|C}\right)$ y $\boldsymbol{\rho}_c = \left(\rho_{11}|c, \cdots, \rho_{J1|c}\right)$ para $c = 1, \ldots, C$

Bajo condiciones normales, los estimadores de máxima verosimilitud para θ resuelven la función score, $\partial \log \prod P\left[\mathbf{y}_1, \mathbf{y}_2\right] / \partial\theta = 0$. Al igual que muchas mezclas finitas, los estimadores de máxima verosimilitud para el LTA pueden ser estimados usando un algoritmo de optimización.

La función de verosimilitud será maximizada usando el algoritmo EM (Esperanza-Maximización). El software Mplus (version 7) de (Muthen y Muthen, 2017) será usado para implementar la estimación del modelo LTA-DINA.

Para el paso E de esperanza, calculamos la probabilidad condicional de que cada individuo es miembro de la clase c_1 en el tiempo $t = 1$ y de la clase c_2 en el tiempo $t = 2$ dadas las respuestas al item $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2)$ y de las estimaciones actuales $\hat{\theta}$ para los parámetros,

$$\hat{\eta}(c_1, c_2) = P\left[C_1 = c_1, C_2 = c_2 | \mathbf{y}_1, \mathbf{y}_2\right] = \frac{\delta_{c_1} \tau_{c_2|c_1} \prod_t \prod_j \prod_{r_j} \rho_{j,r_j|c_t}^{I(y_{j,t}=r_{j,t})}}{\sum_{c_1} \sum_{c_2} \delta_{c_1} \tau_{c_2|c_1} \prod_t \prod_j \prod_{r_j} \rho_{j,r_j|c_t}^{I(y_{j,t}=r_{j,t})}}. \tag{3.14}$$

En el paso M de maximización, actualizamos las estimaciones de los parámetros por

$$\hat{\delta}_{c_1} = \frac{\hat{n}_{c_1}^{(1)}}{n}, \quad \hat{\tau}_{c_2|c_1} = \frac{\hat{n}_{(c_1,c_2)}}{\hat{n}_{c_1}^{(1)}}, \quad \hat{\rho}_{j,r_j|c} = \frac{\hat{n}_{j,r_j|c}^{(1)} + \hat{n}_{j,r_j|c}^{(2)}}{\hat{n}_c^{(1)} + \hat{n}_c^{(2)}}, \tag{3.15}$$

donde:

$$\hat{n}_{(c_1,c_2)} = \sum_j \hat{\eta}_{j(c_1,c_2)}, \hat{n}_{c_1}^{(1)} = \sum_{c_2} \hat{n}_{(c_1,c_2)}, \hat{n}_{c_2}^{(2)} = \sum_{c_1} \hat{n}_{(c_1,c_2)}$$
$$\hat{n}_{j,r_j|c}^{(1)} = \sum_{c_2} \sum_j I\left(y_{j,1} = r_j\right) \hat{\eta}_{(c,c_2)}, \text{ y } \hat{n}_{j,r_j|c}^{(2)} = \sum_{c_1} \sum_j I\left(y_{j,2} = r_j\right) \hat{\eta}_{j(c_1,c)}.$$

La iteración entre estos dos pasos produce una secuencia de parámetros estimados que convergen confiablemente a un máximo local o global de la función de verosimilitud. Los desafios para la inferencia de máxima verosimilitud en muestras pequeñas para el modelo de transición de clases cognitivas se encuentran principalmente debido a los parámetros que han de ser estimados en los límites de un espacio paramétrico(es decir, 0 o 1), lo que dificulta la obtención adecuada de los errores estándar. Aunque las probabilidades de respuesta al item son cercanas a cero o a uno, éstas son altamente deseables desde una perspectiva de medición, cuando algunos de estos parámetros son estimados en el límite, es imposible obtener los errores estándar para la inversa de la matriz hessiana.

El LTA-DINA es ajustado para los datos de la prueba FOC usando el paquete Mplus. Los fueron estimados con 15 puntos de integración, con un criterio de convergencia igual a 10^{-7} para el cambio absoluto en la log-verosimilitud y el numero máximo de iteraciones del EM fué de 100, utilizandose además la descomposición de Cholesky y una cuadratura adaptativa con un tipo de integración estándar y modelo de mixtura con el link logit.

3.2. Inferencia del modelo

A pesar de que las estimaciones por EM y MCMC han hecho al modelo LTA muy popular, la función de verosimilitud para el modelo LTA puede tener características inusuales las cuales pueden afectar seriamente la inferencia. Por ejemplo, puede haber ciertas regiones continuas del espacio de parámetros para los cuales la log verosimilitud es constante, lo que conduce a la indeterminación de algunos parámetros. Para modelos LTA de dos clases se supone que la prevalencia de la clase 1 es cero en el tiempo 1 pero luego la clase 1 emerge al tiempo 2 ($\delta_1 = 0$ y $\sum_{l_1} \delta_{l_1} \tau_{1|l_1} > 0$).

3.3. Criterios para la selección del modelo

Varios estadisticos de ajuste relativos están disponibles para comparar dos o más modelos. En general, los indices de Criterio de información o estadisticos de prueba LR pueden ser usados dentro del contexto del modelamiento de variables latentes. El primero es apropiado para comparar modelos anidados o no anidados, mientras que el otro solo puede ser utilizado para modelos anidados. En este estudio nos basamos en los indices de criterio de información, especificamente en el AIC y BIC. A menudo múltiples indices se usan para comparar el ajuste de los datos al modelo entre un conjunto de modelos para los mismos datos cuando los estimadores de máxima verosimilitud de los parámetros se han obtenido. Los indices de Criterio de información están basados en una forma de penalización de la función de verosimilitud. Valores pequeños de estos indices indican un mejor ajuste. Sin embargo, diferentes indices podrián seleccionar diferentes modelos de ajuste para los mismos datos debido a las diferencias en la función de penalización aplicada a la verosimilitud. El indice AIC está dado por:

$$\mathrm{AIC} = -2 \log L + 2p \tag{3.16}$$

donde p es el número de parámetros estimados y $2p$ es usada como una penalización para la sobreparametrización y L es la función de verosimilitud. Para los modelos de diagnóstico cognitivo, L representa el valor de la función de verosimilutud marginal del modelo, y p

comprende el número total de los parámetros de los itemes y los parámetros estructurales. Uno de los problemas con el AIC es que tiende a seleccionar modelos más complejos. La falta de penalización para el tamaño de muestra conduce a la inconsistencia en el desempeño o actuación del AIC y una tendencia a sobreestimar el correcto número de clases(McLachlan & Peel, 2000). Sin embargo, el AIC es correcto y efciente asintóticamente, si el verdadero modelo no está entre los modelos que se comparan. (Vrieze, 2012). Para tener en cuenta el tamaño de la muestra, el BIC puede ser usado y viene dado por:

$$\mathrm{BIC} = -2\log L + p\ln(N) \tag{3.17}$$

donde L es la verosimilitud del modelo estimado con p parámetros libres y $ln(N)$ es la función logaritmo del tamaño de muestra total N. Como podemos observar en la ecuación 3.18, la función de penalización para el BIC está basada en el número de parámetros estimados asi como también en el tamaño de muestra. El BIC tiende a aplicar mayor penalización a la función de verosimilitud que el AIC cuando modelos complejos son estimados. Como consecuencia, el BIC es más preferido a la hora de seleccionar modelos simples que el AIC debido a la inclusión del tamaño de muestra en la función de penalización. La penalización del BIC con N hace de la significación estadistica más y más dificil de lograr a su vez que el tamaño de muestra aumenta(Vrieze, 2012, p. 233). Como sabemos, a diferencia del AIC, el BIC está hecho para ser asintóticamente consistente(Shao, 1997), lo que significa que a medida que el tamaño de muestra aumenta, el BIC tiende a seleccionar el número correcto de clases mixtas consistentemente si el verdadero modelo está entre los modelos que son comparados. Por el contrario, a medida que el tamaño de muestra aumenta, el AIC tiende a seleccionar un modelo más complejo aun cuando el verdadero modelo esté dentro de los candidatos. Un estudio de Shao acerca de la selección de indices de modelos, muestra que la utilidad del AIC y el BIC dependen básicamente de la estructura del modelo. El BIC se espera que actue mejor cuando el modelo verdadero tiene una estructura simple, y el AIC se espera que actue mejor cuando el modelo verdadero tiene una estructura compleja. Adicionalmente, con respecto al modelo verdadero, la utilidad del AIC y el BIC podría depender de varios factores, incluyendo a la función de pérdida, al estudio del diseño y la pregunta de investigación. En modelos de mixtura de variables latentes, la complejidad del modelo verdadero, la separación de las clases y la proporción de las clases han sido demostradas que afectan la utilidad de estos dos indices de ajuste (Lubke & Neale, 2006; Nylund et al., 2007; Vrieze, 2012). Como mencionamos anteriormente, el BIC actua mejor que el AIC dentro del contexto de modelos de variables latentes.(Jedidi et al., 1997; Li et al., 2009; Nylund et al., 2007; Preinerstorfer & Formann, 2012).

Capítulo 4

Estudio de Simulación

En este capítulo presentaremos un ejemplo simulado de la aplicación LTA-DINA. Con el último modelo como modelo de medición para evaluar el cambio posterior a una intervención instruccional educativa. El estudio se basa en un experimento diseñado multi etápico para evaluar los efectos del tratamiento denominado EAI (instrucción anclada mejorada) en el rendimiento en matemáticas en escuelas primarias con alumnos que presentan dificultades de aprendizaje y con aquellos que no la presentan. Hubieron dos etapas. Una primera aplicación se realizó en el primer semestre académico seguida de una segunda aplicación en el mismo año académico. Los resultados para la segunda evaluación del FOC (Fraction of the Cost) se presentarán en esta simulación. El impacto de la intervención EAI en los estudiantes de matemática fué examinada al evaluar las diferentes probabilidades de transición seguidas de cada una de las dos aplicaciones instruccionales y evaluando también las diferentes probabilidades de transición entre los diferentes grupos de examinados.

Generación de los datos y diseño del estudio

Los estudiantes para este estudio fueron seleccionados de seis salones de matemáticas en cierto colegio de una provincia. La muestra consistirá de 50 varones y 59 mujeres, todos ellos en el sétimo grado. Nueve de los estudiantes fueron tratados con dificultades de aprendizaje mientras que el resto no presentan estas dificultades. La prueba FOC fue aplicada cuatro veces, en las semanas 1, 4, 19 y 24 del año académico. Dos tratamientos instructivos fueron administrados a estos mismos estudiantes durante el año escolar. En el treceavo el instructivo EAI, denominado KK fué enseñado entre las semanas 1 y 4 y en el onceavo dia el instructivo EAI llamado FOC fué aplicado entre las semanas 19 y 24.

Entre la aplicación de estos 2 instructivos, los profesores siguieron su desarrollo regular del sílabo de matemáticas. Cuatro aplicaciones de la prueba FOC fueron realizadas. No hubo en los estudiantes retroalimentación en cuanto a las respuestas correctas acerca de su evaluación con el fin de evitar los efectos posibles de memoria. Así, los resultados de este estudio demuestran que la estrategia fue exitosa.

Mediciones
La prueba de FOC contiene 20 items de preguntas y respuestas cortas diseñadas para evaluar los efectos del instructivo FOC. Los ítems fueron computados dicotómicamente como correctos e incorrectos. La prueba fue aplicada en cada uno de los cuatro tiempos. Los items en el test fueron construidos para medir la capacidad de los estudiantes para realizar diagramas esquemáticos, realizar mediciones, procesar longitudes y determinar costos en la construcción de una rampa para patinadores. La confiabilidad del test fué de 0.80. La confiabilidad basada en las cuatro aplicaciones de este estudio fué de 0.78, 0.79, 0.79, 0.78 respectivamente.

Construcción de la Matriz Q
Este paso es muy importante para modelar el diagnóstico cognitivo y una mala especificación nos dará resultados erróneos. Para la prueba FOC, fueron medidas 4 habilidades cognitivas: Número y Operaciones, Medición, Resolución de problemas y Representación. Un experto en la competencia de educación matemática identificó las 4 habilidades e indicó cuales de estas cuatro habilidades fueron medidas por cada uno de los ítems en la prueba FOC. Los resultados de este análisis fueron usados para cosntruir la matriz Q con elementos q_{jk} que indica si la habilidad k es necesaria para resolver el ítem j.

Tabla 4.1: Matriz Q para el test de FOC

Items	Números y Operaciones	Medición	Solución de Problemas	Representación
1	1	0	0	0
2	1	0	0	0
3	1	1	0	0
4	1	1	0	0
5	1	1	0	0
6	1	1	0	0
7	1	1	0	0
8	1	1	0	0
9	0	1	0	1
10	0	1	0	1
11	1	1	1	1
12	1	1	1	1
13	1	1	1	1
14	1	1	1	1
15	1	1	1	1
16	0	0	1	0
17	1	1	1	1
18	0	1	0	1
19	1	0	0	1
20	1	0	0	0

Preguntas de investigación e hipótesis de investigación basadas en modelos
Describiremos un marco de referencia de hipótesis de tal manera que nos permita determi-

nar:(1)si el estado de la habilidad dominada cambiaron a través de las 4 administraciones de la prueba y si estos cambios subsecuentes a cada intervención fueron iguales o diferentes. (2) si los diferentes grupos de evaluados tienen diferentes probabilidades de transición. Empezamos asumiendo que las cuatro habilidades son independientes y que el crecimiento en la habilidad es también independiente. Más adelante asumimos que las probabilidades de transición son diferentes para cada habilidad. Con respecto a la pregunta 1 hacemos una comparación básica asumiendo una población homogénea con respecto a la matriz de probabilidades de transición.

Tabla 4.2: Matriz de transiciones de probabilidad para dos categorias de clases latentes con cuatro puntos en el tiempo

Test Wave 1 → Test Wave 2 (KK instruction)	Test Wave 2 → Test Wave 3 Regular instruction	Test Wave 3 → Test Wave 4 (FOC instruction)
$p^{21}_{n\|n} p^{21}_{m\|n}$	$p^{32}_{n\|n} p^{32}_{m\|n}$	$p^{43}_{n\|n} p^{43}_{m\|n}$
$p^{21}_{n\|m} p^{21}_{m\|m}$	$p^{32}_{n\|m} p^{32}_{m\|m}$	$p^{43}_{n\|m} p^{43}_{m\|m}$

En la tabla anterior m indica logro y n indica no logrado con relación a la competencia. Como ejemplo consideremos la transición de la primera a la segunda prueba (es decir, el antes y después en el aplicativo KK). La matriz de probabilidad de transición esta compuesta por cuatro componentes: $p^{21}_{n|n}$ (los que permanecen en el estado de no logrado), $p^{21}_{m|m}$ (los que permanecen en el estado de logrado), $p^{21}_{m|n}$ (los que pasan o transicionan de un estado de logro a un estado no logrado), $p^{21}_{n|m}$ (los que pasan o transicionan de un estado de no logrado a un estado de logrado) Las filas de cada matriz de transición son probabilidades condicionales y deben sumar 1, es por eso que para cada matriz de transición, solo se necesitan estimar dos probabilidaes de transición.

Tres matrices de probabilidades de transición se construirán para observar los efectos del evaluativo KK (Etapa 1 a la Etapa 2), instrucción regular (Etapa 2 a la Etapa 3) y la instrucción del FOC (Etapa 3 a la Etapa 4). Cuatro modelos diferentes de estas tres probabilidades de transición fueron probadas para determinar si los efectos en los tres instructivos diferentes fueron aproximadamente iguales o no.

Modelo A

En este modelo, la matriz de probabilidades de transición para cada etapa se asume diferentes que las otras, de este modo indicamos que los efectos de todos los tres instructivos son diferentes.

$$\begin{vmatrix} p^{21}_{n|n} & p^{21}_{m|n} \\ p^{21}_{n|m} & p^{21}_{m|m} \end{vmatrix} \neq \begin{vmatrix} p^{32}_{n|n} & p^{32}_{m|n} \\ p^{32}_{n|m} & p^{32}_{m|m} \end{vmatrix} \neq \begin{vmatrix} p^{43}_{n|n} & p^{43}_{m|n} \\ p^{43}_{n|m} & p^{43}_{m|m} \end{vmatrix}$$

Modelo B

En este modelo, las dos matrices de probabilidades de transición previas al aplicativo FOC se asumen iguales. La transición seguida al FOC, sin embargo se asume que debe ser diferente. Es decir, los efectos del instructivo KK y la aplicación del instructivo de matemática se asumen iguales, pero ambos difieren del efecto del instructivo FOC.

$$\begin{vmatrix} p^{21}_{n|n} & p^{21}_{m|n} \\ p^{21}_{n|m} & p^{21}_{m|m} \end{vmatrix} = \begin{vmatrix} p^{32}_{n|n} & p^{32}_{m|n} \\ p^{32}_{n|m} & p^{32}_{m|m} \end{vmatrix} \neq \begin{vmatrix} p^{43}_{n|n} & p^{43}_{m|n} \\ p^{43}_{n|m} & p^{43}_{m|m} \end{vmatrix}$$

Modelo C

En este modelo, la matriz de transición del pre al post test del instructivo KK se asume que debe ser igual a la matriz de transición del pre al post test del instructivo FOC, pero no así igual que el pre-post del instructivo regular. Es decir, que los dos tratamientos instructivos EAI que son el KK y el FOC, se asumen que deben afectar a las probabilidades de transición en forma similar, pero ambos difieren de los efectos de la instrucción regular en matemáticas.

$$\begin{vmatrix} p^{21}_{n|n} & p^{21}_{m|n} \\ p^{21}_{n|m} & p^{21}_{m|m} \end{vmatrix} = \begin{vmatrix} p^{43}_{n|n} & p^{43}_{m|n} \\ p^{43}_{n|m} & p^{43}_{m|m} \end{vmatrix} \neq \begin{vmatrix} p^{32}_{n|n} & p^{32}_{m|n} \\ p^{32}_{n|m} & p^{32}_{m|m} \end{vmatrix}$$

Modelo D

En este modelo, todas las matrices de probabilidades de transición se asumen iguales, es decir, todos los evaluativos instruccionales se asumen que deben tener el mismo efecto en los estados de dominio o logro de los estudiantes.

$$\begin{vmatrix} p^{21}_{n|n} & p^{21}_{m|n} \\ p^{21}_{n|m} & p^{21}_{m|m} \end{vmatrix} = \begin{vmatrix} p^{32}_{n|n} & p^{32}_{m|n} \\ p^{32}_{n|m} & p^{32}_{m|m} \end{vmatrix} = \begin{vmatrix} p^{43}_{n|n} & p^{43}_{m|n} \\ p^{43}_{n|m} & p^{43}_{m|m} \end{vmatrix}$$

Con respecto a la Pregunta 2 que es, si diferentes grupos de examinados tienen diferentes probabilidades de transición, la restricción fue eliminada en el supuesto de homogeneidad de la población, tal que poblaciones diferentes permitieron tener diferentes matrices de probabilidades de transición. Los grupos más importantes serán los estudiantes con discapacidades y aquellos que no presentan estas. Desafortunadamente, la muestra de estudiantes con dificultades de aprendizaje fue muy pequeña ($N = 9$) lo que produce poca precisión en las estimaciones de las probabilidades de transición para ese grupo.

Estimación de los parámetros de los items en el Modelo DINA

En este estudio, el mismo test fue aplicado cuatro veces. Se asume que el efecto memoria no juega un rol importante, toda vez que la prueba consistió en una evaluación del desempeño

en la cual a los estudiantes se les preguntó como realizar las mediciones, el montaje y otras labores para la construcción del rampa de patinaje, y considerando que los estudiantes no recibieron la retroalimentación acerca de las respuestas correctas después de cada aplicación. Adicionalmente, se asume invarianza en los parámetros del ítem para el modelo DINAa lo largo de las 4 aplicaciones, a pesar de que las proporciones de dominio en la población se consideran cambiables. Mas aún, los parámetros de los items se fueron estimados conjuntamente a lo largo de las 4 aplicaciones. Las proporciones de dominio para cada habilidad fueron estimadas libremente a lo largo de las administraciones.
El modelo basado en hipótesis trata con probabilidades de transición, y como un resultado las estimaciones del parámetro de los ítems no afectan demasiado. Las estimaciones de los parámetros de los ítems para cada modelo hipotético fueron relativamente muy similares unos a otros.
Se presentará en el siguiente cuadro las estimaciones de los parámetros de los ítems para el Modelo A, las dos primeras columnas son las estimaciones de los prametros al item g y s respectivamente. La tercera columna proporciona estimaciones de la calidad de diagnóstico de cada ítem. Esta estimación se obtiene mediante:

$$\frac{(1-s_j)/s_j}{g_j/(1-g_j)} \tag{4.1}$$

El cual es llamado el odds ratio entre responder positivamente condicionado a $\eta_{ij}=1$ y de responder positivamente condicionado a $\eta_{ij}=0$. El ítem con el más alto odds ratio será considerado el más valorable en términos de la diferenciación entre las dos clases latentes, donde las dos clases latentes se definen como una primera en la cual los examinados han dominado todas las habilidades requeridas por el ítem, es decir, $\eta_{ij}=1$) y una segunda en la cual los examinados no han dominado por lo menos una habilidad requerida por ese ítem es decir, $\eta_{ij}=0$)

Método

Se consideró un escenario de evaluación de pre y post test en el cual se aplicaron las mismas pruebas o exámenes en dos tiempos($T=2$). Usando los datos generados, se investigó la recuperación de parámetros del modelo LTA-DINA. Particularmente usamos el modelo de diagnostico cognitivo DINA como modelo de medición. Se generó respuestas de 1000 estudiantes para 20 items diseñados para medir 4 habilidades. La matriz Q se definió con la ayuda de expertos y está dada en la Tabla 1. Los parámetros de adivinación y desliz fueron generados aleatoriamente de una distribución uniforme entre 0.1 y 0.3, y estos permanecieron constantes a través del tiempo.

Las probabilidades de transición, serán tomadas del articulo de Li. El código R que genera los datos está disponible en el Anexo. El modelo fue implementado en el sofware Mplus y el código también se encuentra en el anexo. El criterio de convergencia por defecto usado fue de 0.001 para el cambio absoluto en la logverosimilitud.

Resultados

Tabla 4.3: Estimaciones de los parámetros al item para el modelo LTA-DINA y Odds Ratios.

	True		LTA-DINA	
	Guess	Slip	Guess	Slip
Item1	0,281	0,283	0,20(0,01)	0,23(0,03)
Item2	0,128	0,287	0,20(0,01)	0,23(0,03)
Item3	0,300	0,157	0,20(0,01)	0,23(0,03)
Item4	0,289	0,266	0,20(0,01)	0,23(0,03)
Item5	0,116	0,228	0,20(0,01)	0,23(0,03)
Item6	0,203	0,204	0,20(0,01)	0,23(0,03)
Item7	0,178	0,247	0,20(0,01)	0,23(0,03)
Item8	0,281	0,127	0,20(0,01)	0,23(0,03)
Item9	0,189	0,231	0,20(0,01)	0,23(0,03)
Item10	0,267	0,241	0,20(0,01)	0,23(0,03)
Item11	0,247	0,192	0,20(0,01)	0,23(0,03)
Item12	0,262	0,244	0,20(0,01)	0,23(0,03)
Item13	0,178	0,287	0,20(0,01)	0,23(0,03)
Item14	0,237	0,151	0,20(0,01)	0,23(0,03)
Item15	0,101	0,192	0,20(0,01)	0,23(0,03)
Item16	0,266	0,288	0,20(0,01)	0,23(0,03)
Item17	0,101	0,296	0,20(0,01)	0,23(0,03)
Item18	0,142	0,123	0,20(0,01)	0,23(0,03)
Item19	0,281	0,195	0,20(0,01)	0,23(0,03)
Item20	0,222	0,212	0,20(0,01)	0,23(0,03)

Esto se interpreta como que los items del FOC fueron efecivos al momento de discriminar o distinguir entre los examinados entre las dos clases latentes. Este indice además nos indica que los valores en la matriz Q identifican con certeza las habilidades necesarias para responder bien a los items en la prueba. Los valores grandes de los odds ratios, también fueron debido principalmente a los pequeños parámetros de adivinación. Esto se deduce del hecho de que todos los 20 items de las pruebas FOC fueron de respuestas cortas al item, por consiguiente, la verosimilitud de adivinación fué más baja.

Tabla 4.4: Frecuencia y Proporción de cada habilidad dominada para cada punto en el tiempo.

Habilidades	Tiempo 1	Tiempo 2
Número y Operaciones	500 (0,500)	501 (0,501)
Medición	513 (0,513)	513 (0,513)
Solución de problemas	496 (0,496)	487 (0,487)
Representación	522 (0,522)	507 (0,507)

Probabilidades de dominio a través de los puntos en el tiempo

La tabla 4.4 presenta frecuencias y proporciones de las personas evaluadas que dominaron cada una de las habilidades cognitivas en cada punto en el tiempo. Las probabilidades de dominio para las 4 habilidades no han aumentado substancialmente inclusive para la habilidades de Solución de problemas y Representación ha disminuído. Es decir al parecer el primer instructivo KK no ha sido adecuado

Tabla 4.5: Frecuencias y Proporciones de los patrones de dominio para las cuatro habilidades en cada punto en el tiempo.

Pattern	Time Point 1	Time Point 2
$(0, 0, 0, 0)$	188 (0,188)	189 (0,189)
$(0, 0, 0, 1)$	65 (0,065)	49 (0,049)
$(0, 0, 1, 0)$	49 (0,049)	49 (0,049)
$(0, 0, 1, 1)$	31 (0,031)	30 (0,030)
$(0, 1, 0, 0)$	47 (0,047)	57 (0,030)
$(0, 1, 0, 1)$	41 (0,041)	43 (0,043)
$(0, 1, 1, 0)$	30 (0,030)	30 (0,030)
$(0, 1, 1, 1)$	49 (0,049)	52 (0,052)
$(1, 0, 0, 0)$	45 (0,045)	56 (0,056)
$(1, 0, 0, 1)$	35 (0,035)	43 (0,043)
$(1, 0, 1, 0)$	32 (0,032)	32 (0,032)
$(1, 0, 1, 1)$	42 (0,042)	39 $(0, 039)$
$(1, 1, 0, 0)$	32 (0,032)	35 (0,035)
$(1, 1, 0, 1)$	51 (0,051)	41 (0,041)
$(1, 1, 1, 0)$	55 (0,055)	45 (0,045)
$(1, 1, 1, 1)$	208 (0,208)	210 (0,210)

La Tabla 4.5 presenta frecuencias y probabilidades de los patrones de dominio observados en la data del FOC para cada punto en el tiempo. Un patrón que indica el dominio de las cuatro habilidades en un exámen es representado en la Tabla 4.5 como (1,1,1,1). Igualmente un patrón que indica que el examinado dominó ninguna de las cuatro habilidades es representado en la tabla como (0,0,0,0). Las 4 habilidades pueden generar 2^4 patrones. En datos reales, sin embargo, diferentes dificultades pueden suceder ocasionando en algunos casos que algunos patrones no sean considerados u observados lo cual podria reducir este número de patrones. Observando los patrones podemos decir que (1,1,1,1) fue el más observado a través de las dos mediciones, pero también el patrón (0,0,0,0) se mantuvo con la misma medición en el tiempo $T = 2$.

4.1 Estudio de simulación usando los parámetros recuperados de la base de datos FOC

Para confirmar que nuestras estimaciones pueden ser verdaderas o comprobadas, se condujo un estudio de simulación usando los valores de los parámetros estimados de la base de datos FOC. Se simuló datos con las mismas caracteristicas de la data del FOC con la ayuda del software libre R. Las respuestas de 100 estudiantes a 20 items que miden 4 habilidades cognitivas fueron generadas. Se usó la matriz Q de la tabla 4.1, para los parámetros del modelo utilizamos las estimaciones puntuales obtenidas en la data del FOC. Los parámetros de adivinación y desliz fueron aleatoriamente generados a partir de una distribución de probabilidad uniforme desde 0.1 hasta 0.3 y se mantuvieron constantes a través del tiempo. El modelo LTA-DINA con diferentes probabilidades de transición para cada habilidad y manteniendo constantes los parámetros de adivinación y desliz se muestran en la siguiente tabla:

Tabla 4.6: Probabilidades de transición estimadas para el modelo LTA-DINA

	$\widehat{p}_{m1}$	$\widehat{p}_{m\|n}$	$\widehat{p}_{m\|m}$	$\widehat{p}_{n\|m}$	$\widehat{p}_{n\|n}$
Número y Operaciones	(0,50)	(0,256)	(0,256)	(0,255)	(0,244)
Medición	(0,513)	(0,247)	(0,247)	(0,247)	(0,240)
Solución de problemas	(0,496)	(0,231)	(0,231)	(0,240)	(0,273)
Representación	(0,522)	(0,223)	(0,223)	(0,238)	(0,255)

La tabla 4.6 nos presenta las probabilidades de transición para el modelo LTA-DINA. pm1 indica la probabilidad de dominio de la habilidad en el primer tiempo. pm/n es la probabilidad de transición de un estado de no dominio a un estado de dominio, pm/m es la probabilidad de transición de un estado de dominio a un estado de dominio(es decir, permanencia en el mismo estado), pn/m es la probabilidad de transición de un estado de dominio a un estado de no dominio, pn/n nos indica la probabilidad de transición de un estado de no dominio a un estado de no dominio(permanencia). El modelo además nos dice que para el primer punto en el tiempo, el 50 % de los evaluados han logrado el dominio de la habilidad Número y operaciones, el 51,3 % de los evaluados han alcanzado a dominar la habilidad de Medición, el 49,6 % de los evaluados ha dominado la habilidad Solución de problemas y el 52,2 % ha dominado la habilidad de Representación.

pm1 puede ser interpretado como un ratio o razón de aprendizaje, el cual puede ser utilizado como un indicador por ejemplo cuando se evalua una intervención educativa que se realiza en dos periodos en el tiempo.

Además se evaluó el ajuste del modelo. La tabla 4.7 presenta la siguiente información con respecto al ajuste del modelo:

Tabla 4.7: Información del ajuste del modelo LTA-DINA.

Model	Log-likelihood	Number of parameters	AIC	BIC
LTA-DINA	−16125	52	32355,77	32610,98
LTA-DINA	−16125	52	32355,77	32610,98

4.2 Dominio de las habilidades predichas para el modelo LTA-DINA

Código inicial en R

Código en R que genera datos para el ajuste del modelo LTA-DINA

```
# load required packages
library(MASS)
library(boot)

# number of respondents
J <-1000

# number of items
I <-20

# number of skills
K <-4

# Q- matrix
Q <- t(matrix(c(1, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1,
      1, 0, 0, 1, 1, 0, 0, 1, 1 ,0, 0, 1, 1, 0, 0, 0, 1,
      0, 0, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1 ,1, 1, 1, 1, 1,
      1  1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1, 0, 1, 1 ,1, 1,
      0, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0, 0) ,K,I))
rownames(Q) <- paste0 (" Item ", 1:I)
colnames(Q) <- paste0 ("A", 1:K)

# skill profile patterns
alpha_patt <- as.matrix ( expand.grid (c(0 ,1) ,c(0 ,1) ,c(0 ,1) ,c (0 ,1)))
colnames(alpha_patt) <- paste0 ("A", 1:4)
alpha_patt

# slip and guess
slip <- c(0.192 ,0.260 ,0.119 ,0.291 ,0.143 ,0.182 ,0.237 ,0.209 ,0.134 ,
        0.241 ,0.238 ,0.206 ,0.279 ,0.164 , 0.266 ,0.256 ,0.118 ,0.291
 ,0.210 ,0.264)
guess <- c(0.201 ,0.242 ,0.263 ,0.122 ,0.230 ,0.186 ,0.119 ,0.117 ,0.174 ,
         0.205 ,0.274 ,0.123 ,0.265 ,0.278 ,0.293 ,0.233 ,0.133 ,0.165
 ,0.150 ,0.283)
# generate higher - order latent traits at two time points
```

```r
set.seed (1234)
theta <- mvrnorm(n=J,mu=c(0 ,0.3) , Sigma = matrix (c(1 ,.8 ,.8 ,1) ,2 ,2))

# structural model parameters
lambda0 <- c(1.51 , -1.42 , -0.66 , 0.5)

# generate true skill mastery profiles and sample responses
resp <-array(NA , dim =c(J,I ,2))
A_all <-array(NA , dim=c(J,K ,2))
for (t in 1:2){
 # find the prob of respondent j having skill k
  eta.jk <- matrix(0,J,K)
  for (j in 1:J) {
    for (k in 1:K){
      eta.jk[j, k]<-exp(theta[j,t] + lambda0[k])/(1 + exp(theta [j,t] +
 lambda0 [k]))}
     }
   A <- matrix(0,J,K)
   for (j in 1:J) {
     for (k in 1:K) {
       A[j,k]=rbinom(1,1,eta.jk[j,k])
      }
     }

   # calculate if respondents have all skills needed for each item
   xi_ind <- matrix(0, J, I)

   for (j in 1:J) {
     for (i in 1:I) {
       xi_ind[j,i]<-prod(A[j, ]^Q[i, ])
     }
   }

  # generate prob correct and sample responses
  prob.correct <- matrix(0, J, I)
  y <- matrix(0, J, I)
  for (j in 1:J) {
    for (i in 1:I) {
      prob.correct[j,i] <- ((1 - slip [i])^xi_ind [j,i])*( guess [i ]^(1 -
 xi_ind[j,i]))
      y[j, i] <- rbinom(1, 1, prob.correct[j,i])
     }
    }

  A_all[,,t]<-A
  resp [,,t]<-y
}
skill_data<-cbind(A_all[,,1],A_all[ , ,2])
resp_data<-cbind(resp[,,1], resp[ , ,2])
```

Código Mplus para el ajuste del modelo LTA-DINA

```
TITLE: !LTA-DINA model for T=2
DATA:
FILE IS C:\Users\Usuario\Documents/LTA1_DINA.txt;
VARIABLE:
NAMES ARE
x1 x2 x3 x4 x5 x6 x7 x8 x9 x10 x11 x12
x13 x14 x15 x16 x17 x18 x19 x20 y1 y2
y3 y4 y5 y6 y7 y8 y9 y10 y11 y12 y13
y14 y15 y16 y17 y18 y19 y20;
USEVARIABLES = x1-x20 y1-y20;
CATEGORICAL = x1-x20 y1-y20;
CLASSES = c1(2) c2(2) c3(2) c4(2) c5(2) c6(2) c7(2) c8(2);
ANALYSIS:
TYPE = MIXTURE;
PARAMETERIZATION=PROBABILITY;
STARTS =0;
ALGORITHM = INTEGRATION;
PROCESSORS  =  4;
MODEL:
OVERALL
c5 ON c1;
c6 ON c2;
c7 ON c3;
c8 ON c4;
```