



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA MATEMÁTICA

ESTIMACIÓN DE DESEMPEÑO EN EVALUACIÓN SUMATIVA, CON BASE EN
EVALUACIONES FORMATIVAS USANDO MODELOS ESPACIO ESTADO

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIAS DE LA
INGENIERÍA, MENCIÓN MATEMÁTICAS APLICADAS

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL MATEMÁTICO

OBED AZBEL ULLOA MIRANDA

PROFESOR GUÍA:
ROBERTO ARAYA SCHULZ
PROFESOR CO-GUÍA:
RAÚL GORMAZ ARANCIBIA

COMISIÓN:
ABELINO JIMÉNEZ GAJARDO

Este trabajo ha sido parcialmente financiado por CMM ANID PIA AFB170001 y Proyecto
Fondo Basal FB0003

SANTIAGO DE CHILE
2021

Resumen

En el mundo moderno, en especial en el contexto de la pandemia, la implementación de la *Computer Aided Instruction* o evaluación formativa asistida por computadora (CAI, por sus siglas en inglés) es útil y necesaria. En consecuencia, se ha generado gran interés en determinar qué nivel de desempeño de los alumnos se alcanza al implementar estas estrategias. En este contexto, las evaluaciones sumativas han sido muy cuestionadas y las evaluaciones formativas han sido señaladas como alternativas más efectivas.

Entre la gran diversidad de artículos que buscan predecir el desempeño de los estudiantes, algunos han usado redes neuronales para estimar la correctitud de la respuesta de un alumno, mientras otros usan técnicas como *random forest* o **bayesian knowledge tracing**. También hay casos donde la predicción está enfocada en categorías de desempeño o estimar el eventual abandono de un programa de clases en línea.

Sin embargo, una de las limitaciones de los métodos de predicción utilizados últimamente es la falta de explicabilidad de los modelos que se emplean. Muchos modelos hacen uso de complicadas técnicas de predicción que requieren cierto grado de conocimientos específicos para entenderlas.

Además de la falta de explicabilidad, el porcentaje de estudios enfocados en niveles K-12 es considerablemente bajo. Asimismo, es relativamente bajo el número de artículos donde la predicción está enfocada en una regresión numérica y no en una clasificación.

En esta tesis, se propone modelos lineales y de espacio estado para la predicción del desempeño académico en la prueba nacional estandarizada SIMCE de matemáticas para cuarto básico.

Para analizar el RMSE (*Root Mean Squared Error*) de los modelos, hay que considerar que la desviación de la prueba de 47.8 puntos es la cota superior (RMSE al estimar por la media) y el error de medición de la prueba de 16.56 es la cota inferior; así podemos esperar en el mejor de los casos disminuir 31.24 puntos de error. Dicho esto, el modelo con menor RMSE logra un valor de 26.61 puntos en estimaciones fuera de la base de entrenamiento, esto es, reduce 21.19 de los 31.24 puntos ideales, es decir un 67.8% del error explicable. Por otro lado, el modelo espacio estado reporta un RMSE de 27.52 puntos (64.9% del error explicable), con menos parámetros, más aún, se trata de una formulación que abre eventuales planteamientos para mejorar el diseño de la ejercitación en matemáticas y, por lo tanto, lograr mejores desempeños.

*A mis padres y hermanos.
A Dios a quien le debo todo.*

Agradecimientos

No habría sido posible llevar a cabo esta tesis sin la ayuda de Dios y mis padres José Ulloa y Rachel Miranda, quienes me han apoyado durante toda mi vida y no han dudado sobre mi vocación al elegir mi carrera, sino que me han entregado amor incondicional y mucha paciencia. También estoy muy agradecido con mis hermanos Caleb, Jared y Heber, cada uno entregando su apoyo de distinta forma y sin ellos no podría asegurar que ese trabajo existiría.

A lo largo de mi formación académica muchas figuras me han inspirado y guiado, desde mis profesores de enseñanza básica Juan Acevedo y Ana Meneses, el primero que hizo nacer en mí el amor por las matemáticas y la segunda que fue quien me orientó a ser mejor persona y mejor estudiante. En mi enseñanza media fueron los profesores Juan Carlos Paredes y Rodrigo Gonzáles quienes influyeron en mi maduración académica y como persona, permitiendo que decidiera la carrera de ingeniería matemática.

Durante los años en la universidad, es primeramente mis amigos quienes vienen a mi memoria como parte esencial en llegar hasta donde estoy, aquellos que conocí en Plan Común como Mario Varela, Luis Loyola, Camila Santibañez, Ednar Echevería, Martín Rocha, Bryan Sagredo, Bárabara Blanco y Yasser Nanjari, éste último quien también se decidió por la misma carrera. Durante los años de especialidad como alumno del departamento de ingeniería matemática son muchos los compañeros que me apoyaron en diversas ocasiones, pero me gustaría mencionar con especial apreciación a Ian Letter, Roberto Bobadilla, Simón Piga, Donato Vásquez, Aturo Merino; especialmente aquellos que con quienes compartí muchos momentos en la oficina 435: Kevin Contreras, Felipe Matus, Diego Marchant, Francisco Sanhueza, Vicente Ocqueteau, Pablo Arratia, Reidmen Aróstica, Felipe Olivares, Manuel Suil y Daniel Pereda.

Como estudiante de ingeniería conocí profesores y auxiliares que han sido una motivación y guía durante la carrera, entre los cuales recuerdo con una apreciación especial a Marcelo Leseigneur, quien fue mi profesor y con quien luego trabajaría como miembro del cuerpo docente en más de un curso e incluso en proyectos de ingeniería vinculados con su trabajo, pese a ser aún un estudiante sin experiencia. Otros profesores a quienes me gustaría agradecer son el profesor Rafael Correa, Servet Martínez, Joaquín Fontbona y Gianfranco Liberona.

En cuanto a la realización de esta tesis, le debo gran parte de esta al trabajo y apoyo de mi profesor guía Roberto Araya, quien primero fue mi jefe, cuando me ofreció la oportunidad de trabajar en investigación en educación con él y mi profesor co-guía Raúl Gormaz, gracias a quién llegué a este trabajo. Fueron ellos quienes más tarde me permitieron llevar a cabo mi tesis con ellos en esta misma área de investigación y quienes han aportado constantemente su

experiencia, con lo cual he sido capaz de aprender muchas cosas y crecer como ingeniero. Es trabajando con ellos que llegué a conocer a Abelino Jiménez, a quien agradezco por todas las veces que ha apartado parte de su tiempo para resolver dudas, ofrecer consejos y orientaciones en el trabajo que he realizado; y tuvo la disposición de ser parte de la comisión de esta tesis.

Finalmente, no puedo dejar de mencionar a los funcionarios del departamento que siempre estuvieron dispuestos para ayudar a los alumnos en distintos asuntos, tanto administrativos como académicos. Es así como agradezco a Natacha Astromujoff a quién en más de alguna ocasión di un dolor de cabeza con tanta pregunta y trabajo extra que mis problemas le significaban, también Eterin Jaña que siempre me recibió con una sonrisa y prestó su ayuda cuando lo necesité, al igual que Karen Hernández. Por último, y quien me ayudó enormemente en todo el proceso de titulación, agradezco a Silvia Mariano a quien estuve contactando semanalmente estos últimos meses de trabajo en la tesis.

En resumen, hay mucha gente a quien agradezco y muchos más que no alcanzo a mencionar, pero sin cada uno de ellos las cosas habrían sido muy diferentes y mucho más difíciles. A todos y cada uno de ellos les debo mi eterna gratitud.

Tabla de Contenido

Introducción	1
1. Programa ConectaIdeas y bases de datos	5
1.1. ConectaIdeas	5
1.2. SIMCE	7
1.3. Cuestionarios	10
1.4. Datos ConectaIdeas	11
2. Modelos Lineales para predicción de desempeño	12
2.1. Modelos <i>Gain</i> y <i>Quasi-gain</i>	12
2.2. Selección de Variables históricas	13
2.3. Variables dinámicas	17
2.4. Regresiones lineales	20
2.5. Resultados	24
2.5.1. Modelo sin memoria	32
2.5.2. Modelo glotón	40
2.5.3. Modelo con memoria	50
2.6. Modelo lineal con efecto olvido	57
2.6.1. Resultados en el Modelo con olvido	58
2.7. Resumen	69
3. Modelos espacio estado para la estimación de desempeño	70
3.1. Modelo Espacio Estado	70
3.2. Estados y variable de control	71
3.2.1. Efecto $Bu(t)$ constante.	73
3.2.2. Efecto $Bu(t)$ decreciente.	73
3.2.3. Efecto $Bu(t)$ creciente.	74
3.2.4. Efecto $Bu(t)$ decreciente y constante.	74
3.2.5. Efecto $Bu(t)$ creciente y constante.	75
3.2.6. Efecto $Bu(t)$ decreciente y creciente.	75
3.2.7. Resumen	75
3.3. Estimación de parámetros	76
3.4. Resultados	79
3.4.1. Análisis de resultados	85
3.5. Resumen	89

4. Comparación de los distintos modelos	91
4.1. RMSE en entrenamiento y testeo	92
4.2. Flexibilidad	95
4.3. Explicabilidad	97
4.4. Costo de predicción	97
4.4.1. Modelo sin memoria	98
4.4.2. Modelo glotón	98
4.4.3. Modelo con memoria	99
4.4.4. Modelo con olvido	100
4.4.5. Modelo espacio estado	101
4.4.6. Resumen	102
 Conclusión	 105
 Bibliografía	 108

Índice de Tablas

1.1. Matriz de evaluación teórica de pruebas SIMCE matemática 2015. Nivel 4° básico. Fuente: Informe Técnico SIMCE 2015.	8
1.2. Error de medición de prueba SIMCE para el promedio curso	9
2.1. Variables a nivel alumno que cumplen los criterios de correlación con puntaje SIMCE y Aptus.	14
2.2. Variables a nivel alumno que cumplen los criterios de correlación con puntaje SIMCE y Aptus.	16
2.3. Número de parámetros $\delta^{(m)}$ por modelo, $m \in \{1, 2\}$	23
2.4. Número máximo de parámetros por modelo	24
2.5. Variables que entregaron un mejor desempeño para el modelo sin memoria	32
2.6. RMSE promedio y desviación del RMSE para entrenamiento y testeo en el modelo sin memoria	34
2.7. RMSE promedio del promedio curso y desviación de este RMSE para entrenamiento y testeo en el modelo sin memoria	37
2.8. Efectos $\delta_{k,t}^{(1)}$ para los 5 ejes de contenidos dependiendo de la información en t	39
2.9. Variables que entregaron un mejor desempeño para el modelo glotón	40
2.10. RMSE promedio y desviación del RMSE para entrenamiento y testeo en el modelo glotón	42
2.11. Diferencia entre el RMSE en entrenamiento y el RMSE en testeo para el modelo glotón	42
2.12. RMSE promedio del promedio curso y desviación de este RMSE para entrenamiento y testeo en el modelo glotón	45
2.13. Efectos $\delta_{k,t}^{(1)}$ para los 5 ejes de contenidos dependiendo de la información en t	48
2.14. Efectos $\delta_{k,t}^{(2)}$ para los 5 ejes de contenidos dependiendo de la información en t	49
2.15. Variables que entregaron un mejor desempeño para el modelo con memoria	50
2.16. RMSE promedio y desviación del RMSE para entrenamiento y testeo en el modelo con memoria	51
2.17. RMSE promedio del promedio curso y desviación de este RMSE para entrenamiento y testeo en el modelo con memoria	54
2.18. Efectos $\delta_{k,t}^{(1)}$ para los 5 ejes de contenidos dependiendo de la información en t	57
2.19. Efectos $\delta_{k,t}^{(2)}$ para los 5 ejes de contenidos dependiendo de la información en t	57
2.20. Variables que entregaron un mejor desempeño para el modelo con olvido	58
2.21. RMSE promedio y desviación del RMSE para entrenamiento y testeo en el modelo con olvido	60

2.22. RMSE promedio del promedio curso y desviación de este RMSE para entrenamiento y testeo en el modelo con olvido	63
2.23. Efecto descuento para cada mes	67
2.24. Efectos $\delta_{k,t}^{(1)}$ para los 5 ejes de contenidos dependiendo de la información en t	67
2.25. Ejemplo de un caso con desempeño sobre la media en todos los meses	68
2.26. Efectos $\delta_{k,t}^{(1)}$ para los 5 ejes de contenidos dependiendo de la información en t	69
3.1. Variables que entregaron un mejor desempeño para el modelo espacio estado	80
3.2. RMSE promedio y desviación del RMSE para entrenamiento y testeo en el modelo espacio estado	81
3.3. RMSE promedio del promedio curso y desviación de este RMSE para entrenamiento y testeo en el modelo espacio estado	84
3.4. Parámetros del modelo espacio estado	85
4.1. Resumen de los modelos	91
4.2. RMSE promedio y desviación del RMSE para testeo en modelos con olvido y espacio estado	95
4.3. Costos de memoria y cómputo de modelo sin memoria	98
4.4. Costos de memoria y cómputo de modelo glotón	99
4.5. Costos de memoria y cómputo de modelo con memoria	100
4.6. Costos de memoria y cómputo de modelo con olvido	101
4.7. Costos de memoria y cómputo de modelo espacio estado	101
4.8. Costo en memoria y número de operaciones requeridas para cada modelo en t	102
4.9. Orden de memoria y número de operaciones requeridas para cada modelo en t	102
4.10. Costo en memoria y número de operaciones requeridas para cada modelo hasta T	102
4.11. Orden de memoria y número de operaciones requeridas para cada modelo hasta T	103

Índice de Ilustraciones

1.1. Alumnos trabajando en la plataforma de ConectaIdeas durante una sesión.	5
2.1. RMSE en el baseline con nota en t	26
2.2. RMSE en el baseline con nota hasta t	27
2.3. RMSE en el baseline con nota mensual	27
2.4. RMSE en el baseline con pretest	28
2.5. RMSE en el baseline con midtest	28
2.6. RMSE del promedio curso en el modelo baseline con nota en t	30
2.7. RMSE del promedio curso en el modelo baseline con nota hasta t	30
2.8. RMSE del promedio curso en el modelo baseline con nota mensual	31
2.9. RMSE del promedio curso en el modelo baseline con pretest	31
2.10. RMSE del promedio curso en el modelo baseline con midtest	32
2.11. RMSE en el modelo sin memoria	33
2.12. Cuatro casos en entrenamiento y sus distintas predicciones según el mes t de información usada	35
2.13. Cuatro casos en testeo y sus distintas predicciones según el mes t de información usada	36
2.14. RMSE del promedio curso en el modelo sin memoria	37
2.15. Efecto γ_t que acompaña la variable pretest, para distintos meses t de información.	38
2.16. Efecto $\beta_{1,t}$ asociado a la variable <i>simce</i> histórico, para distintos meses t de información.	38
2.17. Efecto $\beta_{2,t}$ asociado a la variable “Me va bien en matemáticas”, para distintos meses t de información.	39
2.18. RMSE en el modelo glotón	41
2.19. Cuatro casos en entrenamiento y sus distintas predicciones según el mes t de información usada	43
2.20. Cuatro casos en testeo y sus distintas predicciones según el mes t de información usada	44
2.21. RMSE del promedio curso en el modelo glotón	45
2.22. Efecto γ_t que acompaña la variable pretest, para distintos meses t hasta el cual se tiene información.	46
2.23. Efecto $\beta_{1,t}$ asociado a la variable “Mi problema es un comportamiento para el profesor de matemática”, para distintos meses t de información.	46
2.24. Efecto $\beta_{2,t}$ asociado a la variable nota de torneos espiral, para distintos meses t de información.	47
2.25. RMSE en el modelo con memoria	51

2.26. Cuatro casos en entrenamiento y sus distintas predicciones según el mes t de información usada	52
2.27. Cuatro casos en testeo y sus distintas predicciones según el mes t de información usada	53
2.28. RMSE del promedio curso en el modelo con memoria	54
2.29. Efecto γ_t que acompaña la variable pretest, para distintos meses t hasta el cual se tiene información.	55
2.30. Efecto $\beta_{1,t}$ asociado a la variable since histórico, para distintos meses t de información.	55
2.31. Efecto $\beta_{2,t}$ asociado a la variable “Mis padres dicen que soy buen estudiante”, para distintos meses t de información.	56
2.32. RMSE en el modelo con olvido	59
2.33. Cuatro casos en entrenamiento y sus distintas predicciones según el mes t de información usada	61
2.34. Cuatro casos en testeo y sus distintas predicciones según el mes t de información usada	62
2.35. RMSE del promedio curso en el modelo con olvido	63
2.36. Efecto γ_t que acompaña la variable pretest, para distintos meses t hasta el cual se tiene información.	64
2.37. Efecto $\beta_{1,t}$ asociado a la variable “Mi problema es un comportamiento para el profesor de matemática”, para distintos meses t de información.	65
2.38. Efecto $\beta_{2,t}$ asociado a la variable “Matemáticas es fácil para mí”, para distintos meses t de información.	66
2.39. Efecto ρ de olvido.	67
3.1. RMSE en el modelo espacio estado	81
3.2. Cuatro casos en entrenamiento y sus distintas predicciones según el mes t de información usada	82
3.3. Cuatro casos en testeo y sus distintas predicciones según el mes t de información usada	83
3.4. RMSE del promedio curso en el modelo espacio estado	84
4.1. RMSE para estimaciones de puntaje para los 5 modelos en el grupo de testeo	92
4.2. RMSE para estimaciones del promedio curso de puntaje para los 5 modelos en el grupo de testeo	93
4.3. Diferencia del RMSE y SEM para estimaciones de puntaje para los 5 modelos en el grupo de testeo	94
4.4. Diferencia del RMSE y SEM para estimaciones del promedio curso de puntaje para los 5 modelos en el grupo de testeo	94
4.5. Costos en $t = 8$ para distinto número n de alumnos	103
4.6. Costos totales a $T = 8$ para distinto número n de alumnos	103
4.7. Costos en t para $n = 100,000$ de alumnos	104
4.8. Costos totales a T para $n = 100,000$ de alumnos	104

Introducción

En la actualidad, existe una amplia gama de recursos tecnológicos disponibles para gran parte de la población, así lo reporta el informe preparado por la Subsecretaría de Telecomunicaciones de Chile en diciembre de 2017. Dicho informe indica que 87.4% de los hogares declara tener acceso propio y pagado a internet [9].

Un alto nivel de acceso a la tecnología significa que ésta estará disponible para los diversos aspectos de la vida diaria, lo que presenta tanto desafíos como oportunidades en todos los ámbitos, entre estos, la educación. Por ello, en variadas ocasiones se han implementado metodologías de clase que buscan incorporar el uso de estos recursos para ayudar en el aprendizaje de contenidos. Una de las principales ventajas de la enseñanza que se apoya en herramientas tecnológicas es la recolección de información acerca del estudiante, como ya lo mencionaba John A. Self en 1973[33]. En dicho artículo, este autor analiza las ventajas y desventajas de la instrucción asistida por computadora.

Desde 1973 a la fecha, la forma en que el computador puede asistir en el aula ha cambiado considerablemente y con el tiempo se ha consolidado la utilidad de dicha herramienta. Esto ha permitido realizar meta-análisis, como el de Qing Li y Xin Ma en 2010[30], un análisis de 46 estudios, el cual se enfocó en el efecto de aprendizaje en matemáticas y reportó que la tecnología tenía un efecto positivo significativo en los logros en matemáticas. Ese mismo año, Kebritch, Hurumi y Bai[21] estudiaron el efecto que un juego de matemáticas en computadora tiene en el desempeño en matemáticas y encontraron una diferencia positiva y significativa en los resultados del *posttest* entre el grupo de control de 76 alumnos y el grupo de tratamiento de 117 alumnos.

Esta tesis trabaja con datos de un programa llamado ConectaIdeas, implementado por primera vez en 2002. En esta tesis trabajamos con los datos del año 2017, del cual Araya *et al.* (2019)[3] reportaron un efecto positivo de 0.27 desviaciones estándar en el *posttest* para los alumnos que participaron en el grupo de tratamiento; esto equivale a un poco más de 12 puntos SIMCE, o bien un 48% más del aprendizaje que se suele lograr en cuarto básico, que son las 0.56 desviaciones que reportan Carolyn J. Hill *et al.* (2008)[19]. Sin embargo, este efecto es el obtenido en el grupo de tratamiento y para futuras instancias del programa, será de interés estimar el efecto que tendrá el programa en un establecimiento en particular, en un curso o inclusive en un alumno. Por esto, es necesario construir un estimador de desempeño dentro del programa que permita obtener predicciones para futuras realizaciones de ConectaIdeas.

Con distintas herramientas de estadística y *machine learning* se han realizado estimaciones

del desempeño de estudiantes, ya sea estimando el puntaje en un determinado test o actividad, o estimando una categoría de desempeño. Por ejemplo, Yang *et al.* (2017)[39] utilizaron redes neuronales de series de tiempo para la estimación de las calificaciones de alumnos en cursos abiertos masivos en línea. Por otra parte, Badrinath *et al.* (2021)[2] diseñaron una librería en python de modelos de rastreo de conocimiento Bayesiano (*Bayesian Knowledge Tracing*), la cual permite crear, inicializar, ajustar, predecir, evaluar y hacer validación cruzada con estos modelos, a fin de estimar la maestría cognitiva de los estudiantes en el contexto de sistemas de tutoría computacional.

También es importante considerar qué desempeño se está estimando: muchos artículos están enfocados en determinar categorías de desempeño, como los estudios de Xinhua Wang *et al.* (2020)[37], Hasan *et al.* (2018)[17] o el de Hamsa *et al.* (2016)[17]. Mientras tanto, un porcentaje menor se centra en estimar directamente un puntaje o calificación. Además, aquellos casos donde se busca predecir el puntaje suelen ser estudios con alumnos universitarios y hacer uso de indicadores de desempeño alcanzado en cursos de primer y segundo año de universidad para estimar el rendimiento en años siguientes, o bien, basarse en antecedentes socio-económicos de los alumnos, así como en puntajes en pruebas de admisión a las instituciones. Otros casos de estudio en los últimos años han sido los *Massive Open Online Course* o cursos abiertos en línea (MOOC por sus siglas en inglés), pero también en estos los participantes en su mayoría cuentan con más de 18 años de edad.

Un caso especialmente importante, es el trabajo realizado por Van der Molen (2013)[27], quien usando datos de la plataforma ConectaIdeas del año 2011, planteó estrategias para estimar el puntaje que obtendría un alumno, en un conjunto de ejercicios futuros dentro de esta misma plataforma. En su tesis, usó árboles de decisión, *Support Vector Machines* (SVM), factorización matricial y factorización de tensores. Los resultados reportados indican que dependiendo los contenidos, el desempeño de los métodos de predicción varía, sin mostrar alguna prevalencia de un método en general, sin embargo los métodos de factorización obtienen un mejor ajuste en la mayoría de los casos; con un RMSE de 0.43099 en testeo para el mejor caso, para puntajes en escala [0,1].

Esta tesis propone modelos que predicen un puntaje en la prueba nacional estandarizada SIMCE, para alumnos de cuarto año básico en matemáticas. Se usa poca información pre-existente sobre el alumno y gran parte de los datos usados para la predicción se obtienen mediante la implementación de sesiones con ejercicios en una plataforma computacional. Es decir, se recopila información de forma similar que en un MOOC, pero para alumnos K-12, en particular para alumnos entre 8 y 10 años.

La elección del puntaje SIMCE como el objetivo de la estimación tiene dos principales motivos. El primero obedece a una necesidad actual de conocer antes de la prueba cuál es el puntaje esperado; es decir, que funcione como un predictor para los profesores, de modo que puedan tener una expectativa de los resultados. La segunda razón es la búsqueda de una alternativa confiable para el SIMCE: si los modelos planteados en esta tesis lograsen estimaciones similares al SIMCE, bien podrían sustituirlo. En vez de una evaluación sumativa, se usaría una evaluación formativa, es decir, se evaluaría el proceso a lo largo del año y no el desempeño en un solo momento.

La propuesta de este trabajo consiste en usar *State Space Models* o modelos espacio es-

tado (SSM por sus siglas en inglés), también llamados *Input-Output Hidden Markov Models* (IOHMM). Reciben este nombre por la estructura que considera coordenadas para variables de entrada (*input*) y variables de salida (*output*), así como estados ocultos que funcionan como series de tiempo con un factor de “pérdida de memoria”.

Este enfoque es similar al de Feni Mi y Dit-Yan Yeung (2015)[25], quienes usaron información de MOOC –como Coursera y edX– y modelos de espacio estado para predecir el nivel de deserción en este tipo de cursos. Otro estudio con un acercamiento similar es el de C. Schatten y L. Schmidt-Thieme (2017)[32], que hace uso de modelos de espacio estado para la actualización temporal de los parámetros de un modelo de factorización matricial, con el objetivo de estimar el desempeño de los alumnos en determinadas tareas.

En ambos casos, se hace uso de modelos espacio estado, pero en el primero, es con estudiantes primariamente mayores de 18 años y se estima una variable binaria; mientras que en el segundo, los alumnos tienen entre 6 y 14 años, y el modelo es usado como una herramienta para agregar una componente temporal a modelos de factorización matricial. Cabe señalar que este último modelo solo estima parámetros asociados a los alumnos y a las tareas que ellos trabajaron, lo cual limita la realización de predicciones fuera de la muestra.

Igualmente, Feng Wang y Li Chen (2016)[36] implementaron modelos espacio estado que también usaban datos de MOOC para estimar el nivel de deserción, pero sus modelos eran no lineales. De modo semejante, Manie Tadayon y Greg Pottie (2019)[34] usaron los HMM para predecir el desempeño de alumnos de tercero y cuarto grado en cada nivel del juego educacional de matemáticas *Save PATCH* y clasificaron su desempeño en una escala discreta de 3 niveles. Ahora bien, este último modelo no es propiamente un modelo espacio estado puesto que no tiene una componente del tipo Input-Output, sino que las estimaciones se hacen con base en 3 reglas de predicción propuestas en el estudio.

Por otro lado, existen estudios que presentan mejor rendimiento en cuanto a poder predictivo, como Delianidi *et al.* (mayo 2021)[22], en el que se emplearon redes neuronales para estimar la correctitud de la respuesta de un alumno. Otro ejemplo es la investigación de Arif *et al.* (marzo 2021)[24] que estima las categorías de desempeño, al igual que Lalitha *et al.* (mayo 2021)[38]. Los estudios mencionados reportan buenos grados de predicción. Sin embargo, son modelos que o bien hacen uso de estructuras matemáticas más complejas que son difíciles de interpretar, como las redes neuronales; o bien, son modelos que requieren simplificar el valor a predecir trabajando con categorías de desempeño en lugar de puntajes.

Esta situación es uno de los temas abordados en el meta-estudio de Alamry y Alharbi (febrero 2021)[1]. Estos autores analizaron una serie de artículos sobre modelos de predicción y pusieron especial énfasis en la poca explicabilidad que tienen muchos de estos modelos de predicción, pues no es directo ni sencillo obtener una lectura sobre las componentes que constituyen los predictores utilizados.

La contribución de este trabajo recae principalmente en el uso de modelos espacio estado para la estimación de desempeño en una evaluación sumativa estandarizada, usando datos de ejercitación en línea de evaluaciones formativas para alumnos de cuarto año de enseñanza básica. El objetivo es mostrar que estos modelos se pueden usar en este contexto y obtener RMSE similares a modelos lineales, usando menos parámetros, entregando un mayor grado de

flexibilidad que estos para estimaciones fuera de la muestra y, al mismo tiempo, sin sacrificar la fácil interpretación y comprensión de estos modelos.

Además, los modelos espacio estado reconocen que no es posible medir directamente el aprendizaje o la habilidad de los alumnos, sino que se miden los resultados de pruebas y ejercicios. Es así que se modela el concepto de un estado latente u oculto del aprendizaje o habilidad. Esto permite interpretar las evaluaciones como una medición de la dinámica interna del aprendizaje del alumno.

En el primer capítulo de esta tesis, presentamos el contexto en que los datos fueron obtenidos, describiendo brevemente la plataforma de ejercitación en línea de la cual se recopiló la mayoría de los datos, así como las evaluaciones sumativas y cuestionarios realizados durante el transcurso del año escolar. El segundo capítulo, describe la metodología para la selección de variables y se presenta 4 modelos para la predicción de puntaje usando los datos descritos en el primer capítulo. El siguiente capítulo, es el que presenta el modelo espacio estado y su implementación, junto con un análisis de los resultados reportados con este modelo. Finalmente, el cuarto capítulo consiste en un resumen de los resultados obtenidos con cada modelo y un análisis comparativo de estos, considerando la ventajas y desventajas de cada uno.

Capítulo 1

Programa ConectaIdeas y bases de datos

Los datos que se presentan en este capítulo posteriormente pasarán por un proceso de selección de variables para luego ser usados con entradas para modelos de estimación del desempeño de los alumnos en la prueba SIMCE. En este contexto resulta importante entender de dónde provienen los diversos datos y al mismo tiempo tener en mente para qué serán usados.

1.1. ConectaIdeas

El programa ConectaIdeas es una plataforma de ejercitación en matemáticas con la asistencia de computadora desarrollado el año 2002 en el Centro de Investigación Avanzada en Educación (CIAE) de la Universidad de Chile. Se han realizado iteraciones del programa desde el año 2003 al 2021, agregando nuevas herramientas a lo largo de los años.

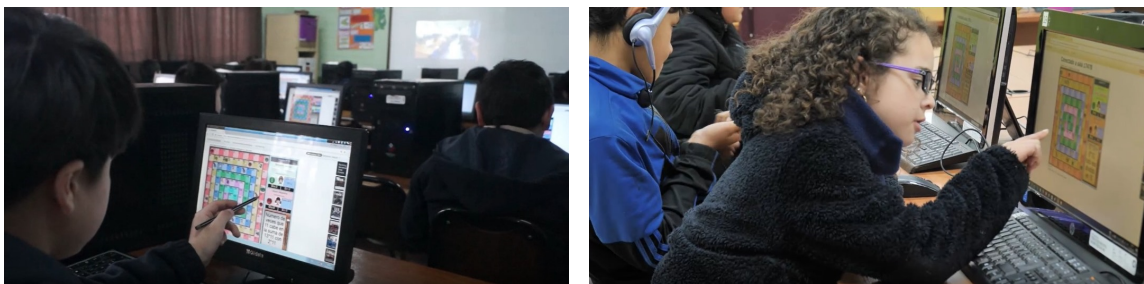


Figura 1.1: Alumnos trabajando en la plataforma de ConectaIdeas durante una sesión.

Una de las entidades particularmente interesada en el uso de herramientas y plataformas virtuales dentro del aula, es el Banco Interamericano del Desarrollo (BID), entidad que declara la importancia en que las escuelas tengan recursos adecuados y sean capaces de utilizarlos para el aprendizaje como una de las 5 dimensiones del éxito en educación[11]. Por esto ha llevado a cabo programas piloto en diversos países, entre ellos Chile, por lo que en conjunto con el Centro de Investigación Avanzada en Educación (CIAE), se llevó a cabo el año 2017 el programa **ConectaIdeas**. Realizado en 24 establecimientos municipales de la región metropolitana, es un programa de ejercitación en matemáticas, el cual consiste en realizar la

ejercitación de contenidos mediante una plataforma virtual donde los alumnos deben resolver el ejercicio que aparece en pantalla para recibir retroalimentación en forma de sonido de aplausos cuando responden correctamente, y luego así avanzar al siguiente ejercicio. Esto permite capturar información sobre el número de intentos que requiere el niño para resolver el ejercicio presentado, además del tiempo requerido; más aún, asociado a un número identificador (id) para cada ejercicio, es posible determinar a qué eje de contenidos corresponde y el objetivo de aprendizaje (OA) asociado a dicho eje. Este programa fue implementado con alumnos de cuarto básico, nivel para el cual los ejes de contenido y objetivos de aprendizaje se desglosan de la siguiente forma:

1. Números y operaciones

- *OA 01*: Representar números del 0 a 10,000
- *OA 02*: Aplicar estrategias de cálculo mental
- *OA 03*: Comprender relación suma y resta, resolución problemas
- *OA 04*: Aplicar propiedades del 0 y 1 para la multiplicación y división
- *OA 05*: Comprender la multiplicación de números de 3 dígitos
- *OA 06*: Utilizar estrategias para dividir
- *OA 07*: Resolver problemas en contexto cotidiano
- *OA 08*: Demostrar que comprende las fracciones
- *OA 09*: Resolver suma y resta de fracciones
- *OA 10*: Representar fracciones y números mixtos
- *OA 11*: Representar y ordenar decimales
- *OA 12*: Resolver suma y resta empleando valor posicional

2. Patrones y álgebra

- *OA 13*: Identificar patrones numéricos
- *OA 14*: Resolver ecuaciones que involucren suma y resta

3. Geometría

- *OA 15*: Describir localización de objetos en mapas
- *OA 16*: Determinar las vistas de figuras 3D
- *OA 17*: Identificar y crear figuras simétricas
- *OA 18*: Trasladar, rotar y reflejar figuras 2D
- *OA 19*: Construir ángulos con el transportador

4. Medición

- *OA 20*: Leer y registrar mediciones del tiempo
- *OA 21*: Realizar conversiones de tiempo
- *OA 22*: Medir longitudes con unidades estandarizadas
- *OA 23*: Comprender área de una figura 2D
- *OA 24*: Comprender concepto de volumen

5. Datos y probabilidades

- *OA 25*: Realizar encuestas y analizar sus resultados
- *OA 26*: Realizar experimentos aleatorios y tabular

- *OA 27*: Leer e interpretar gráficos y pictogramas

Teniendo información referente a la ejercitación que realiza cada niño en los distintos contenidos de matemática, es natural pensar en la relación que esta información tiene con el desempeño, o más aún, la evolución de dicho desempeño en base a dicha ejercitación, por lo cual se realizaron distintas pruebas durante el año para ir midiendo el avance que lograban los niños que fueron parte del programa. De igual modo, como es común en este tipo de estudios, se trabajó con grupos de tratamiento y control, los cuales fueron seleccionados a nivel curso, de manera aleatoria dentro de cada establecimiento, es decir, en cada colegio se determinaba de manera aleatoria entre los cursos A y B, cuál sería parte del grupo tratamiento y cuál sería parte del grupo control.

Independientemente si un curso era parte del grupo tratamiento o del grupo control, todos los alumnos de cuarto básico de los 24 establecimientos rindieron las pruebas mencionadas en el párrafo anterior. Las evaluaciones realizadas transversalmente entre tratamiento y control son:

1. “Pretest” o “Baseline” realizado a principio de año (marzo-abril) para medir el desempeño de los alumnos antes de la participación en el programa. Este fue llevado a cabo por la agencia Aptus.
2. “Midtest” realizado a mediados del año escolar (principios de agosto) para evaluar el avance luego de 4 meses usando el sistema ConectaIdeas. Este se realizó en conjunto con el centro de medición MideUC.
3. SIMCE, la prueba nacional estandarizada que se realiza todos los años cerca de fin de año (noviembre), bajo el control de la agencia de la calidad de la educación.
4. “Posttest” que se hizo a fines de año (diciembre) para medir resultados luego de haber realizado el SIMCE. Al igual que el “Midtest”, el encargado de esta prueba fue el centro de medición MideUC.

De estas 4 mediciones, son de particular interés el **pretest**, pues nos da un punto de referencia inicial; y el **SIMCE**, puesto que al ser realizado todos los años y en todo el país, es un parámetro de medición del desempeño que se maneja en todos los establecimientos de Chile.

1.2. SIMCE

El Sistema de Medición de la Calidad de la Educación (SIMCE), es una prueba estandarizada llevada a cabo por la agencia de la calidad de la educación, que por ley se estipula que son los encargados de elaborar y aplicar esta evaluación con el fin de “evaluar y orientar el sistema educativo para que este propenda al mejoramiento de la calidad y equidad de las oportunidades educativas”[10]. El SIMCE es básicamente una serie de pruebas en Matemáticas y otras asignaturas, que se aplican a alumnos de cuarto, sexto, y octavo años de enseñanza básica, y a segundo año de enseñanza media. Solo el año 2015 se realizó de manera experimental, un SIMCE a alumnos de segundo básico.

En el caso particular de la prueba de matemáticas para alumnos de cuarto año de enseñanza básica, esta es constituida por 40 preguntas, las cuales corresponden a cada uno de los

cinco ejes de contenidos mencionados anteriormente. La distribución de preguntas por cada eje para el año 2015 (último año con reporte técnico disponible)[16], se detalla en la tabla a continuación:

Eje	Porcentaje de preguntas
Números y operaciones	45 %
Patrones y álgebra	5 %
Geometría	20 %
Medición	20 %
Datos y probabilidades	10 %

Tabla 1.1: Matriz de evaluación teórica de pruebas SIMCE matemática 2015. Nivel 4° básico.

Fuente: Informe Técnico SIMCE 2015.

Esta proporción de preguntas por eje es la misma para las 5 formas distintas de las prueba. A su vez, para garantizar que estas formas sean equivalentes en cuanto a dificultad entre si, para que constituya un buen instrumento de medición de habilidad en matemáticas, se realiza cada año pruebas experimentales y con el uso de modelos de teoría de respuesta de ítem (IRT por Item Response Theory)[31], se puede seleccionar preguntas dentro de parámetros apropiados de dificultad para la elaboración de las distintas formas. Al igual que con las pruebas experimentales, los resultados de cada año son evaluados mediante IRT, con lo que se puede obtener una estimación de habilidad en matemáticas dado el conjunto de preguntas contestadas, por lo tanto, el puntaje SIMCE es una estimación, y en consecuencia, posee un error asociado. Al mismo tiempo, al no corresponder a una escala definida de puntaje, no existe un valor máximo exacto, ni tampoco un mínimo, pero para el año 2017, el rango de puntajes es entre 100.93 y 378.19 puntos, con media 260.95 y desviación estándar 47.80.

Considerando el rango de valores posibles para el SIMCE, se puede determinar el error estándar de medición (SEM por Standard Error of Measurement)[8] usando la confiabilidad α_c reportada para la prueba, que para el año 2017 en la prueba de matemáticas de cuarto básico estuvo entre 0.88 y 0.93 dependiendo de la forma de la prueba[15]. Esta confiabilidad corresponde al alfa de Cronbach, es decir:

$$\alpha_c = \frac{m}{m-1} \left(1 - \frac{\sum_{i=1}^m Var(x_i)}{Var\left(\sum_{i=1}^m x_i\right)} \right)$$

donde x_i son los puntajes a una pregunta o ítem $i \in \{1, \dots, m\}$ calculados mediante el modelo IRT. Luego, con esto se calcula el alfa de Cronbach para las m preguntas de una forma, para así obtener una medida de qué tan consistentes son estas en conjunto como indicador de la habilidad de cada alumno. A su vez, para obtener el SEM, se hace uso de la siguiente relación

$$SEM = s\sqrt{1 - \alpha_c}$$

donde s es la desviación estándar de la prueba. Sin embargo, es importante no confundir el SEM con el Error estándar de estimación (SEE por Standard Error of the Estimate), que

estará asociado a la estimación de puntaje mediante algún método predictivo como los que se usaron en el desarrollo de esta tesis. Siendo la fórmula del SEE

$$SEE = s\sqrt{1 - r_{xy}}$$

donde r_{xy} es el cuadrado de la correlación entre el puntaje de la prueba y los regresores x usados para predecirla. Luego, siempre que la correlación entre el SIMCE y los regresores usados para estimarlo sea menor a la confiabilidad de la prueba, se tendrá que

$$\begin{aligned} \alpha_C &\geq r_{xy} \\ \implies 1 - \alpha_C &\leq 1 - r_{xy} \\ \implies s\sqrt{1 - \alpha_C} &\leq s\sqrt{1 - r_{xy}} \\ \implies SEM &\leq SEE \end{aligned}$$

Es decir, que nuestro error de estimación estará acotado inferiormente por el error de medición del instrumento. Para el caso del SIMCE, como ya conocemos la confiabilidad, y dado que la desviación estándar de puntajes para el año 2017 fue de 47.80, podemos decir

$$\alpha_c \in [0.88, 0.93] \implies SEM \in [12.65, 16.56] \quad (1.1)$$

Por su parte, para las estimaciones del promedio de puntajes de un curso, será importante considerar de forma similar al cálculo recién hecho. Considerando un curso que tiene n alumnos, se puede determinar el SEM a nivel curso para esta prueba.

$$\{x_i\}_{i=1}^n, SEM(x_i) = s \implies SEM\left(\frac{1}{n}\sum_{i=1}^n x_i\right) = SEM(\bar{x}) = \frac{1.96}{\sqrt{n}}s$$

por esto, podemos calcular cuál será el SEM esperado para la media curso dependiendo del número de alumnos del curso.

Número de alumnos	SEM curso
10	[7.84, 10.26]
15	[6.40, 8.38]
20	[5.54, 7.26]
25	[4.96, 6.49]
30	[4.53, 5.93]
35	[4.19, 5.49]
40	[3.92, 5.13]
45	[3.69, 4.84]

Tabla 1.2: Error de medición de prueba SIMCE para el promedio curso

1.3. Cuestionarios

En conjunto con las pruebas SIMCE, también se realizan tres cuestionarios, para estudiantes, profesores y padres. De todo esto se obtiene una serie de antecedentes de cada alumno, de su profesor, sus padres y del establecimiento, y el puntaje obtenido en cada área respectiva. Datos que serán importantes a considerar como candidatos para la predicción de puntaje. El número de preguntas obtenidas de cada cuestionario fue

1. Cuestionario de profesores, 5 preguntas (28 ítems)
2. Cuestionario de padres, 3 preguntas (3 ítems)
3. Cuestionario de estudiantes, 3 preguntas (10 ítems)

donde los ítems corresponden al desglose de algunas preguntas. Las preguntas analizadas para el año 2017, el cuestionario del que se obtuvieron y las posibles respuestas se detallan en el anexo.

Además del cuestionario SIMCE, existe otro cuestionario realizado como parte del programa, el cual se enfocó en preguntas sobre el ambiente en clases, la percepción que tienen los alumnos sobre la inteligencia y sobre su disposición a estudiar matemáticas. Para este cuestionario, al ser realizado como parte del programa, se dispone acceso a la totalidad de preguntas, que fue un total de 10 sin contar la pregunta de ejemplo. Entre estas diez preguntas, se obtiene 41 ítems. De igual modo que antes, el detalle de las preguntas de este cuestionario se incluye en el anexo.

Otro cuestionario, son dos preguntas que se realizó a los estudiantes durante cuatro torneos de ejercicios, llamados “Torneos espiral”, a lo largo del año. La primera, realizada solamente en el primer torneo, planteaba el enunciado “Para aprender matemáticas prefieres:” y se presentaba las siguientes seis alternativas al alumno para completar la frase

1. Ejercicios de juego Espiral con rival de otra escuela
2. Ejercicios de juego Espiral con rival de mi curso
3. Ejercicios de juego Espiral con Bot
4. Ejercicios de conexión en el computador
5. Ejercicio en el cuaderno
6. Ejercicio en el pizarrón

Mientras que la segunda pregunta, hecha en los cuatro torneos, era “¿qué nota le pones al torneo?” y se presentaba como alternativas los valores enteros de uno a siete.

Con estos tres grupos de cuestionarios podremos extraer una serie de variables, tanto a nivel alumno, como a nivel curso y establecimiento; y llamaremos este conjunto de variables como **variables históricas**, dado el carácter fijo (o poco variable) a lo largo del año que presentan estas variables.

1.4. Datos ConectaIdeas

En el programa ConectaIdeas, al ser llevado a cabo mediante una plataforma virtual, es posible recopilar información sobre cada ejercicio realizado, como la fecha y hora, número de intentos antes de responder correctamente, respuestas a preguntas de abiertas (también llamadas preguntas de desarrollo), y una calificación entre 1.0 y 7.0 asignada por el sistema a cada ejercicio en función del número de intentos y alternativas de la pregunta. Para los fines del trabajo realizado, se agrupó la información a nivel mensual y con esto se construyó las siguientes variables:

1. Número de ejercicios contestados.
2. Número de ejercicios contestados en el primer intento.
3. Número de ejercicios contestados en más de un intento.
4. Diferencia entre el número de ejercicios contestados en el primer intento y los contestados en más de un intento.
5. Precisión de respuesta en el primer intento, calculada como el porcentaje de ejercicios resueltos en el primer intento.
6. Calificación promedio en los ejercicios.
7. Número de preguntas abiertas contestadas.
8. Total de palabras contestadas en preguntas abiertas.
9. Promedio de palabras por pregunta abierta contestada.
10. Calificación promedio en preguntas abiertas.

Obteniendo cada una de estas a nivel mensual y desglosadas por cada uno de los 5 ejes. Llamaremos este conjunto de variables provenientes de la plataforma como **variables dinámicas**, en referencia a cómo estas cambian durante el transcurso del año en que se llevó a cabo el programa.

Capítulo 2

Modelos Lineales para predicción de desempeño

En este capítulo se describe la metodología usada para seleccionar las variables, se presenta 4 modelos para la estimación del puntaje SIMCE, la estrategia usada para elegir aquellas variables que permiten que cada modelo tenga el menor RMSE posible y se analiza los efectos asociados a estas variables en cada modelo. Los modelos usados logran valores de RMSE que para el último mes antes del SIMCE, llegan hasta los 26.61.

2.1. Modelos *Gain* y *Cuasi-gain*

Es común en estudios donde se busca evaluar el desempeño antes y después de tratamiento, trabajar con modelos lineales de tipo *gain* o análisis de ganancia de puntaje (GSA por *Gain Score Analysis*) y *cuasi-gain* o análisis de ajuste por covarianza (CAA por *Covariance adjustment analysis*), donde la diferencia de ambos es en cómo se incorpora el desempeño en el pretest dentro de la ecuación.

Primeramente, un típico modelo lineal consiste de una variable observada $y \in \mathbb{R}^n$, que será estimada usando variables independientes $X \in \mathbb{R}^{n \times m}$, y un error asociado $\varepsilon \sim \mathcal{N}(0, \Sigma)$, $\Sigma \in \mathbb{R}^{n \times n}$. Entonces se formula mediante la ecuación

$$y = \beta X + \varepsilon$$

donde $\beta \in \mathbb{R}^m$ será el parámetro a estimar.

Los modelos de tipo *gain* son aquellos en que la variable a predecir es la diferencia de desempeño entre el posttest y el pretest del estudio, es decir, la ganancia en términos de puntaje. Para esto, sin embargo, es necesario que ambas pruebas estén en la misma escala de puntajes, o bien, que dichos puntajes sean normalizadas previamente para que así se pueda hacer la diferencia entre ambos. Esto es, teniendo un pretest p y un posttest y , y variables independientes X , se plantea la fórmula

$$y - p = \beta X + \varepsilon$$

donde ε es el error, que al igual que un modelo lineal tradicional, se asume con distribución normal, de varianza $\sigma > 0$. En este caso se puede definir $\tilde{y} := y - p$, puesto que la variable dependiente para modelos *gain* es la diferencia de los test, y con esta notación es simplemente escribir un modelo lineal tradicional.

Entonces la distinción se produce, cuando con las mismas variables se plantea la estimación como un modelo del tipo *cuasi-gain*, los cuales consideran el pretest como otro de los regresores, y por ende aparecerá un efecto asociado a estimar, el cual se puede llamar γ para diferenciarlo del efecto β de la ecuación anterior. Con esto un modelo *cuasi-gain* sería de la forma

$$y = \gamma p + \beta X + \varepsilon$$

En cuanto a la prevalencia de un modelo sobre otro, no existe una regla clara, sino mas bien dependerá del enfoque particular del estudio para el cual sean usados [35]. En el caso del programa ConectaIdeas y los posteriores análisis de resultados que se llevaron a cabo, nos interesará los modelos *cuasi-gain* por dos motivos, primero porque al tener un parámetro más (γ), permite ajustar de mejor manera el modelo y con esto obtener mejores predicciones; y segundo, porque el interés de nuestra predicción de desempeño será el puntaje SIMCE que obtendrá un estudiante más que en el incremento de desempeño respecto del pretest.

2.2. Selección de Variables históricas

En el contexto de la estimación de resultados SIMCE usando información de ejercitación y proveniente de los distintos cuestionarios mencionados en la introducción, es necesario elegir qué variables usaremos; puesto que, si bien, un acercamiento glotón promete reducir el error de predicción, también nos acercará a obtener modelos sobre ajustados, y al mismo tiempo significará un mayor costo en recopilar la información a la hora de realizar predicciones de desempeño en futuras instancias del programa. Por esto, buscaremos elegir una cantidad limitada de variables, idealmente referentes a datos tanto de los alumnos como de sus padres, el establecimiento o el curso, que sean de fácil obtención para futuras implementaciones de ConectaIdeas.

Para seleccionar las variables, analizamos cómo estas se correlacionan con el puntaje SIMCE, y esto se realizó tanto a nivel individual, como a nivel curso, pues existe diferencias entre un curso y otro, que muchas veces son visibles al observar la media y/o mediana curso en determinadas componentes y su relación con el puntaje promedio de cada curso. Al mismo tiempo, como se posee información del pretest, buscamos que las variables seleccionadas tampoco sean tales que expliquen lo que el pretest ya incluye como variable.

Cuando analizamos las variables a nivel alumno, consideramos también la correlación con el pretest Aptus, puesto que idealmente nos gustaría que estas variables puedan ser un sustituto al pretest, dado que son de una obtención más fácil y económica que la realización de un pretest.

Primeramente, al considerar variables a nivel alumno y calcular la correlación de estas con el puntaje SIMCE y Aptus de cada estudiante, se estableció tres criterios mínimos para considerar dicha variable como candidato para ser usada en las regresiones con las que

esperamos predecir el puntaje SIMCE.

1. La correlación con el puntaje SIMCE debe ser mayor a 0.2 (en valor absoluto).
2. La correlación con el puntaje Aptus debe ser mayor a 0.2 (en valor absoluto).
3. Significancia del 1 % de la correlación en ambos casos anteriores.

Con estos dos criterios, las variables seleccionadas y la correlación con los puntajes SIMCE y Aptus de estas se detalla en la siguiente tabla. Aquellas variables que corresponden con el nivel de acuerdo del estudiante con una afirmación presentada en alguno de los cuestionarios, se presentarán como frases entre comillas.

Variable	Correlación con:	
	SIMCE	Aptus
“Matemáticas es fácil para mi”	0.2676	0.2157
“Normalmente me va bien en matemáticas”	0.2153	0.2106
“Me va bien en matemáticas”	0.3031	0.2675
“Me gusta matemáticas”	0.2386	0.2064
“Me gusta estudiar matemáticas”	0.2279	0.2037
“Me gustaría no tener que estudiar matemáticas”	-0.2645	-0.2248
“Matemáticas es aburrida”	-0.2080	-0.2149
“No soy bueno en matemáticas”	-0.3090	-0.3351
“Para mi, matemáticas es más difícil que para muchos otros estudiantes en mi clase”	-0.2848	-0.2912
“Estar en clase de matemáticas me hace sentir triste o enojado”	-0.2363	-0.2270
“Mi comportamiento es un problema para el profesor de matemáticas”	-0.2834	-0.2421
Años de educación formal de la madre (o pareja del padre) del estudiante	0.2412	0.2274
Expectativas de los padres sobre años de educación que tendrá el estudiante	0.2352	0.2058

Tabla 2.1: Variables a nivel alumno que cumplen los criterios de correlación con puntaje SIMCE y Aptus.

En donde no se indicó la significancia de cada teste de correlación dado que todos son menores al criterio del 5 %, es más, el p-valor más alto entre las correlaciones estimadas en estas 13 variables, es de $3.24 \cdot 10^{-6}$, por lo que no es relevante incluir estos valores para cada tests.

Además consideramos la variable indicando el sexo del alumno por el interés que significa esta variable ante una eventual diferencia en desempeño por sexo, pese a que la correlación de esta variable no cumple con los criterios anteriores, pues tiene una correlación de 0.1177 con el puntaje SIMCE y 0.1437 con el puntaje Aptus. Con estas variable y las de la tabla

anterior, tendríamos 14 variables históricas que varían de alumno en alumno.

Por otro lado, para variables a nivel curso, como nuestro objetivo es que estas expliquen el comportamiento del SIMCE a nivel curso/establecimiento, y no las variaciones individuales que ya captura el pretest y otras variables, se consideró dos modelos incluyendo efectos fijos a nivel establecimiento/curso, y el pretest. En estos se tiene el puntaje SIMCE y_i de un alumno i en un establecimiento j , con un puntaje de pretest p_i :

$$y_i = \beta_0 + \sum_{j=1}^{24} \beta_j E_{i,j} \quad (2.1a)$$

$$y_i = \beta_0 + \sum_{j=1}^{24} \beta_j E_{i,j} + \gamma p_i \quad (2.1b)$$

donde $E_{i,j}$ indica si el alumno i pertenece al establecimiento j , es decir

$$E_{i,j} = \begin{cases} 1 & \text{el estudiante } i \text{ está en el establecimiento } j \\ 0 & \sim \end{cases}$$

Luego, dependiendo si estamos considerando la media curso de las variables o la mediana, fijamos criterios para seleccionar las variables:

1. La correlación con la media curso del SIMCE debe tener un valor absoluto mayor a 0.25.
2. La correlación con los efectos fijos curso del modelo 2.1a debe tener un valor absoluto mayor a 0.25.
3. La correlación con los efectos fijos curso del modelo 2.1b debe tener un valor absoluto mayor a 0.25.
4. Significancia del 10% de la correlación en los 3 casos anteriores.

donde en los casos que tanto media como mediana de una variable cumplan con los criterios, se tomará aquella que tenga mayor correlación con el promedio SIMCE.

Con esto, se obtiene dos grupos de variables, aquellas que cumplen los criterios a nivel media y aquellas que lo cumplen con la mediana. El conjunto de estos grupos está compuesto de las siguientes variables:

Variable	Correlación de media con:			Correlación de mediana con:		
	media SIMCE	β_j en 2.1a	β_j en 2.1b	media SIMCE	β_j en 2.1a	β_j en 2.1b
“Me va bien en matemáticas”	0.5096 (0.0110)	0.5218 (0.0089)	0.5365 (0.0069)	0.4368 (0.0328)	0.4609 (0.0234)	0.4729 (0.0196)
“Siento temor que las pruebas sean difíciles para mí”	-0.4886 (0.0154)	-0.4944 (0.0141)	-0.4611 (0.0233)	-0.3545 (0.0892)	-0.3897 (0.0598)	-0.3274 (0.1184)
“La inteligencia es algo que no se puede cambiar mucho”	-0.4046 (0.0498)	-0.3801 (0.0669)	-0.4372 (0.0326)	-0.1977 (0.3545)	-0.2099 (0.3249)	-0.2161 (0.3104)
“Mis padres dicen que soy buen estudiante”	0.4550 (0.0255)	0.3683 (0.0766)	0.4863 (0.0160)	0.4030 (0.0508)	0.3464 (0.0973)	0.4127 (0.0451)
“Mis padres dicen que soy capaz de aprender”	0.6189 (0.0013)	0.6368 (0.0008)	0.5992 (0.0020)	0.2210 (0.2994)	0.1452 (0.4983)	0.1841 (0.3892)
“Mi comportamiento es un problema para el profesor de matemáticas”	-0.6763 (0.0003)	-0.6882 (0.0002)	-0.6899 (0.0002)	-0.4612 (0.0233)	-0.4900 (0.0151)	-0.4374 (0.0326)
“Disfruto teniendo en cuenta diferentes perspectivas”	0.4606 (0.0235)	0.5107 (0.0108)	0.4877 (0.0156)	0.4102 (0.0465)	0.4225 (0.0397)	0.4453 (0.0292)
“Prefiero trabajar como parte de un equipo a trabajar solo”	0.3246 (0.1217)	0.3349 (0.1097)	0.2569 (0.2255)	0.4059 (0.0491)	0.3673 (0.0774)	0.3463 (0.0973)
“Leo en mi casa todos los días”	0.2103 (0.3241)	0.1590 (0.4581)	0.1802 (0.3994)	0.5212 (0.0090)	0.5381 (0.0067)	0.5036 (0.0121)
Nota a torneos espiral	0.4134 (0.0446)	0.3460 (0.0976)	0.4298 (0.0361)	0.3175 (0.1306)	0.3041 (0.1485)	0.3525 (0.0911)
Expectativas de los padres sobre años de educación que tendrá el estudiante	0.5547 (0.0049)	0.4568 (0.0249)	0.5513 (0.0052)	0.2985 (0.1566)	0.3311 (0.1140)	0.3006 (0.1535)
SIMCE histórico (Promedio SIMCE del establecimiento en los años 2014-2016)	0.6822 (0.0002)	0.5940 (0.0022)	0.6814 (0.0002)	0.6822 (0.0002)	0.5940 (0.0022)	0.6814 (0.0002)
Años del profesor como docente	0.3682 (0.0767)	0.4014 (0.0519)	0.3570 (0.0868)	0.3682 (0.0767)	0.4014 (0.0519)	0.3570 (0.0868)

Tabla 2.2: Variables a nivel alumno que cumplen los criterios de correlación con puntaje SIMCE y Aptus.

Luego, a estas 13 variables agregamos otra variable a nivel establecimiento que es el IVE (Índice de Vulnerabilidad Escolar) por ser una variable de interés al igual que lo era el sexo a nivel alumno.

Con esto tenemos un conjunto de **28 variables históricas** que probamos para determinar cuáles resultan mejores regresores para un modelo lineal.

2.3. Variables dinámicas

Del capítulo anterior, tenemos 10 variables dinámicas, las cuales para un alumno i , de un establecimiento j , para el eje de contenidos k en el mes t serán:

1. $e_{i,j,k,t}$: Número de ejercicios contestados.
2. $f_{i,j,k,t}$: Número de ejercicios contestados en el primer intento.
3. $g_{i,j,k,t}$: Número de ejercicios contestados en más de un intento.
4. $d_{i,j,k,t}$: Diferencia entre el número de ejercicios contestados en el primer intento y los contestados en más de un intento.
5. $a_{i,j,k,t}$: Precisión de respuesta en el primer intento, calculada como el porcentaje de ejercicios resueltos en el primer intento.
6. $n_{i,j,k,t}$: Calificación promedio en los ejercicios.
7. $o_{i,j,k,t}$: Número de preguntas abiertas contestadas.
8. $p_{i,j,k,t}$: Total de palabras contestadas en preguntas abiertas.
9. $q_{i,j,k,t}$: Promedio de palabras por pregunta abierta contestada.
10. $r_{i,j,k,t}$: Calificación promedio en preguntas abiertas.

Cada una de estas variables será normalizada para cada mes y eje, lo cual será importante tener en cuenta para los análisis posteriores.

A estas variables, se suma una creada usando las 10 anteriores (ya normalizadas). Para esto, se considera un mes dado y se realiza un análisis de componentes principales (ACP), del cual se selecciona las primeras 5 componentes principales. Con esto tendremos además:

11. $pc_{i,j,k,t}$ Componente principal i k -ésima usando las variables dinámicas en el mes t .

Para entender cómo estas componentes están constituidas, procedimos a realizar la separación de los datos en grupos de entrenamiento y testeo 100 veces, para luego analizar en promedio qué variables contribuyen más a las primeras cinco componentes. Veremos esto mediante un enfoque en las 10 variables que tienen mayor participación en la primera componente, pues esta componente concentra la mayor parte de la varianza de la nube de datos.

Veamos los gráficos que muestran el grado de contribución en las primeras 5 componentes para las 10 variables con mayor contribución en la primera componente.

Contribución por variable en $t = 1$

	C.P. 1	C.P. 2	C.P. 3	C.P. 4	C.P. 5
Número de ejercicios – Geometría	8.98	0.04	0.51	0.21	1.77
N. de ej. en 1 intento – Geometría	8.91	0.11	4.06	1.09	0.17
Precisión de respuesta – Geometría	8.91	0.11	4.08	1.09	0.17
Precisión de respuesta – Num. y op.	8.59	0.04	0.61	0.27	2.27
N. de ej. en 1 intento – Num. y op.	8.16	0.03	1.06	1.05	0.55
Número de ejercicios – Num. y op.	8.05	0.06	1.45	1.17	0.63
Nota promedio en ejercicios – Num. y op.	7.32	1.17	0.72	0.26	1.69
Diferencia de ej. en 1 y más intentos – Geometría	7.28	0.21	9.47	3.15	2.07
Nota promedio en ejercicios – Geometría	6.89	0.3	3.22	1.11	0.2
Diferencia de ej. en 1 y más intentos – Num. y op.	6.62	0.38	0.57	0.79	0.42

(a) Contribución de variables al ACP realizado con las variables dinámicas en $t = 1$

Contribución por variable en $t = 2$

	C.P. 1	C.P. 2	C.P. 3	C.P. 4	C.P. 5
Número de ejercicios – Pat. y alg.	9.26	1.75	0.37	0.58	0.13
Precisión de respuesta – Pat. y alg.	8.75	1.67	0.34	0.63	0.11
N. de ej. en 1 intento – Pat. y alg.	8.29	1.88	0.53	1.04	0.17
N. de ej. en más de 1 intento – Pat. y alg.	7.72	1.17	0.16	0.17	0.11
Palabras totales en p. abiertas – Pat. y alg.	7.2	0.35	0.41	0.43	0.5
Número de preguntas abiertas – Pat. y alg.	6.92	0.66	0.57	0.79	0.64
Palabras promedio en p. abiertas – Pat. y alg.	6.92	0.66	0.57	0.79	0.64
N. de ej. en 1 intento – Geometría	4.64	0.38	4.69	8.19	2.9
Número de ejercicios – Geometría	4.58	0.32	4.11	6.94	2.86
Diferencia de ej. en 1 y más intentos – Geometría	4.21	0.41	4.8	8.59	2.66

(b) Contribución de variables al ACP realizado con las variables dinámicas en $t = 2$

Contribución por variable en $t = 3$

	C.P. 1	C.P. 2	C.P. 3	C.P. 4	C.P. 5
N. de ej. en 1 intento – Num. y op.	7.12	0.98	0.59	1.59	5.69
Número de ejercicios – Num. y op.	6.69	0.76	0.41	2.4	6
N. de ej. en 1 intento – Datos y prob.	5.32	1.39	4.75	0.62	2.47
Diferencia de ej. en 1 y más intentos – Num. y op.	5.2	1.11	3.73	1	3.73
Nota promedio en ejercicios – Datos y prob.	4.95	0.78	0.24	1.43	3.99
Precisión de respuesta – Medición	4.88	3.1	0.8	0.99	0.41
Precisión de respuesta – Datos y prob.	4.6	0.87	2.66	1.03	4.26
N. de ej. en 1 intento – Geometría	4.31	1.49	4.04	2.68	2.03
Diferencia de ej. en 1 y más intentos – Datos y prob.	4.27	0.35	1.02	1.84	5.62
Palabras totales en p. abiertas – Num. y op.	3.84	0.44	0.88	5.45	0.92

(c) Contribución de variables al ACP realizado con las variables dinámicas en $t = 3$

Contribución por variable en $t = 4$

	C.P. 1	C.P. 2	C.P. 3	C.P. 4	C.P. 5
N. de ej. en 1 intento – Num. y op.	6.93	0.57	0.29	0.7	0.55
Diferencia de ej. en 1 y más intentos – Num. y op.	5.83	0.27	0.64	0.57	2.55
Número de ejercicios – Num. y op.	5.13	0.7	1.23	2.72	3.18
Precisión de respuesta – Num. y op.	4.56	0.28	0.97	1.13	3.73
Nota promedio en ejercicios – Num. y op.	4.32	0.25	0.31	0.83	4.69
N. de ej. en 1 intento – Pat. y alg.	4.23	3.16	1.9	2.22	0.41
Diferencia de ej. en 1 y más intentos – Medición	3.87	0.95	0.66	0.82	0.57
Precisión de respuesta – Medición	3.61	0.1	0.17	0.61	1.47
N. de ej. en 1 intento – Medición	3.54	2.91	3.43	2.35	3.06
Diferencia de ej. en 1 y más intentos – Geometría	3.16	1.88	0.59	1.03	1.26

(d) Contribución de variables al ACP realizado con las variables dinámicas en $t = 4$

Contribución por variable en $t = 5$

	C.P. 1	C.P. 2	C.P. 3	C.P. 4	C.P. 5
N. de ej. en 1 intento – Num. y op.	5.99	0.35	0.68	3.84	1.47
N. de ej. en 1 intento – Geometría	5.94	1.24	0.16	0.5	4.58
N. de ej. en 1 intento – Medición	5.23	2.28	1.68	2.74	0.34
N. de ej. en 1 intento – Pat. y alg.	5.16	0.12	4.63	0.28	1.49
Diferencia de ej. en 1 y más intentos – Num. y op.	4.86	1.3	0.16	1.01	1.92
Número de ejercicios – Num. y op.	4.77	1.79	0.96	4.75	1.02
Número de ejercicios – Geometría	4.3	4.34	0.17	0.48	2.34
Número de ejercicios – Medición	4.14	5.45	1.21	1.92	0.15
Nota promedio en ejercicios – Num. y op.	4.02	2.37	0.21	0.99	1.31
Diferencia de ej. en 1 y más intentos – Geometría	3.82	1.01	0.09	0.6	6.25

(e) Contribución de variables al ACP realizado con las variables dinámicas en $t = 5$

Contribución por variable en $t = 6$

	C.P. 1	C.P. 2	C.P. 3	C.P. 4	C.P. 5
N. de ej. en 1 intento – Medición	6.65	1.91	1.65	1.34	0.74
Diferencia de ej. en 1 y más intentos – Medición	6.01	0.22	0.46	2.1	1.48
Número de ejercicios – Medición	5.41	3.37	2.73	0.91	0.84
Nota promedio en ejercicios – Medición	4.85	0.3	0.32	1.28	0.75
Precisión de respuesta – Medición	4.73	0.58	0.56	1.36	1.18
Número de preguntas abiertas – Medición	4.64	0.3	2.37	1.95	3.28
Nota promedio en ejercicios – Geometría	3.82	0.59	0.75	0.3	0.83
Palabras totales en p. abiertas – Medición	3.8	0.65	2.82	1.55	2.41
Número de preguntas abiertas – Num. y op.	3.7	0.93	0.3	1.62	3.7
Palabras promedio en p. abiertas – Medición	3.7	0.23	1.99	1.98	3.26

(f) Contribución de variables al ACP realizado con las variables dinámicas en $t = 6$

Contribución por variable en $t = 7$

	C.P. 1	C.P. 2	C.P. 3	C.P. 4	C.P. 5
N. de ej. en 1 intento – Pat. y alg.	6.12	1.18	2.04	2.48	1.03
Diferencia de ej. en 1 y más intentos – Pat. y alg.	6.08	0.88	2.04	1.74	1.29
Número de ejercicios – Pat. y alg.	5.7	1.54	2.17	3.03	0.84
N. de ej. en 1 intento – Medición	5.3	3.37	0.81	0.76	0.5
Palabras promedio en p. abiertas – Medición	4.92	5.15	0.37	0.3	0.44
Palabras totales en p. abiertas – Medición	4.86	5.82	0.4	0.19	0.42
Número de preguntas abiertas – Medición	4.84	5.15	0.39	0.29	0.46
Número de ejercicios – Medición	4.67	4.85	1.2	0.64	0.41
Nota promedio en ejercicios – Pat. y alg.	4.24	1.7	1.86	0.56	0.43
Precisión de respuesta – Pat. y alg.	3.69	1.67	1.78	0.61	0.58

(g) Contribución de variables al ACP realizado con las variables dinámicas en $t = 7$

Contribución por variable en $t = 8$

	C.P. 1	C.P. 2	C.P. 3	C.P. 4	C.P. 5
N. de ej. en 1 intento – Datos y prob.	4.96	1.8	0.64	0.2	1.73
N. de ej. en 1 intento – Medición	4.65	0.69	1.9	0.54	3.95
N. de ej. en 1 intento – Geometría	4.26	0.61	2	3.01	5.02
N. de ej. en 1 intento – Num. y op.	4.15	0.98	3.5	4.28	0.97
Diferencia de ej. en 1 y más intentos – Geometría	3.97	1.84	0.67	0.17	1.34
Diferencia de ej. en 1 y más intentos – Datos y prob.	3.92	0.38	0.15	1.02	1.96
Nota promedio en ejercicios – Datos y prob.	3.84	1.61	0.17	0.34	0.34
Número de ejercicios – Datos y prob.	3.71	3.61	1.09	0.57	1.22
Número de ejercicios – Num. y op.	3.67	0.82	3.34	6.71	1.2
Nota promedio en ejercicios – Num. y op.	3.65	3.54	1.42	0.14	0.62

(h) Contribución de variables al ACP realizado con las variables dinámicas en $t = 8$

Luego, con estos resultados podremos analizar el comportamiento de un modelo que haga uso de esta variable.

Un primer análisis es observar aquellas variables que frecuentemente están entre las 10 que más contribuyen a la primera componente, es decir aquellas que se repiten para distintos valores de t . Entre ellas destaca

1. Número de ejercicios contestados en el primer intento - Medición (6 veces).
2. Número de ejercicios contestados en el primer intento - Patrones y álgebra (5 veces).

3. Número de ejercicios contestados en el primer intento - Geometría (5 veces).
4. Número de ejercicios contestados - Números y operaciones (5 veces).
5. Número de ejercicios contestados - Medición (4 veces).
6. Número de ejercicios contestados en el primer intento - Números y operaciones (4 veces).
7. Diferencia entre el número de ejercicios contestados en el primer intento y los contestados en más de un intento - Números y operaciones (4 veces).
8. Diferencia entre el número de ejercicios contestados en el primer intento y los contestados en más de un intento - Geometría (4 veces).

donde es evidente que el número de ejercicio que contesta cada alumno es lo que genera mayor variabilidad en los datos, en especial el número de ejercicios contestados en el primer intento.

2.4. Regresiones lineales

Para evaluar la combinación de variables que muestre el menor error de estimación de puntaje SIMCE, se hará uso de 3 grupos de variables: el pretest, las variables históricas y las variables dinámicas. Nuevamente se presenta la dificultad de tener muchas variables como para considerar todas las combinaciones posibles de estas, además que queremos encontrar variables que de alguna forma se complementen con la información recopilada mediante la plataforma, es decir, con las variables ConectaIdeas. Con este objetivo se limita el número de variables históricas a un máximo de 2, igualmente las variables dinámicas. Debido a esto, consideramos modelos de la siguiente forma:

$$\hat{y}_{i,j,t} = \gamma_t p_i + \beta_0 + \beta_{1,t} \Theta_{i,j}^{(1)} + \beta_{2,t} \Theta_{i,j}^{(2)} + \delta_t^{(1)} U_{i,j,t}^{(1)} + \delta_t^{(2)} U_{i,j,t}^{(2)} \quad (2.2)$$

donde:

- $\hat{y}_{i,j,t}$ es la estimación de puntaje SIMCE del alumno i en un establecimiento j , con la información disponible en t .
- p_i es el puntaje en el pretest del alumno i .
- $\Theta_{i,j}^m$ corresponde a la m -ésima variable histórica en el modelo, $m \in \{1, 2\}$, para el alumno i del establecimiento j .
- $U_{i,j,t}^m$ corresponde a la m -ésima variable dinámica en el modelo, $m \in \{1, 2\}$, para el alumno i del establecimiento j , en el instante t .

Es decir, consideraremos modelos con el pretest, un intercepto β_0 , 2 variables históricas y 2 variables dinámicas disponibles en el mes t .

Se debe hacer la distinción entre el superíndice m de las variables históricas y dinámicas, indicando si es la primera o segunda variable histórica (o dinámica) del modelo y esto no indicará cuál de las 28 variables históricas (u 11 variables dinámicas) es la que está siendo considerada en el modelo.

Es importante notar que debido a que la estimación $\hat{y}_{i,j,t}$ depende de la información disponible, o hasta qué mes t esta está disponible, para cada uno de los modelos, se puede construir el modelo para cada $t \in \{1, \dots, 8\}$, por lo que habría 8 sub-modelos, uno para cada t . Sin embargo, queremos que nuestro intercepto se mantenga constante independiente de t , y es por eso que este es el único parámetro que no depende de t .

Para obtener estos parámetros, consideremos que existe un $\tilde{\beta}_0$ dado y queremos estimar el modelo para un t dado. Entonces se resuelve:

$$(P_t) \quad \begin{aligned} \min_{\theta} \quad & Error(\tilde{\beta}_0, \theta_t) = \sum_{i=1}^N (y_i - \hat{y}_{i,j,t})^2 \\ \text{s.a:} \quad & \theta_t = \left(\gamma_t, \beta_{1,t}, \beta_{2,t}, \delta_t^{(1)}, \delta_t^{(2)} \right) \\ & \hat{y}_{i,j,t} = \gamma_t p_i + \tilde{\beta}_0 + \beta_{1,t} \Theta_{i,j}^{(1)} + \beta_{2,t} \Theta_{i,j}^{(2)} + \delta_t^{(1)} U_{i,j,t}^{(1)} + \delta_t^{(2)} U_{i,j,t}^{(2)} \end{aligned} \quad (2.3)$$

donde N es el número de estudiantes y recordamos que y_i es el puntaje SIMCE del alumno i (como en 2.1). Luego, para la estimación de β_0 se resuelve:

$$(P) \quad \begin{aligned} \min_{\beta_0} \quad & \frac{1}{8} \sum_{t=1}^8 \sqrt{\frac{1}{N} Error(\beta_0, \theta_t)} \\ \text{s.a:} \quad & \theta_t \in \underset{\tilde{\theta}_t}{\text{argmin}}(P_t) \end{aligned} \quad (2.4)$$

Lo que será importante considerar, es el hecho que las variables dinámicas están desglosadas por eje $k \in \{1, \dots, 5\}$ (por los 5 ejes de contenidos), es decir variables de forma $u_{i,j,k,t}$, por lo que las variables $U_{i,j,t}^{(m)}$ serán vectores de tamaño 5, pero por simpleza y facilidad a la hora de trabajar con distintas realizaciones de cada variable, preferiremos usar una notación donde sea explícito el eje asociado mediante un subíndice k . A su vez esto nos permitirá considerar la variable de dos formas:

1. $u_{i,j,k,t}$, será la variable dinámica para el alumno i del establecimiento j , asociada al eje k en el mes t .
2. $u_{i,j,k,1:t}$, será la variable dinámica para el alumno i del establecimiento j , asociada al eje k hasta el mes t .

Esto inducirá diferencias en algunos casos. Por ejemplo, si la variable dinámica usada en el modelo es el “Número de ejercicios resueltos” ($e_{i,j,k,t}$), ocurre que

$$e_{i,j,k,1:t} = \sum_{s=1}^t e_{i,j,k,s}$$

mientras que, si por ejemplo la variable es la “Precisión de respuesta en el primer intento” ($a_{i,j,k,t}$), esto no ocurre puesto que

$$a_{i,j,k,1:t} \neq \sum_{s=1}^t a_{i,j,k,s}$$

dado que la precisión de respuesta al primer intento la definimos como

$$\text{Precisión de respuesta} = \frac{\text{Número de ejercicios contestados en el primer intento}}{\text{Número de ejercicios contestados}}$$

entonces al considerar $a_{i,j,k,1:t}$ sería este cociente considerando el total de ejercicios hasta t , mientras que la suma será de las proporciones mensuales.

Esta formulación se justifica por el hecho que a la hora de realizar estimaciones del puntaje SIMCE en futuras instancias de ConectaIdeas, no se tendrá a disposición la información de todos los meses entre el instante 0 y el instante del SIMCE T , que al ser meses, $T = 9$ (noviembre). Por esto es importante considerar estimaciones que posean información limitada hasta cierto mes $t < T$. Entonces ¿qué representa el producto $\delta_t^m U_{i,j,t}^m$? Como respuesta a esta pregunta, consideramos 3 alternativas:

$$\delta_t^{(m)} U_{i,j,t}^{(m)} = \sum_{k=1}^5 \delta_{k,t}^{(m)} u_{i,j,k,t}^{(m)} \quad (2.5a)$$

$$\delta_t^{(m)} U_{i,j,t}^{(m)} = \sum_{s=1}^t \sum_{k=1}^5 \delta_{k,s}^{(m)} u_{i,j,k,s}^{(m)} \quad (2.5b)$$

$$\delta_t^{(m)} U_{i,j,t}^{(m)} = \sum_{k=1}^5 \delta_{k,t}^{(m)} u_{i,j,k,1:t}^{(m)} \quad (2.5c)$$

donde $u_{i,j,k,1:t}^m$ representa la variable dinámica m -ésima del modelo ($m \in \{1, 2\}$), para el alumno i del establecimiento j , en el eje k , desde el mes 1 al mes t , es decir durante el transcurso de todos esos meses.

Luego, considerando las expresiones 2.5a - 2.5c en 2.2 se obtiene los 3 modelos:

$$\widehat{y}_{i,j,t} = \gamma_t p_i + \beta_0 + \beta_{1,t} \Theta_{i,j}^{(1)} + \beta_{2,t} \Theta_{i,j}^{(2)} + \sum_{k=1}^5 \delta_{k,t}^{(1)} u_{i,j,k,t}^{(1)} + \sum_{k=1}^5 \delta_{k,t}^{(2)} u_{i,j,k,t}^{(2)} \quad (\text{sin memoria})$$

$$\widehat{y}_{i,j,t} = \gamma_t p_i + \beta_0 + \beta_{1,t} \Theta_{i,j}^{(1)} + \beta_{2,t} \Theta_{i,j}^{(2)} + \sum_{s=1}^t \sum_{k=1}^5 \delta_{k,s}^{(1)} u_{i,j,k,s}^{(1)} + \sum_{s=1}^t \sum_{k=1}^5 \delta_{k,s}^{(2)} u_{i,j,k,s}^{(2)} \quad (\text{glotón})$$

$$\widehat{y}_{i,j,t} = \gamma_t p_i + \beta_0 + \beta_{1,t} \Theta_{i,j}^{(1)} + \beta_{2,t} \Theta_{i,j}^{(2)} + \sum_{k=1}^5 \delta_{k,t}^{(1)} u_{i,j,k,1:t}^{(1)} + \sum_{k=1}^5 \delta_{k,t}^{(2)} u_{i,j,k,1:t}^{(2)} \quad (\text{con memoria})$$

Notamos entonces que el modelo sin memoria considera solo la información del mes t del cual se tenga información disponible, con efectos δ por cada eje; mientras que el modelo glotón considera la información de todos los meses hasta t , considerando un efecto δ asociado para cada mes y a cada eje, mientras que el modelo con memoria considera la información hasta t , pero los efectos δ son tan solo uno por cada eje.

Se debe considerar también que los efectos $\beta_{1,t}$ y $\beta_{2,t}$ que acompañan las variables históricas y el efecto γ_t que va junto al pretest, también dependen de t y tan solo el intercepto β_0 es transversal en el tiempo. Por esto, podemos hacer el siguiente análisis:

$$\begin{array}{r} \text{Modelos 2.6} \\ \left\{ \begin{array}{ll} \beta_0 & \rightarrow 1 \\ \beta_{1,1}, \dots, \beta_{1,8} & \rightarrow 8 \\ \beta_{2,1}, \dots, \beta_{2,8} & \rightarrow 8 \\ \gamma_1, \dots, \gamma_8 & \rightarrow 8 \end{array} \right. \end{array} \quad \begin{array}{l} \text{Parámetros} \\ \\ \\ \\ \text{Total: 25} \end{array}$$

Entonces, todos los modelos podrán tener 25 efectos dados por el intercepto, pretest y las hasta dos variables históricas.

Para las variables dinámicas, por ejemplo si consideramos la información disponibles hasta $t = 8$, esto es hasta octubre, justo antes del SIMCE, los modelos sin memoria y con memoria usan 5 efectos δ por variable dinámica, mientras que glotón usará 40 de estos. Ahora, esto es considerando el caso para un t dado, mientras que al ver el modelo de manera transversal para distintas capacidades de la disponibilidad de la información, es decir en la medida que aumenta el valor del mes t hasta el cual se dispone información, la cantidad de parámetros totales de los modelos, es no despreciable.

Entonces hay que tener en cuenta la cantidad de variables dinámicas que se requiere en cada mes, en especial en caso del modelo glotón que irá aumentando puesto el número de efectos $\delta^{(m)}$ en este dependen del instante t , como se puede notar a continuación en el análisis que se hace para uno de los efectos $\delta^{(m)}$ de los 3 modelos.

Modelo	Mes t de información								Total de parámetros
	1	2	3	4	5	6	7	8	
Modelo sin memoria	5	5	5	5	5	5	5	5	40
Modelo glotón	5	10	15	20	25	30	35	40	180
Modelo con memoria	5	5	5	5	5	5	5	5	40

Tabla 2.3: Número de parámetros $\delta^{(m)}$ por modelo, $m \in \{1, 2\}$

Esto es algo que tendremos que tener en consideración como una de las eventuales ventajas o desventajas de cada modelo a la hora de determinar la prevalencia de uno por sobre otro, puesto que si bien todos comparten el mismo número de parámetros acompañando las variables históricas, el pretest y el intercepto, son los efectos de las variables dinámicas las que generan la gran diferencia en el número de parámetros que requiere cada modelo, que al ser dos variables dinámicas, esta diferencia es aun mas acentuada. En resumen

Modelo	Parámetro Intercepto	Parámetro Pretest	Parámetros V. Históricas	Parámetros V. Dinámica	Total
	b_0	γ_t	$\beta_{m,t}$	$\delta_{k,t}^{(m)}$	
sin memoria	1	8	16	80	105
glotón	1	8	16	360	385
con memoria	1	8	16	80	105

Tabla 2.4: Número máximo de parámetros por modelo

Ahora, este número de parámetros por modelo será el máximo posible dado la cantidad de variables históricas y dinámicas que estamos considerando, sin embargo, en combinaciones donde tan solo se elige una variable histórica y/o una variable dinámica, la cantidad de parámetros decrece acorde a esto.

2.5. Resultados

De los 500 alumnos en el grupo tratamiento, de los cuales se tiene información de pretest, SIMCE y ejercitación del programa, se elige separar los grupos de entrenamiento y testeo para evaluar el desempeño de los modelos. Esto se hace para probar la predicción en un grupo de datos distinto del empleado para entrenarlo. Más aun, para que dicha prueba sea cercana a lo que serán estimaciones fuera de la muestra, la selección de grupo de entrenamiento y testeo será por establecimiento, de modo que los datos con los cuales después probaremos el desempeño de un modelo sean efectivamente de colegios distintos. Entonces, se fija un grupo de entrenamiento correspondiente al 75 % de los establecimientos y por consecuencia un grupo de testeo del 25 %, es decir, 18 colegios formarán el grupo de entrenamiento mientras que los 6 restantes serán el grupo de testeo.

Teniendo el grupo de entrenamiento, se usa la media y desviación de cada una de sus variables para normalizar tanto los datos de entrenamiento como de testeo. Al realizar esta normalización se tiene en cuenta que para normalizar las variables históricas que corresponden al promedio curso de alguna característica, dicha normalización es realizada restando el promedio entre los 18 establecimientos en entrenamiento y dividiendo por la desviación de estos. Por otro lado, el resto de las variables, al variar por alumno, se normalizan tomando el promedio y desviación entre todos los alumnos del grupo de entrenamiento.

La única variable que no estará normalizada para estos modelos, será la variable a predecir y_i , el puntaje SIMCE; el resto de las variables serán normalizadas y por lo tanto el tamaño de los efectos se deberá interpretar en términos de desviaciones estándar de dicha variable.

Para elegir la mejor combinación de variables, en modelos lineales se suele utilizar técnicas de selección de variables como stepwise regression, sin embargo debido a la particular estructura de los modelos y la existencia de variables que son transversales en el tiempo como el intercepto, mientras las otras varían según t , no es directo implementar estas técnicas para evaluar estos modelos.

La estrategia usada para elegir las mejores combinaciones de variables fue fijar un par de

entrenamiento y testeo, para luego simplemente considerar las distintas combinaciones con y sin pretest, con hasta 2 variables históricas y 1 o 2 de las primeras 10 variables dinámicas, o la variable dinámica de componentes principales, la cual siempre será considerada sola y no con otra variable dinámica por colinealidad.

$$\underbrace{2}_{\text{Pretest}} \cdot \underbrace{\left[\binom{28}{2} + 28 + 1 \right]}_{\text{V. Históricas}} \cdot \underbrace{\left[\binom{10}{2} + 10 + 1 \right]}_{\text{V. Dinámicas}} = 2 \cdot 407 \cdot 56 = 45584$$

Luego, se consideró 45,584 combinaciones para cada uno de los 3 modelos. Luego, se eligió las 200 combinaciones que tuvieron un mejor desempeño entre estas combinaciones y se llevó a cabo 100 validaciones cruzadas para confirmar que el desempeño sea estable, luego de esto se eligió para cada modelo la combinación de variables que en el promedio de las validaciones cruzadas haya tenido un menor RMSE en el grupo de testeo. Entonces, para cada modelo la combinación óptima de variables fueron las que se detallan a continuación.

Antes de analizar los resultados de cada modelo, definimos 5 baselines como medida de comparación para los modelos que consideramos en este trabajo. Para definir los primeros 3 baselines hicimos uso de las notas normalizadas que ConectaIdeas asigna a los ejercicios, las cuales usamos de la siguiente forma:

1. $N_{i,j,t}$, el promedio de nota de todos los ejercicios que hizo el alumno en t , independiente del eje.
2. $N_{i,j,1:t}$, el promedio de nota de todos los ejercicios que hizo el alumno hasta t , independiente del eje.
3. $\bar{N}_{i,j,t}$, el promedio de nota de todos los ejercicios que hizo el alumno en cada mes, promediado de 1 a t , independiente del eje.

Consideramos además

4. p_i es el puntaje en el pretest del alumno i .
5. m_i serán los resultados del alumno i en el test MideUc de agosto

Teniendo esto, los baselines considerados fueron:

$$\begin{aligned} \hat{y}_{i,j,t} &= \sigma N_{i,j,t} + \mu && \text{(baseline con nota en } t) \\ \hat{y}_{i,j,t} &= \sigma N_{i,j,1:t} + \mu && \text{(baseline con nota hasta } t) \\ \hat{y}_{i,j,t} &= \sigma \bar{N}_{i,j,t} + \mu && \text{(baseline con nota mensual)} \\ \hat{y}_{i,j,t} &= \gamma_t p_i + \beta_0 + \beta_{1,t} \Theta_{i,j}^{(1)} + \beta_{2,t} \Theta_{i,j}^{(2)} && \text{(baseline con pretest)} \\ \hat{y}_{i,j,t} &= \gamma_t m_i + \beta_0 + \beta_{1,t} \Theta_{i,j}^{(1)} + \beta_{2,t} \Theta_{i,j}^{(2)} && \text{(baseline con midtest)} \end{aligned}$$

donde μ es la media SIMCE nacional 2017 y σ la desviación estándar del mismo año. Las variables históricas consideradas son:

1. $\Theta_{i,j}^{(1)} = sh_{ij}$, since histórico del establecimiento j al cual pertenece el alumno i .
2. $\Theta_{i,j}^{(2)} = iv_{ij}$, índice de vulnerabilidad escolar (IVE) del establecimiento j al cual pertenece el alumno i .

Los resultados de estos modelos se presentan a continuación

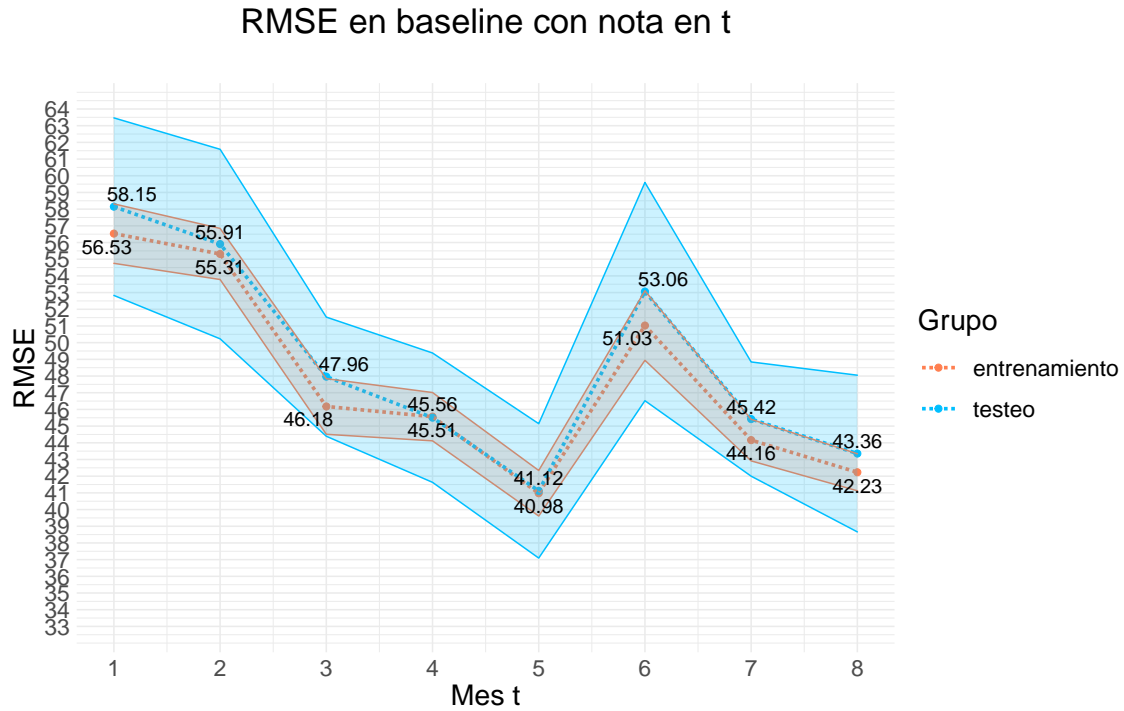


Figura 2.1: RMSE en el baseline con nota en t

RMSE en baseline con nota hasta t

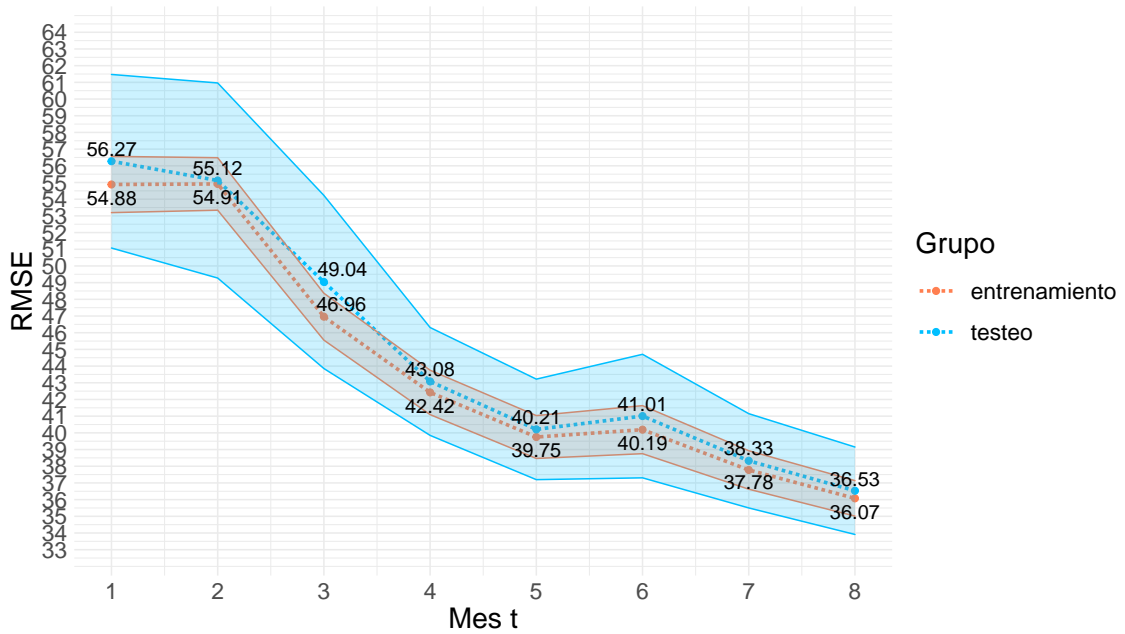


Figura 2.2: RMSE en el baseline con nota hasta t

RMSE en baseline con nota mensual

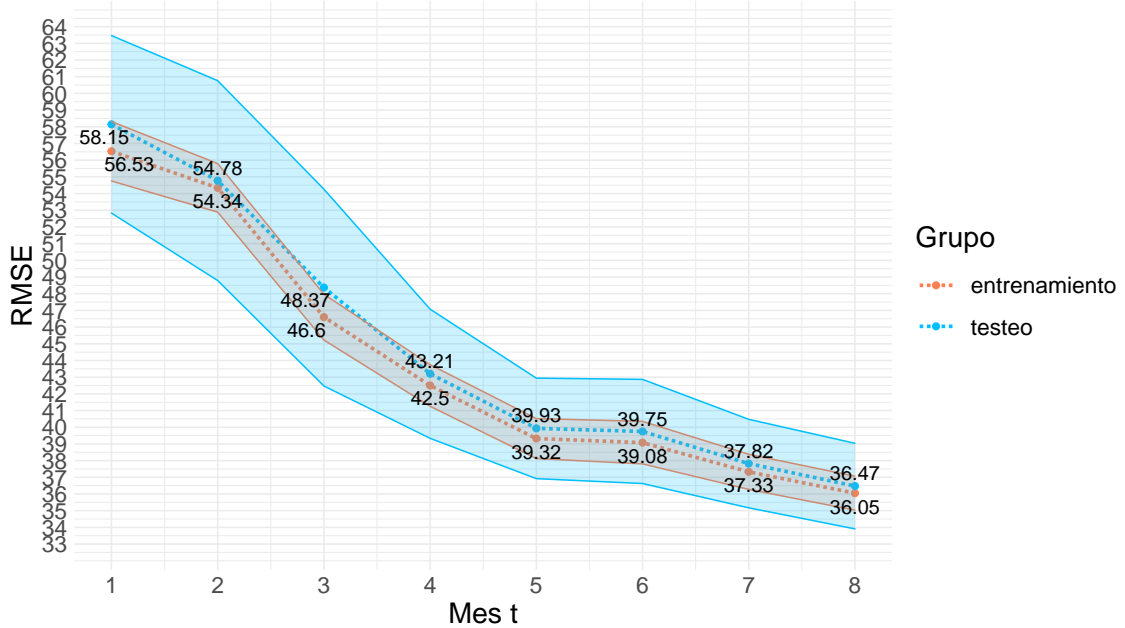


Figura 2.3: RMSE en el baseline con nota mensual

RMSE en baseline con pretest

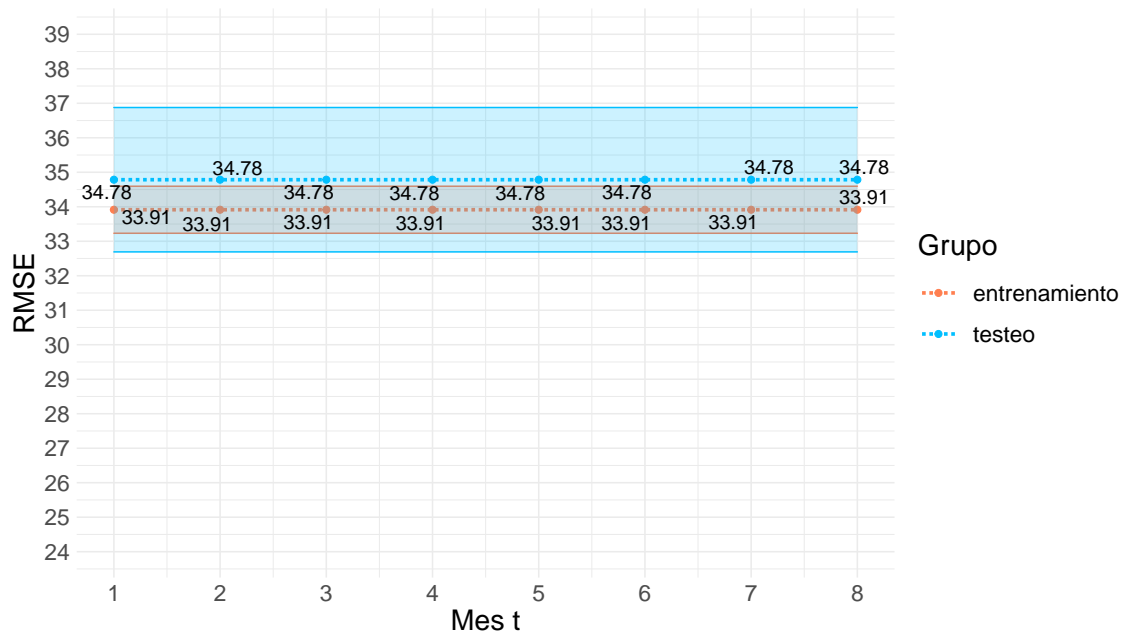


Figura 2.4: RMSE en el baseline con pretest

RMSE en baseline con midtest

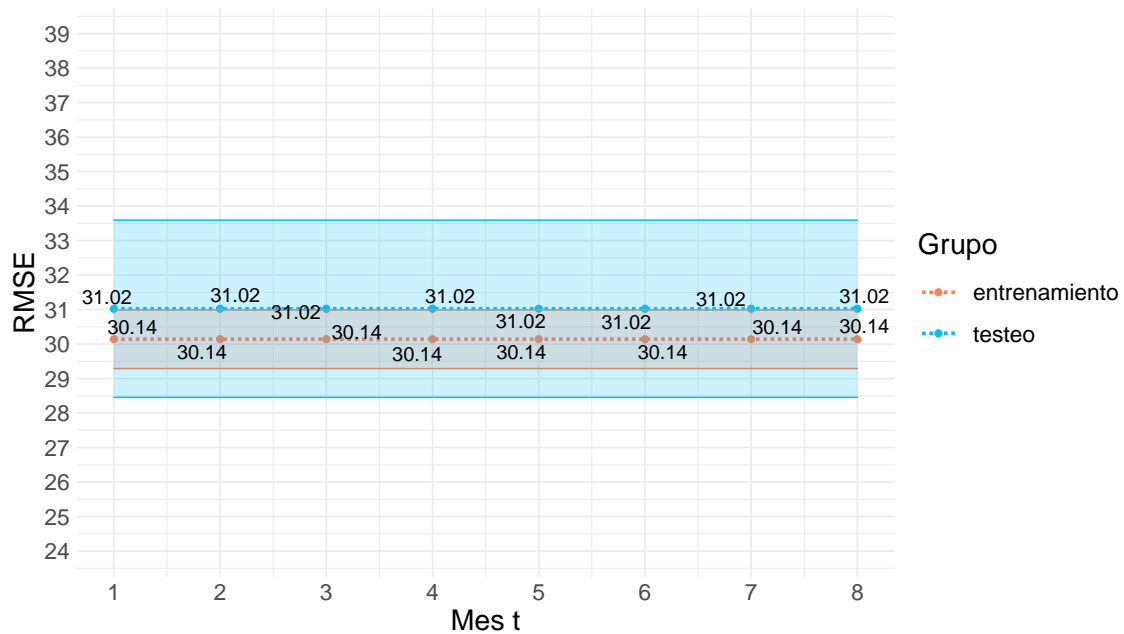


Figura 2.5: RMSE en el baseline con midtest

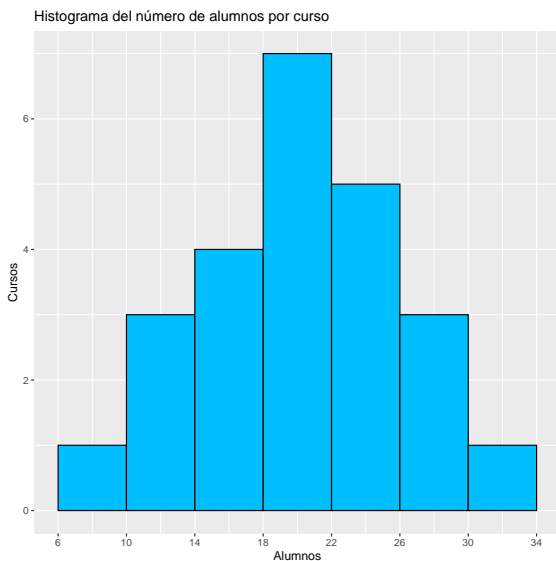
Los resultados en estos baselines nos permite deducir principalmente dos cosas; primero, que simplemente reescalar las notas en los ejercicios no es tan buen predictor para el puntaje SIMCE, pero en la medida que el número de meses de información aumenta (baseline con

nota hasta t y baseline con nota mensual), la predicción mejora significativamente. Por otro lado, los test estandarizados como el pretest de principios de año (baseline con pretest) o el midtest de agosto (baseline con midtest) son buenos predictores comparado con las notas de ejercicios, pero es importante considerar el costo extra que significa llevar a cabo estas mediciones.

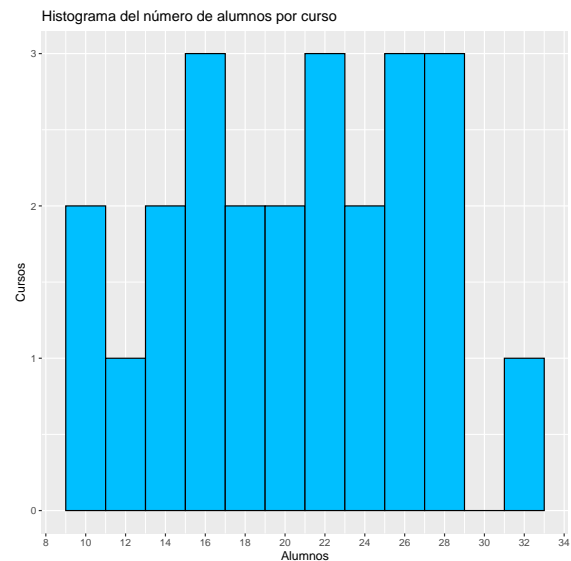
El error en testeo del baseline con pretest será una meta a superar, pues si se considerará el pretest como regresor junto con otras variables, sería indeseable que las otras variables no mejoren los resultados en el grupo de testeo. Al mismo tiempo, los resultados del baseline con midtest serán una meta a superar con la información a partir de agosto ($t=6$) en nuestros modelos. Sin embargo, estaremos acotados inferiormente por el SEM del SIMCE.

De igual modo, usando estos modelos se puede promediar las predicciones por curso de cada uno y comparar con el promedio de los puntajes reales. Haciendo esto se puede obtener una estimación del puntaje curso y calcular el RMSE asociado, el cual también esperamos supere el RMSE del promedio curso para los baselines y que a su vez estará acotado por el SEM del promedio curso, donde será importante considerar el número de alumnos de cada curso.

En la mayoría de los casos la cantidad de alumnos será menor a 30 por varios motivos, pero principalmente porque la cantidad de alumnos que rindieron tanto el pretest de marzo como el posttest (SIMCE) en noviembre es entre 15 y 25 alumnos. Veamos un histograma del número de alumnos por curso en la base de datos de los 24 establecimientos.



(a) Histograma del número de alumnos por curso con bandas de ancho 4



(b) Histograma del número de alumnos por curso con bandas de ancho 2

Donde en promedio los cursos tienen 20.83 alumnos. Esto quiere decir que en promedio deberíamos tener un SEM de $[5.54, 7.26]$, por lo que nuestro RMSE por curso estará acotado inferiormente por estos valores.

Veamos ahora los gráficos del RMSE cursos para los baselines.

RMSE curso en baseline con nota en t

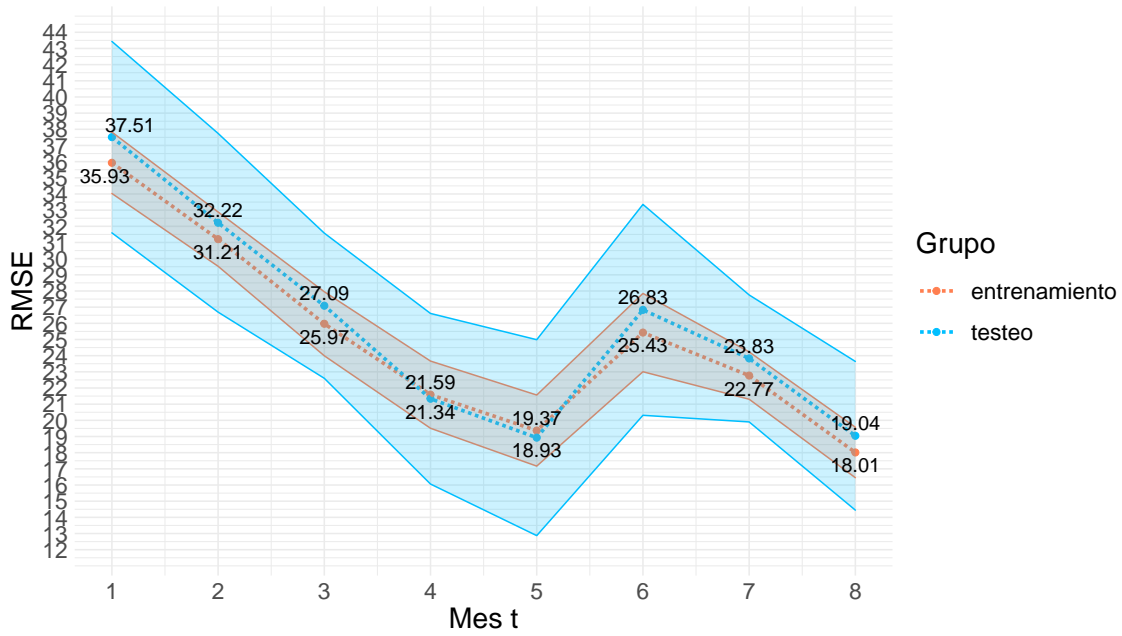


Figura 2.6: RMSE del promedio curso en el modelo baseline con nota en t

RMSE curso en baseline con nota hasta t

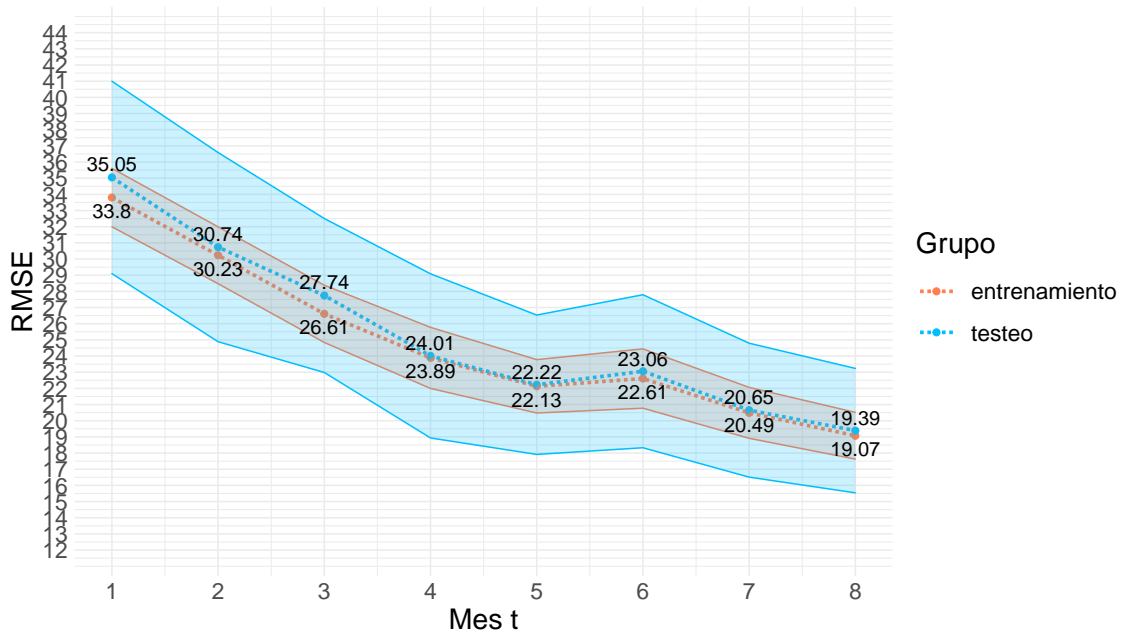


Figura 2.7: RMSE del promedio curso en el modelo baseline con nota hasta t

RMSE curso en baseline con nota mensual

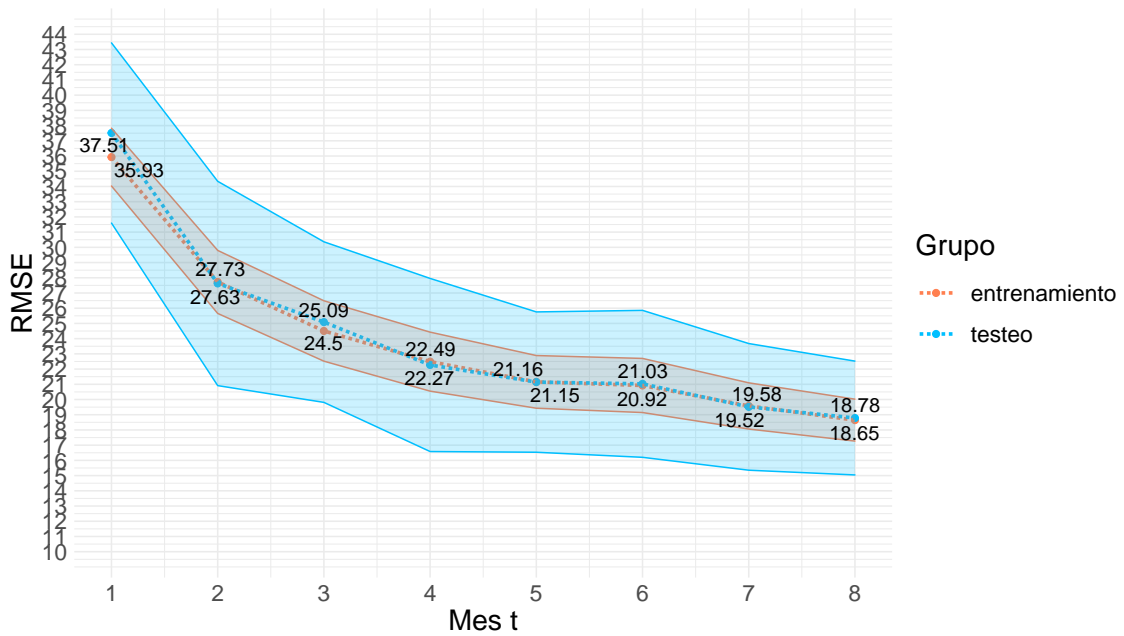


Figura 2.8: RMSE del promedio curso en el modelo baseline con nota mensual

RMSE curso en baseline con pretest

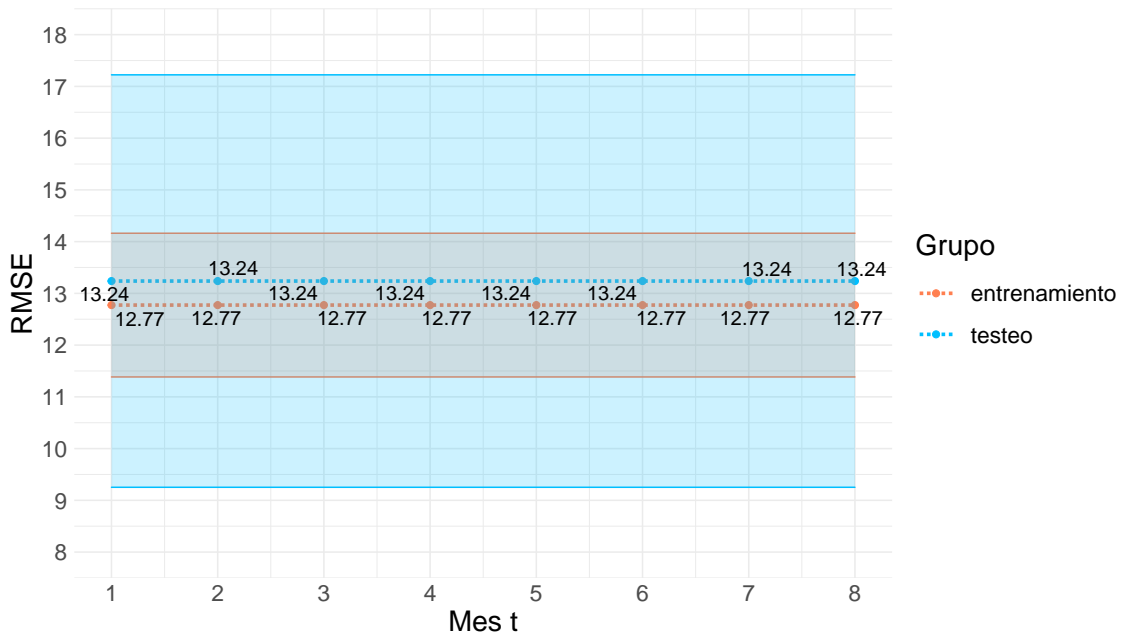


Figura 2.9: RMSE del promedio curso en el modelo baseline con pretest

RMSE curso en baseline con midtest

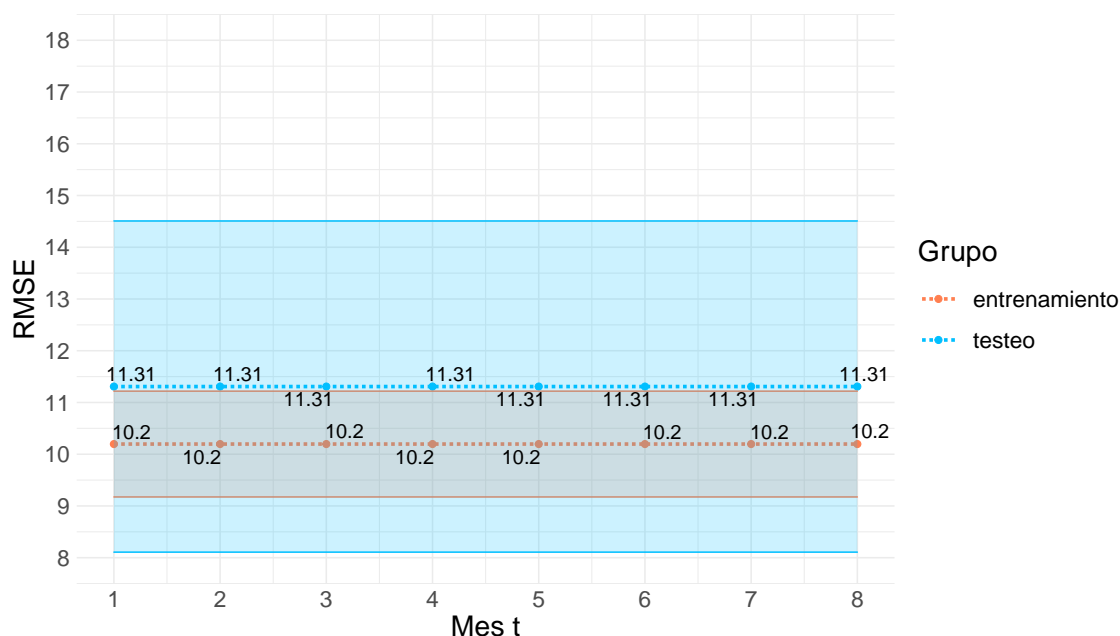


Figura 2.10: RMSE del promedio curso en el modelo baseline con midtest

Con esto podremos tener un punto de referencia para el RMSE del promedio curso.

2.5.1. Modelo sin memoria

Para el primer modelo, las variables que entregaron mejores resultados fueron:

Pretest	Variables históricas	Variables dinámicas
Si	Simce histórico - media curso	Componentes Principales
	“Me va bien en matemáticas” (respuesta por alumno)	

Tabla 2.5: Variables que entregaron un mejor desempeño para el modelo sin memoria

Es decir, si se define:

- $\Theta_{i,j}^{(1)} = sh_{i,j}$, simce histórico del establecimiento j al cual pertenece el alumno i .
- $\Theta_{i,j}^{(2)} = mb_{i,j}$, promedio curso de las respuestas sobre el nivel de acuerdo sobre “Me va bien en matemáticas”, de los alumnos del establecimiento j .
- $u_{i,j,k,t}^{(1)} = pc_{i,j,k,t}$, la componente principal k en el mes t para el alumno i del establecimiento j .
- $u_{i,j,k,t}^{(2)} = 0$.

donde la segunda variable dinámica se dejó en cero porque el modelo con mejor desempeño solo considera una variable dinámica, es decir, el modelo será de la siguiente forma

$$\hat{y}_{i,j,t} = \gamma_t p_i + \beta_0 + \beta_{1,t} sh_{i,j} + \beta_{2,t} mb_{i,j} + \sum_{k=1}^5 \delta_{k,t}^{(1)} pc_{i,j,k,t} \quad (2.8)$$

Esto nos dice que podemos estimar el desempeño SIMCE usando información del pretest, el puntaje simce promedio de los últimos 3 años del establecimiento, el promedio curso de las respuestas sobre el nivel de acuerdo a las afirmaciones “Me va bien en matemáticas”, junto con las primeras 5 componentes principales creadas usando las variables dinámicas obtenidas de la plataforma.

Si analizamos las estimaciones de puntaje realizadas en 100 validaciones cruzadas, error en entrenamiento y testeo se puede ver en el siguiente gráfico.

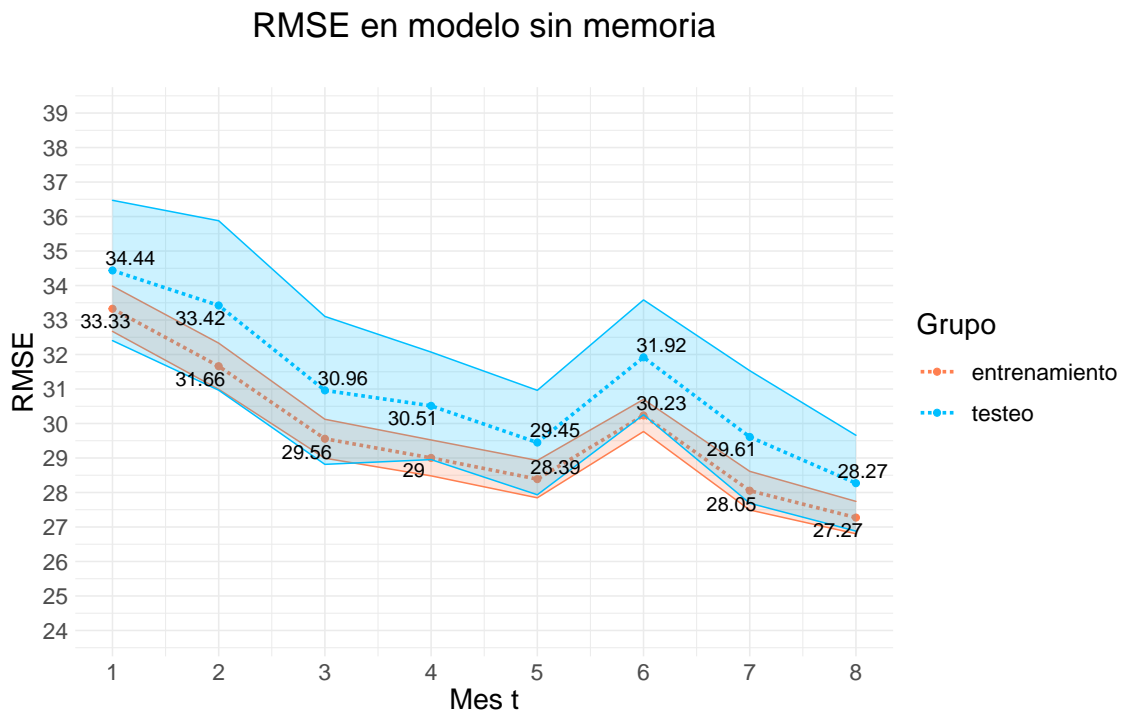


Figura 2.11: RMSE en el modelo sin memoria

donde las bandas de error corresponden a una desviación estándar de los RMSE obtenidos en las 100 validaciones cruzadas. Esto se puede notar en mayor detalle observando los datos de la siguiente tabla.

Meses	1	2	3	4	5	6	7	8
RMSE promedio en entrenamiento	33.33	31.66	29.56	29.00	28.39	30.23	28.05	27.27
RMSE promedio en testeo	34.44	33.42	30.96	30.51	29.45	31.92	29.61	28.27
Desviación RMSE entrenamiento	0.66	0.67	0.56	0.52	0.54	0.47	0.56	0.47
Desviación RMSE testeo	2.03	2.46	2.14	1.56	1.51	1.66	1.92	1.39

Tabla 2.6: RMSE promedio y desviación del RMSE para entrenamiento y testeo en el modelo sin memoria

Primero es importante notar cómo en la medida que aumenta el número de meses t , el error tiende a disminuir, con una diferencia de 6.17 puntos en RMSE entre el primer y último mes en testeo. Esto nos dice que, la información de los meses más cercanos al SIMCE tiende a ser mejor regresor, salvo algunos casos como en $t = 6$, que se aprecia un incremento respecto del mes anterior en entrenamiento; esto quiere decir que el mes de agosto en particular es peor regresor que julio ($t=5$) o septiembre ($t=7$), y una posible explicación a esto es el regreso de vacaciones de invierno y el inicio del segundo semestre.

Ahora, es también de interés comprender la interacción que tiene cada uno de los efectos del modelo en la predicción, para esto primero vamos a apreciar casos de entrenamiento y testeo, y ver cómo la estimación va cambiando en la medida que cambia el mes t de información que están usando las variables dinámicas para predecir el puntaje.

En los siguientes gráficos se muestra algunos casos en cómo este modelo (sin memoria) se implementa, mostrando en una recta horizontal roja el puntaje SIMCE del alumno, y la otra curva ilustra $\hat{y}_{i,j,t}$, es decir, para cada mes t se obtiene una predicción del SIMCE usando la información del mes t indicado. Observamos primero el desempeño en el grupo de entrenamiento.

Casos de entrenamiento

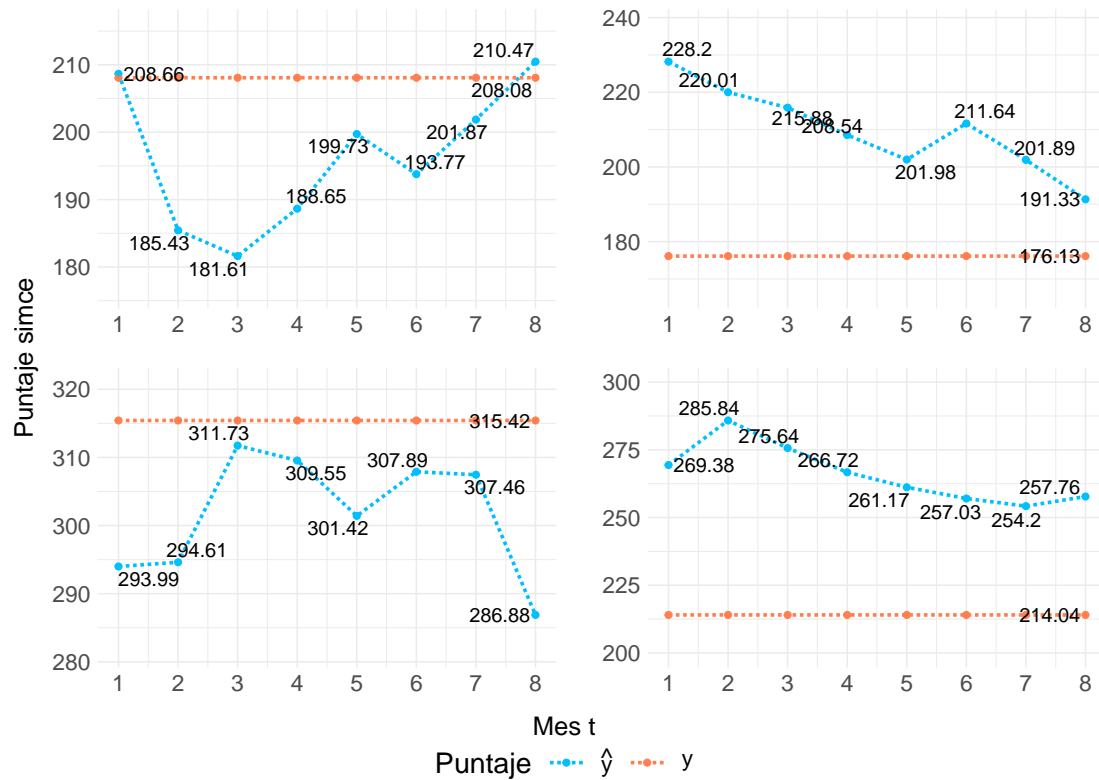


Figura 2.12: Cuatro casos en entrenamiento y sus distintas predicciones según el mes t de información usada

Es posible apreciar que hay una relativa tendencia a obtener mejores predicciones en la medida que nos acercamos al mes del SIMCE ($t = 9$), pero tampoco es muy marcado y en algunos casos simplemente esto no ocurre. Más aún, si apreciamos ahora casos del grupo de testeo, podremos notar que la tendencia no es muy notoria.

Casos de testeo

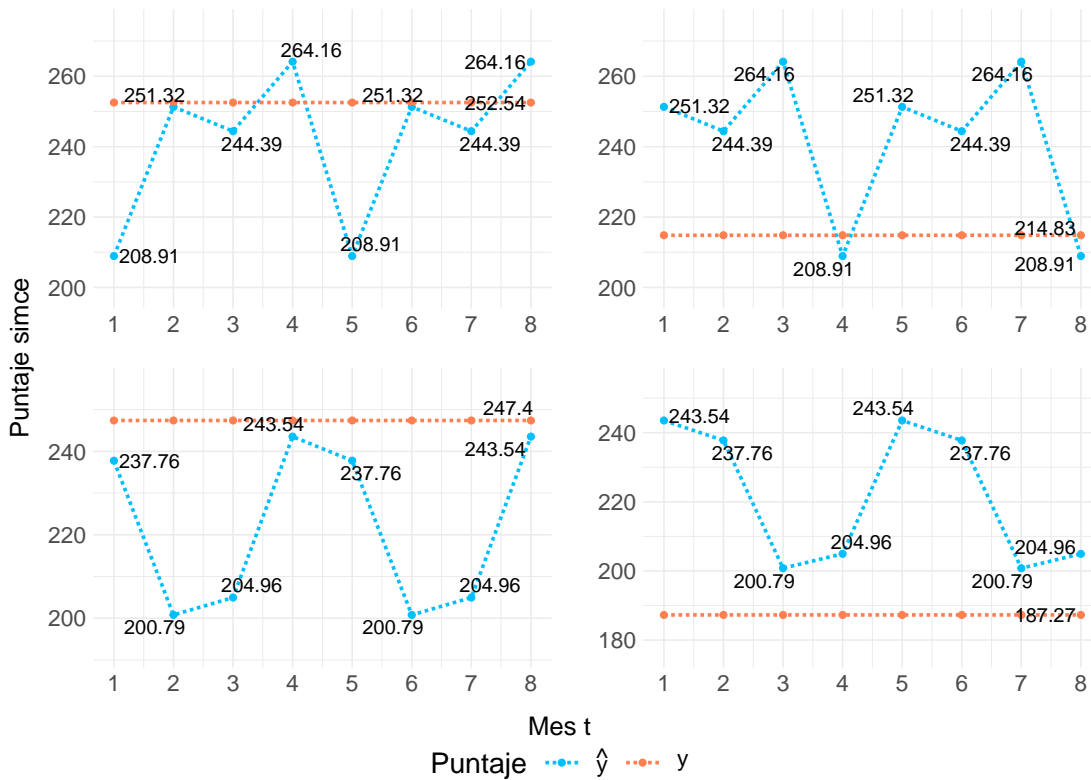


Figura 2.13: Cuatro casos en testeo y sus distintas predicciones según el mes t de información usada

Sin embargo, esto no es ninguna sorpresa puesto que al usar tan solo un mes, es solo ver el desempeño del momento y no permite obtener una visión comprensiva del desempeño del estudiante. Entonces ¿por qué al estar más cerca de la fecha de aplicación del SIMCE la predicción mejora? Esto se debe a que la prueba evalúa los cinco ejes de contenidos, y es en los meses más cercanos a la evaluación cuando el currículo y programa de cada profesor contempla la inclusión de ejercicios de todos los ejes, entonces a medida que el mes sea más cercano a noviembre, el mes de aplicación de la prueba, las variables dinámicas comienzan a presentar mayor diversidad entre ejes, lo que ayuda a una mejor estimación del puntaje.

Ahora, al igual que con el RMSE obtenido de las estimaciones para el puntaje de cada alumno, también calculamos el RMSE para las estimaciones del promedio curso, donde obtenemos el siguiente gráfico y tabla.

RMSE por curso en modelo sin memoria

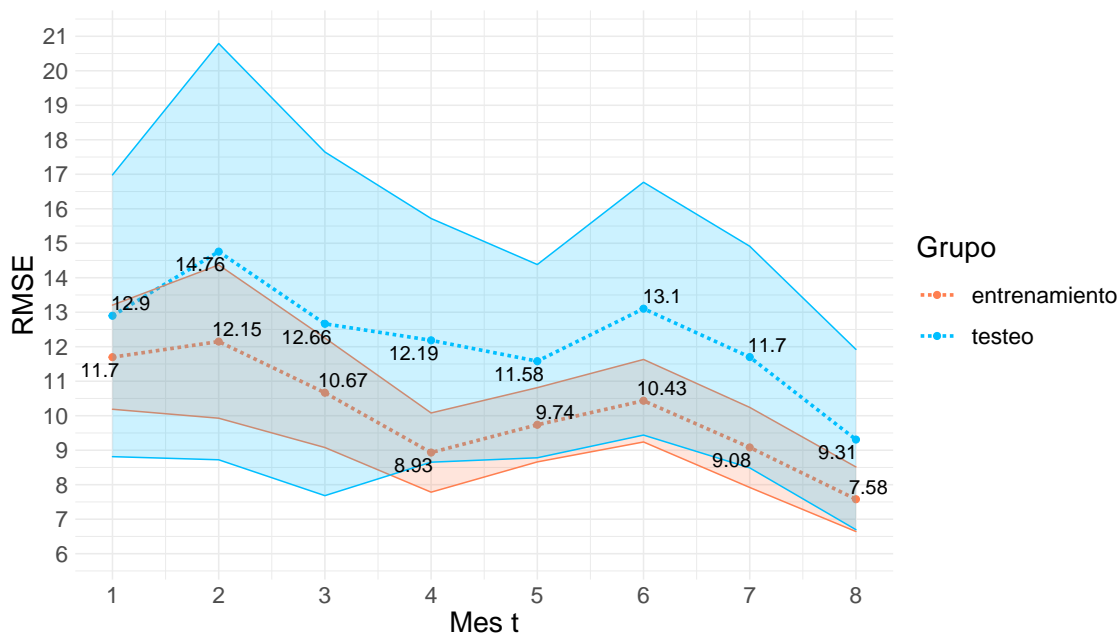


Figura 2.14: RMSE del promedio curso en el modelo sin memoria

Meses	1	2	3	4	5	6	7	8
RMSE promedio en entrenamiento	11.70	12.15	10.67	8.93	9.74	10.43	9.08	7.58
RMSE promedio en testeo	12.90	14.76	12.66	12.19	11.58	13.10	11.70	9.31
Desviación RMSE entrenamiento	1.51	2.22	1.59	1.15	1.08	1.20	1.16	0.94
Desviación RMSE testeo	4.08	6.03	4.98	3.53	2.80	3.67	3.21	2.61

Tabla 2.7: RMSE promedio del promedio curso y desviación de este RMSE para entrenamiento y testeo en el modelo sin memoria

Donde podemos notar que estamos bastante cerca del SEM para el promedio curso (a 2.05 puntos).

En cuanto a los valores estimados para cada parámetro del modelo, podemos observar el promedio de las estimaciones realizadas en las validaciones cruzadas. Entonces obtenemos los siguientes resultados para el efecto γ_t , que acompaña el puntaje en el pretest.

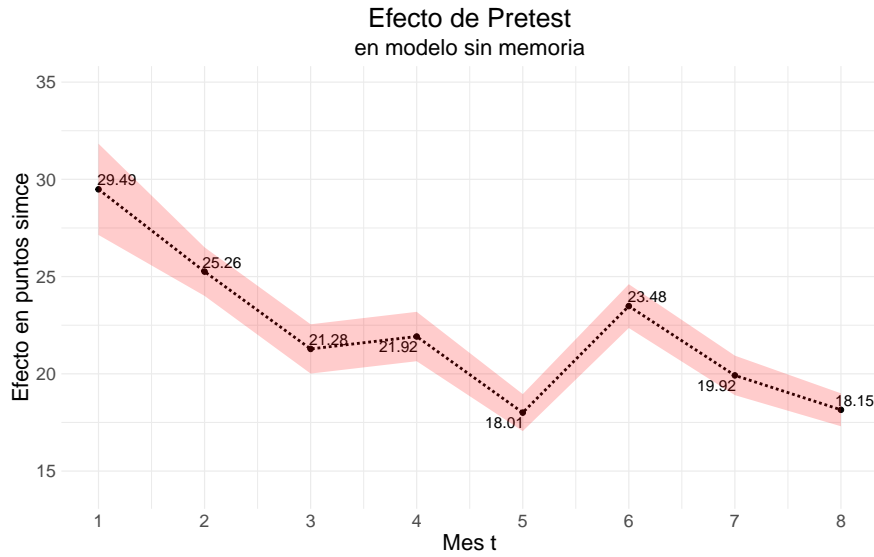


Figura 2.15: Efecto γ_t que acompaña la variable pretest, para distintos meses t de información.

Esto indica que en la medida que se tiene información más cercana a noviembre, el tamaño del efecto pretest se acerca cero, es decir, el modelo considera que el pretest tiene menor peso como regresor a la hora de estimar el puntaje. En una situación donde el desempeño de los estudiantes sea muy similar entre los últimos meses del año y el SIMCE, se esperaría que el pretest tenga un efecto no significativo, es decir, que no sea distinguible de cero.

Por otro lado, las variables históricas muestran un comportamiento diferente entre ellas. Primero el efecto asociado a la variable simce histórico, muestra lo siguiente

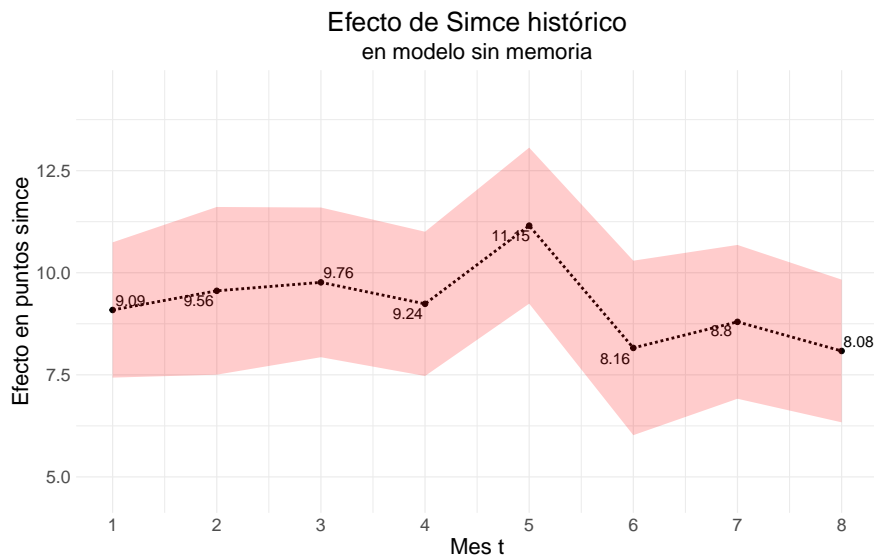


Figura 2.16: Efecto $\beta_{1,t}$ asociado a la variable simce histórico, para distintos meses t de información.

donde se aprecia que el valor estimado oscila entre los 8.1 y 11 puntos simce, pero sin

mostrar una tendencia al alza o a la baja. Esto podría indicar que esta variable tiende a tener un efecto independiente del mes t y su estimación tan solo se ve afectada por las fluctuaciones en las variables dinámicas. En cambio, el efecto que acompaña la variable “Me va bien en matemáticas” presenta un comportamiento similar al pretest, como se aprecia en la siguiente figura

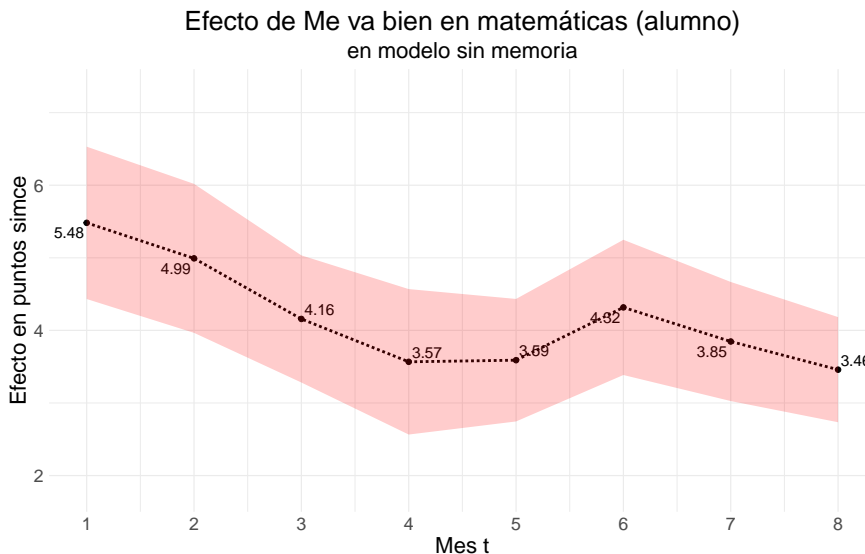


Figura 2.17: Efecto $\beta_{2,t}$ asociado a la variable “Me va bien en matemáticas”, para distintos meses t de información.

esto indicaría que en la medida que se tiene mayor información sobre el rendimiento de los alumnos mediante las componentes principales de las variables dinámicas, la percepción que tiene cada estudiante sobre su desempeño, va pasando a segundo plano.

Por otro lado, los efectos de las variables dinámicas para este modelo, los veremos para cada eje y para cada t , recordando que estos corresponderán a las primeras cinco componentes principales de un ACP realizado con las otras 10 variables dinámicas para dicho mes t .

Los efectos $\delta_{k,t}^{(1)}$, de las componentes principales (CP), son apreciables en la siguiente tabla.

Mes \ Efecto	1	2	3	4	5	6	7	8
Primera CP	1.0555	-0.0384	1.4972	5.4001	4.8967	2.4602	2.1803	3.6146
Segunda CP	-0.9259	0.0281	0.0583	-0.0493	0.1149	0.0486	-0.1498	0.3740
Tercera CP	-0.2373	-0.0168	-0.0795	-0.0239	-0.1131	0.0366	0.1125	0.0299
Cuarta CP	0.5081	-0.0527	0.1956	0.0614	0.0271	-0.0032	0.0481	-0.0310
Quinta CP	-0.0385	0.0493	0.0561	0.0195	0.1228	0.0158	0.0237	0.0743

Tabla 2.8: Efectos $\delta_{k,t}^{(1)}$ para los 5 ejes de contenidos dependiendo de la información en t .

En este caso, recordando que en cada mes solo se considera la información de ese mes, estos efectos nos indicarán cómo de manera aislada las primeras 5 componentes de la

información en ese mes son un buen regresor para el desempeño en el SIMCE (en $t = 9$). Primero es importante destacar que como sería de esperar, la primera componente es aquella que tiene un efecto que es de mayor mayor magnitud que las demás, y de igual manera las siguientes componentes tienen efectos de menor magnitud, aunque en algunos meses la cuarta o quinta componente tienen un peso mayor que la segunda y tercera, esto podría indicar que la variabilidad de la nube de datos en esos meses es tal que aun en las cuarta y quinta componentes, la varianza capturada por estos es en parte la que permite estimar de mejor manera el puntaje SIMCE.

Finalmente, más allá del signo de los efectos estimados o de la tendencia que tengan estos según los valores de t , ¿cómo se entiende el valor de cada efecto?

En este caso particular, como la variable dinámica que entregó el mejor desempeño son las primeras 5 componentes principales, para tener una comprensión detallada de lo que significa cada efecto, habría que analizar cómo están construidas estas componentes. Analizando las figuras 2.1a a la 2.1h, notamos que en general las principales fuentes de varianza en la nube de datos, en particular la primera componente, son el número de ejercicios resueltos.

Teniendo en cuenta esto, el efecto en cada componente en cada mes se interpreta como los puntos que el modelo estima tiene un alumno que esté a distancia uno del origen en la componente respectiva. Por ejemplo, en el mes 8, la primera componente tiene un efecto de 3.61. Esto indica que al realizar un análisis de componentes principales sobre las otras 10 variables dinámicas en $t = 8$, un alumno que en la primera componente esté a distancia uno del origen, se estima debería tener 3.61 puntos más que uno que esté en el origen de dicha componente, y como las variables usadas para crear esta componente fueron normalizadas, esto quiere decir que mientras más lejos del origen se esté en esta componente, será un alumno que se distingue del resto en varias dimensiones.

Luego, estos 3.61 puntos serán para un alumno que esté lejos de la media principalmente en cuanto al número de ejercicios que haya realizado, es decir, para la primera componente el modelo está asignando efectos positivos de puntaje para los alumnos que han tenido un alto nivel de ejercitación en términos del número de ejercicios que ha contestado.

2.5.2. Modelo glotón

Para el segundo modelo, las variables que entregaron mejores resultados fueron:

Pretest	Variables históricas	Variables dinámicas
Si	“Mi comportamiento es un problema para el profesor de matemáticas” (media curso)	Precisión de respuesta en el primer intento
	Nota a torneos espiral (respuesta por alumno)	Promedio de notas en preguntas abiertas

Tabla 2.9: Variables que entregaron un mejor desempeño para el modelo glotón

Es decir, si se define:

- $\Theta_{i,j}^{(1)} = com_{i,j}$, promedio curso de las respuestas sobre el nivel de acuerdo sobre “Mi comportamiento es un problema para el profesor de matemáticas”, de los alumnos del establecimiento j .
- $\Theta_{i,j}^{(2)} = nt_{i,j}$, promedio curso de la nota que los alumnos del establecimiento j pusieron a los torneos espiral.
- $u_{i,j,k,t}^{(1)} = a_{i,j,k,t}$, la precisión de respuesta del alumno i en el primer intento a ejercicios del eje k en el mes t .
- $u_{i,j,k,1:t}^{(2)} = r_{i,j,k,t}$:, calificación promedio del alumno i en preguntas abiertas en el eje k en el mes t .

el modelo será de la siguiente forma

$$\hat{y}_{i,j,t} = \gamma_t p_i + \beta_0 + \beta_{1,t} com_{i,j} + \beta_{2,t} mb_{i,j} + \sum_{s=1}^t \sum_{k=1}^5 \delta_{k,t}^{(1)} a_{i,j,k,s} + \sum_{s=1}^t \sum_{k=1}^5 \delta_{k,t}^{(2)} r_{i,j,k,s} \quad (2.9)$$

Esto muestra nuevas variables históricas como las óptimas para esta estructura de modelo. Pues ahora la estimación se realiza considerando la percepción promedio que tiene el curso sobre cómo la conducta de ellos como alumnos afecta al profesor de matemáticas, cómo los estudiantes evalúan los torneos realizados como parte del programa, y como variable dinámica, qué porcentaje de ejercicios el alumno es capaz de resolver sin cometer errores y la calificación de los alumnos en preguntas abiertas.

El desempeño de este modelo con las variables ya mencionadas se puede apreciar en el siguiente gráfico.

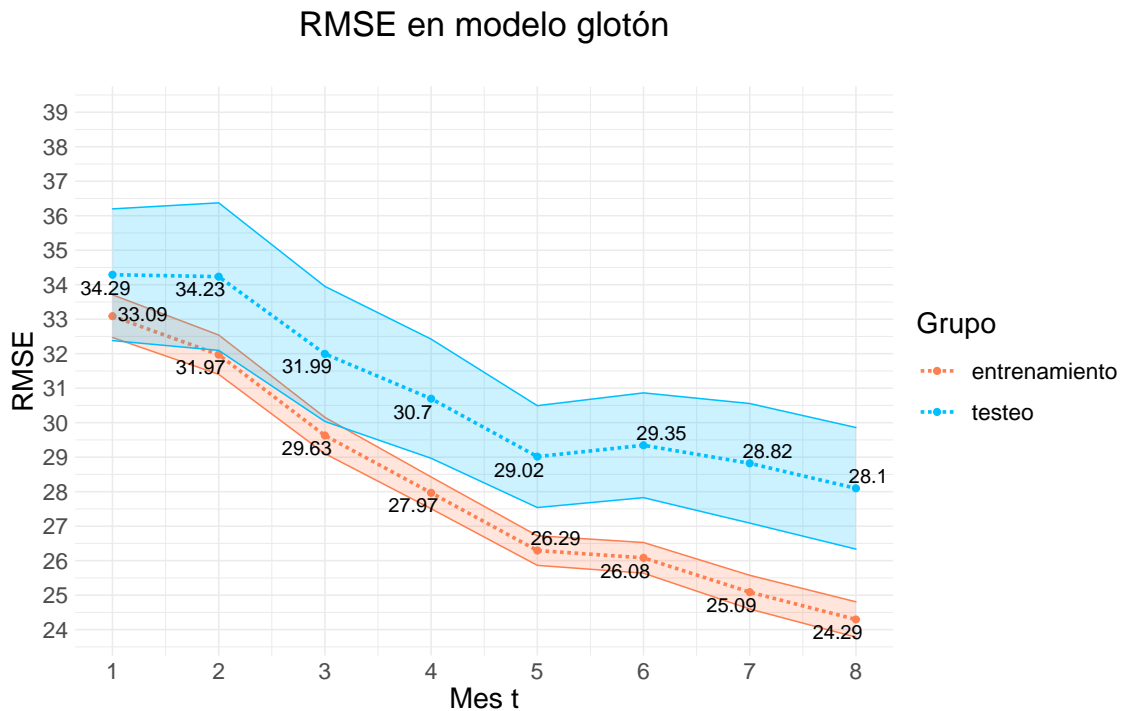


Figura 2.18: RMSE en el modelo glotón

donde las bandas de error corresponden a una desviación estándar de los RMSE obtenidos en las 100 validaciones cruzadas. Esto se puede notar en mayor detalle observando los datos de la siguiente tabla.

Meses	1	2	3	4	5	6	7	8
RMSE promedio en entrenamiento	33.09	31.97	29.63	27.97	26.29	26.08	25.09	24.29
RMSE promedio en testeo	34.29	34.23	31.99	30.70	29.02	29.35	28.82	28.10
Desviación RMSE entrenamiento	0.62	0.57	0.53	0.46	0.43	0.44	0.49	0.51
Desviación RMSE testeo	1.91	2.14	1.96	1.73	1.48	1.52	1.73	1.76

Tabla 2.10: RMSE promedio y desviación del RMSE para entrenamiento y testeo en el modelo glotón

En este caso se nota que el error en entrenamiento es increíblemente bajo comparado con el caso anterior, llegando a 24.29 puntos, esto es 2.98 puntos más bajo; estando aproximadamente a 7.7 puntos del error de medición del SIMCE para $t = 8$; sin embargo, así como el error en el grupo de entrenamiento baja considerablemente en la medida que aumenta los meses de información disponible para el modelo, esto no es lo que ocurre en el grupo de testeo, que si bien el error baja, no pasa la barrera de los 28 puntos; lo cual indica que esta forma de modelos son propensos a sobreajustes. Al observar la diferencia entre el RMSE en entrenamiento y testeo, vemos que esta aumenta con t , lo que refuerza lo ya mencionado.

Meses	1	2	3	4	5	6	7	8
Diferencia	1.20	2.26	2.36	2.73	2.73	3.27	3.73	3.81

Tabla 2.11: Diferencia entre el RMSE en entrenamiento y el RMSE en testeo para el modelo glotón

Esto no es un comportamiento deseable a la hora de construir un modelo que busca realizar predicciones en instancias futuras fuera de la muestra, pero los efectos que arroja, nos permitirán obtener una comprensión más profunda de las dinámicas entre ejes para la precisión de respuestas, puesto que al ser un modelo más ajustado a los datos de entrenamiento, tendrá un carácter más descriptivo en relación a los datos.

Al igual que para el modelo sin memoria, ahora observaremos la misma serie de casos de entrenamiento y testeo para observar si en esta ocasión se aprecia una mejor estimación del puntaje SIMCE en la medida que varía el t al cual está asociado el modelo.

Casos de entrenamiento

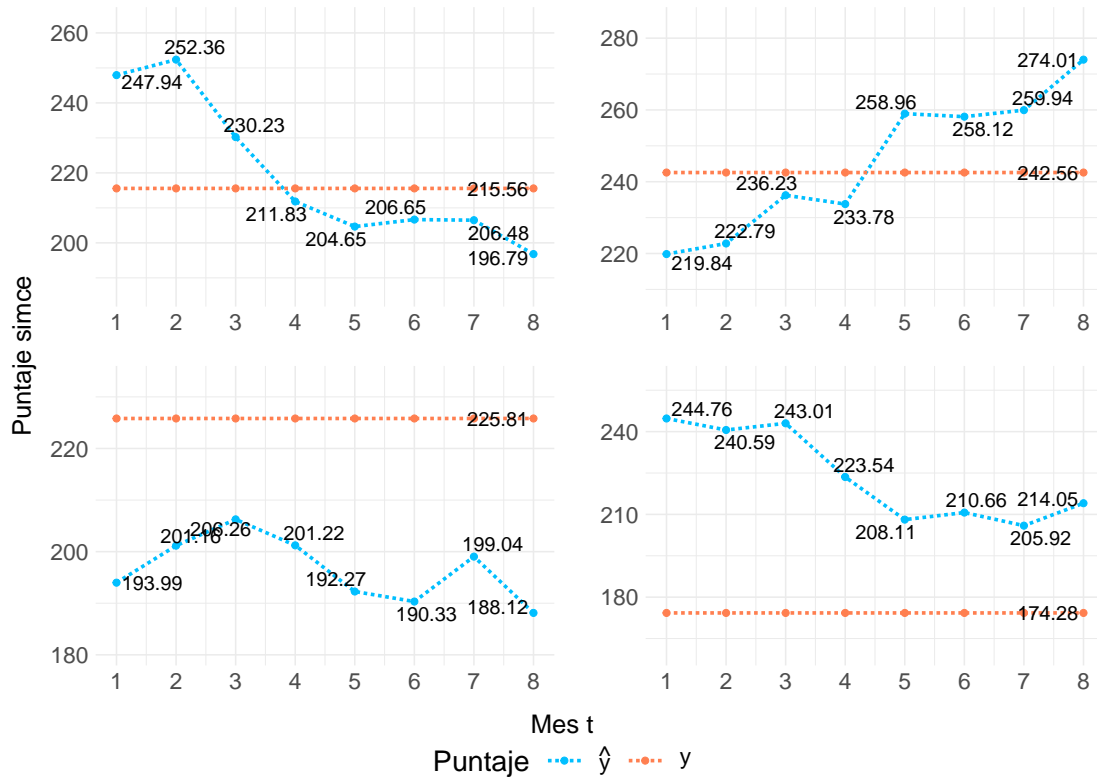


Figura 2.19: Cuatro casos en entrenamiento y sus distintas predicciones según el mes t de información usada

Tal y como es de esperar, en estos modelos la predicción va mejorando en la medida que aumenta el mes t de información disponible, esto porque a diferencia del modelo anterior en que solo se disponía del mes t para las variables dinámicas, en este caso avanzar en t significa agregar más variables al modelo. Así, aumenta el número de parámetros al aumentar el valor de t y por ende es esperable que mejore la predicción, pero como es el caso de todo modelo con más parámetros, se tiende a generar un sobreajuste, que es lo que ya observamos en la tabla 2.10.

Comprobemos esto observando los casos en testeo, que al igual que para entrenamiento, estaremos usando los mismos casos que para el modelo anterior.

Casos de testeo

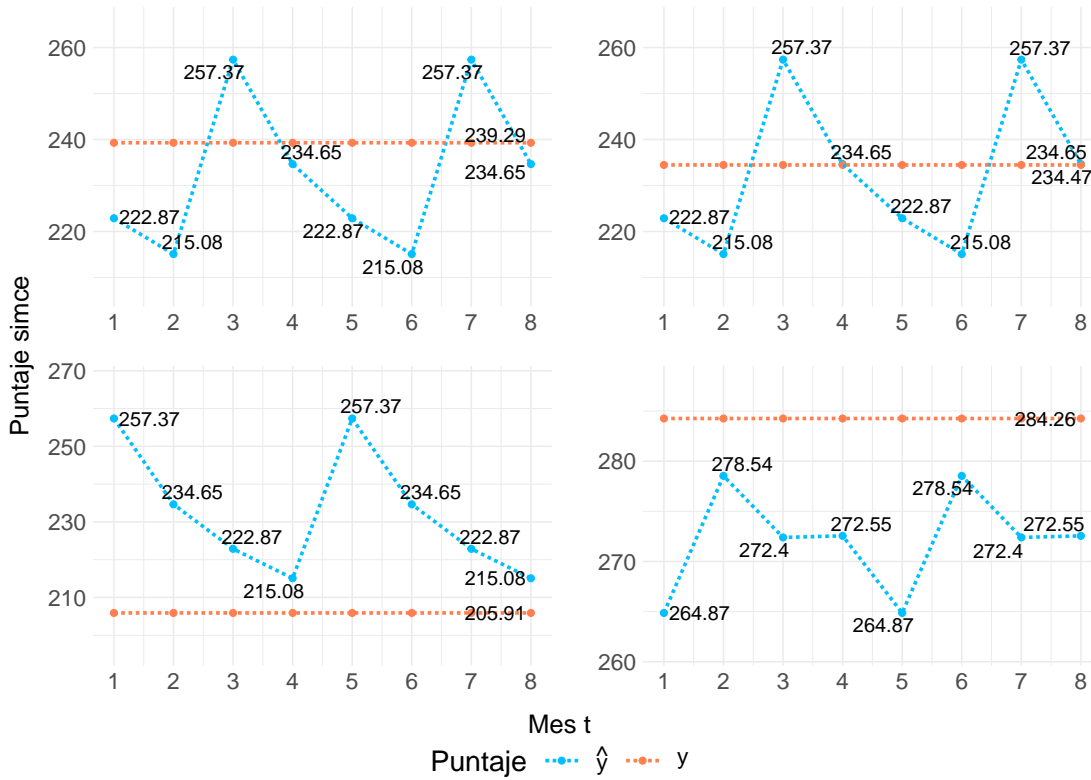


Figura 2.20: Cuatro casos en testeo y sus distintas predicciones según el mes t de información usada

Efectivamente, el error es de mayor magnitud en los casos de testeo, lo cual es particularmente evidente en el caso de la esquina inferior derecha, donde el puntaje predicho oscila más en torno al puntaje real en vez de ir acercándose a este valor de forma paulatina.

A nivel curso, el error de estimación es el siguiente

RMSE por curso en modelo glotón

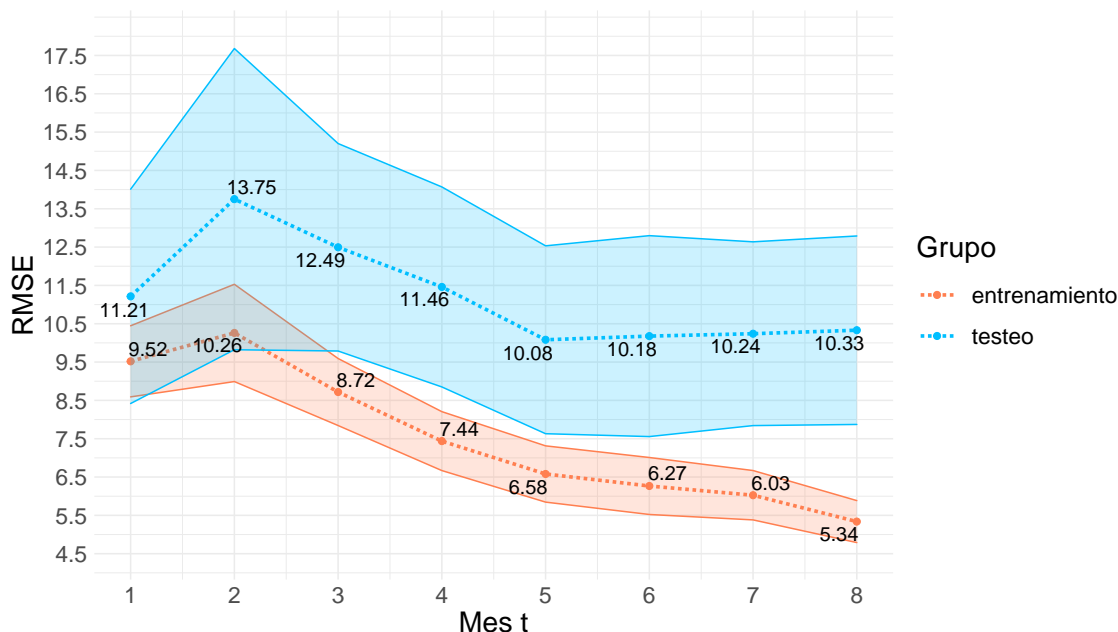


Figura 2.21: RMSE del promedio curso en el modelo glotón

Meses	1	2	3	4	5	6	7	8
RMSE promedio en entrenamiento	9.52	10.26	8.72	7.44	6.58	6.27	6.03	5.34
RMSE promedio en testeo	11.21	13.75	12.49	11.46	10.08	10.18	10.24	10.33
Desviación RMSE entrenamiento	0.93	1.27	0.87	0.77	0.73	0.74	0.64	0.55
Desviación RMSE testeo	2.80	3.93	2.71	2.61	2.45	2.62	2.40	2.46

Tabla 2.12: RMSE promedio del promedio curso y desviación de este RMSE para entrenamiento y testeo en el modelo glotón

donde se mantiene el comportamiento observado en las estimaciones por alumno, esto es, que el RMSE en entrenamiento baja 4.18 entre el primer y último mes, mientras que el error en testeo baja solo 0.88 puntos en este mismo intervalo, y en este caso, el RMSE es mayor que el error en el modelo anterior en testeo.

Por su parte si observamos el efecto γ_t para este modelo, esperaríamos un decrecimiento que sea como mínimo más pronunciado que el modelo anterior, puesto que ahora se tiene más información en la medida que nos acercamos a noviembre con la información disponible, por lo que con esta se debería ser capaz de caracterizar mejor el desempeño de cada estudiante sin tener tanta dependencia de los resultados iniciales que capturó el pretest. Veamos si estas expectativa resulta ser cierta.

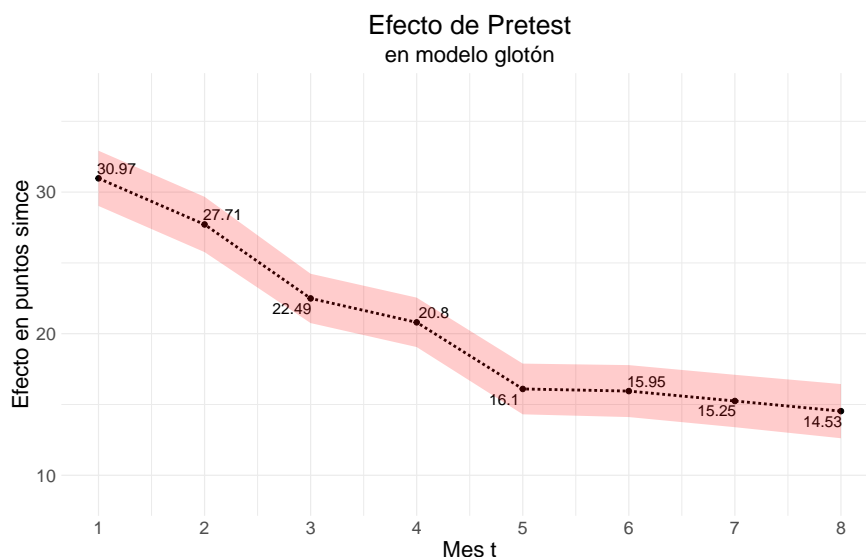


Figura 2.22: Efecto γ_t que acompaña la variable pretest, para distintos meses t hasta el cual se tiene información.

donde podemos comprobar que las suposiciones mencionadas se cumplen y efectivamente el efecto γ_t experimenta un decrecimiento más pronunciado y suavizado que antes. ¿ocurrirá lo mismo con las variables históricas o se observará un comportamiento diferente? El siguiente gráfico responde a esa pregunta.

Respecto al efecto de la variable “Mi problema es un comportamiento para el profesor de matemática”, obtenemos los siguientes resultados

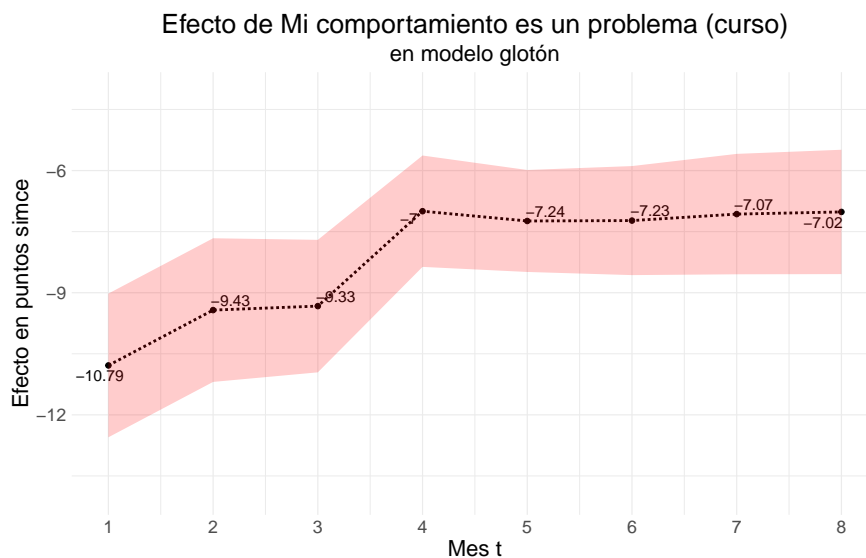


Figura 2.23: Efecto $\beta_{1,t}$ asociado a la variable “Mi problema es un comportamiento para el profesor de matemática”, para distintos meses t de información.

donde podemos notar que en la medida que el curso en promedio tiene una percepción

que su comportamiento es problemático para el profesor, esto tendrá un efecto negativo en el desempeño, o al menos se estima que así será. Esto se puede interpretar como un indicador del “clima negativo” del curso y mientras más acentuado se vea, sería de esperar un peor desempeño. En este caso particular se aprecia claramente una tendencia a estabilizar el peso de este efecto en torno a los -7 puntos simce, pues a partir de $t = 4$ no se aprecian mayores variaciones en el valor estimado.

Por su parte, el efecto de la evaluación que los niños daban de los torneos tiene un comportamiento más parecido al que observamos anteriormente con el simce histórico, con un valor que oscila, en este caso entre 2.2 y 4.6 puntos simce como se aprecia a continuación:

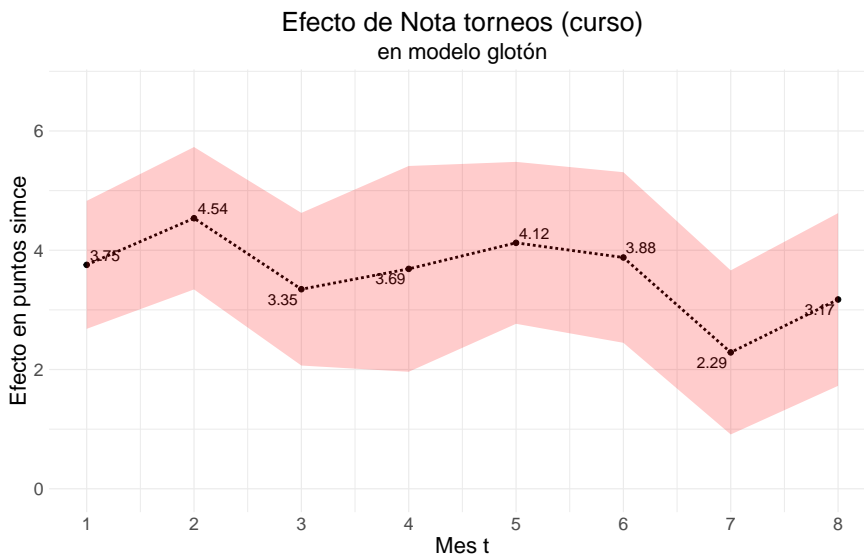


Figura 2.24: Efecto $\beta_{2,t}$ asociado a la variable nota de torneos espiral, para distintos meses t de información.

Entonces la apreciación sobre este efecto será similar a lo ya mencionado, esto es, que esta variable solo experimenta variaciones relativas producto de perturbaciones inducidas por la variable dinámica.

Estos gráficos refuerzan nuestras observaciones anteriores, sobre variables que claramente pasan a tener menor protagonismo como predictores en la medida que se obtiene mayor información con las variables dinámicas, pero a su vez existe variables que están capturando efectos sobre el desempeño de los alumnos que escapan de los alcances medibles por las variables dinámicas consideradas y por lo tanto aunque se tenga mucha información de la ejercitación de cada alumno, el peso de estos efectos seguirá siendo distinto de cero.

Veamos los efectos de las variables dinámicas, que para este modelo nos centraremos solo en el caso $t = 8$ (recordando que solo este caso ya tiene 80 parámetros de variables dinámicas), pues en ese caso podremos observar de mejor manera como el modelo se ajusta a los datos y nos ayuda a comprender las dinámicas entre los ejes. Comenzando por los efectos $\delta_{k,t}^{(1)}$, de la precisión de respuesta en cada eje.

Mes Efecto	1	2	3	4	5	6	7	8
Núm. y op.	4.4675	-1.0301	3.2987	2.2838	3.9401	-0.1387	2.8662	2.9043
Pat. y álg.	0.9474	-0.1792	-0.7465	0.0845	6.5891	-2.1536	2.3119	3.3656
Geometría	-2.0663	-0.0645	-0.6051	3.4812	-0.4389	0.9224	0.8208	2.6147
Medición	-0.6869	-3.0152	1.7863	2.2524	0.3735	-0.7999	0.4052	1.5918
Datos y Prob.	0.5565	0.6844	1.8359	-1.0961	2.9242	-0.8593	3.0295	4.0189

Tabla 2.13: Efectos $\delta_{k,t}^{(1)}$ para los 5 ejes de contenidos dependiendo de la información en t .

Uno de los primeros patrones que salta a la vista, es que los efectos en los últimos dos meses son todos positivos, esto nos indica que este modelo en particular entrega una gran importancia y un efecto positivo a la precisión en los 2 meses previos al SIMCE, mientras que por contrario, la precisión en los primeros meses tiene un efecto negativo, lo cual se puede explicar por la baja precisión en los primeros meses, por lo que el efecto de estar una desviación por sobre la media en esos meses, palidece en comparación con una alta precisión en los últimos meses.

Esto tiene sentido considerando el siguiente ejemplo de 4 alumnos con distintos niveles de desempeño en el eje de geometría ($k = 3$) en los meses 1 y 8. Recordando que al estar las variables normalizadas, un alumno que esté una desviación por sobre la media en $a_{i,j,k,t}$, será tal que $a_{i,j,k,t} = 1$, si su desempeño es en la media, $a_{i,j,k,t} = 0$ y bajo la media ya son valores negativos $a_{i,j,k,t} < 0$.

1. Alumno con rendimiento una desviación sobre la media en $t = 1$ ($a_{i,j,3,1} = 1$) y rendimiento una desviación sobre la media en $t = 8$ ($a_{i,j,3,8} = 1$).

En este caso el efecto de estos dos meses en este eje será el siguiente

$$\underbrace{-2.07}_{\delta_{3,1}^{(1)}} \cdot 1 + \underbrace{2.61}_{\delta_{3,8}^{(1)}} \cdot 1 = -2.07 + 2.61 = 0.54$$

Luego, un alumno que se mantiene una desviación por sobre la media entre estos dos meses, el modelo estima que tiene 0.54 puntos más.

2. Alumno con rendimiento una desviación sobre la media en $t = 1$ ($a_{i,j,3,1} = 1$) y rendimiento en la media (o bajo la media) en $t = 8$ ($a_{i,j,3,8} = 0$ ó $a_{i,j,3,8} < 0$, respectivamente).

En cambio, este caso, el efecto es

$$\underbrace{-2.07}_{\delta_{3,1}^{(1)}} \cdot 1 + \underbrace{2.61}_{\delta_{3,8}^{(1)}} \cdot 0 = -2.07$$

o bien, menor que -2.07 en caso que el rendimiento en $t = 8$ haya sido bajo la mediana. Entonces el efecto negativo en $t = 1$ lo que logra hacer es penalizar que el alumno haya bajado en sus posición relativa respecto de la media curso en su precisión de respuesta.

3. Alumno con rendimiento en la media (o bajo la media) en $t = 1$ ($a_{i,j,3,1} = 0$ ó $a_{i,j,3,1} < 0$, respectivamente). y rendimiento una desviación sobre la media en $t = 8$ ($a_{i,j,3,8} = 1$).

Ahora, en este caso el efecto es

$$\underbrace{-2.07}_{\delta_{3,1}^{(1)}} \cdot 0 + \underbrace{2.61}_{\delta_{3,8}^{(1)}} \cdot 1 = 2.61$$

y en caso que el rendimiento en $t = 1$ fuese bajo la media, el efectos será mayor a 2.61. Es decir, que el modelo entrega al menos 2.61 puntos más a un alumno que pasó de estar en la media o bajo esta en el mes inicial y pasó a estar una desviación por sobre esta en el último mes.

4. Alumno con rendimiento en la media (o bajo la media) en $t = 1$ ($a_{i,j,3,1} = 0$ ó $a_{i,j,3,1} < 0$, respectivamente) y rendimiento en la media (o bajo la media) en $t = 8$ ($a_{i,j,3,8} = 0$ ó $a_{i,j,3,8} < 0$, respectivamente).

Finalmente en este caso, el efecto es

$$\underbrace{-2.07}_{\delta_{3,1}^{(1)}} \cdot 0 + \underbrace{2.61}_{\delta_{3,8}^{(1)}} \cdot 0 = 0$$

En caso de mantenerse en la media, el modelo entrega 0 puntos, sin embargo, si su posición bajo la media en $t = 1$ es más baja que su posición en $t = 8$, es posible que aún así el modelo estime que este alumno tenga un efecto positivo.

En resumen, el signo en los efectos funciona de tal modo que el mayor efecto lo experimentan alumnos que mejoran su posición respecto de la media entre los primeros y los últimos meses, mientras que aquellos que mantienen un desempeño alto no reciben muchos puntos estimados de desempeño por esta variable y aquellos que bajan en desempeño reciben una penalización en el puntaje estimado.

Mientras, al mismo tiempo, los efectos del promedio en preguntas abierta de cada eje es el siguiente:

Mes Efecto	1	2	3	4	5	6	7	8
Núm. y op.	-	-	3.4011	-1.4155	-0.6102	-	-	0.7481
Pat. y álg.	-	-	-	-	-	-	-	-
Geometría	-	-	-	-	-	-	-	-
Medición	-	-	-	-	-	-	-	-
Datos y Prob.	-	-	-	-	-	-	-	-

Tabla 2.14: Efectos $\delta_{k,t}^{(2)}$ para los 5 ejes de contenidos dependiendo de la información en t .

En este caso ocurre que, debido al bajo número de preguntas abiertas realizadas en la mayoría de los ejes, esas dimensiones no fueron incluidas en las regresiones, ya sea porque su poca variación entraba en colinealidad con las otras variables o los efectos de estas carecían de un efecto significativamente distinto de cero, por lo que fueron excluidos del modelo. De todos modos, es interesante notar que el eje que tiene suficiente varianza para la estimación de los efectos, es el primero, de números y operaciones. Esto tiene sentido pues es el eje más prevalente en la ejercitación durante todo el año, pero aún así existen meses (6 y 7) donde la ejercitación no es suficiente para la estimación del efecto.

2.5.3. Modelo con memoria

Para el tercer modelo, las variables que entregaron mejores resultados fueron:

Pretest	Variables históricas	Variables dinámicas
Si	Simce histórico - media curso	Precisión de respuesta en el primer intento
	“Mis padres dicen que soy buen estudiante” - media curso	Promedio de notas en preguntas abiertas

Tabla 2.15: Variables que entregaron un mejor desempeño para el modelo con memoria

Es decir, tenemos el mismo conjunto de variables que para el modelo reflm1, entonces tenemos que

- $\Theta_{i,j}^{(1)} = sh_{i,j}$, simce histórico del establecimiento j al cual pertenece el alumno i .
- $\Theta_{i,j}^{(2)} = pb_{i,j}$, promedio curso de las respuestas sobre el nivel de acuerdo sobre “Me va bien en matemáticas”, de los alumnos del establecimiento j .
- $u_{i,j,k,1:t}^{(1)} = a_{i,j,k,1:t}$, la precisión de respuesta del alumno i en el primer intento a ejercicios del eje k entre los meses 1 y t .
- $u_{i,j,k,1:t}^{(2)} = r_{i,j,k,1:t}$, calificación promedio del alumno i en preguntas abiertas en el eje k entre los meses 1 y t .

el modelo será de la siguiente forma

$$\hat{y}_{i,j,t} = \gamma_t p_i + \beta_0 + \beta_{1,t} sh_{i,j} + \beta_{2,t} pb_{i,j} + \sum_{k=1}^5 \delta_{k,t}^{(1)} a_{i,j,k,1:t} + \sum_{k=1}^5 \delta_{k,t}^{(2)} r_{i,j,k,1:t} \quad (2.10)$$

cuyo error en entrenamiento y testeo se puede ver en el siguiente gráfico

RMSE en modelo con memoria

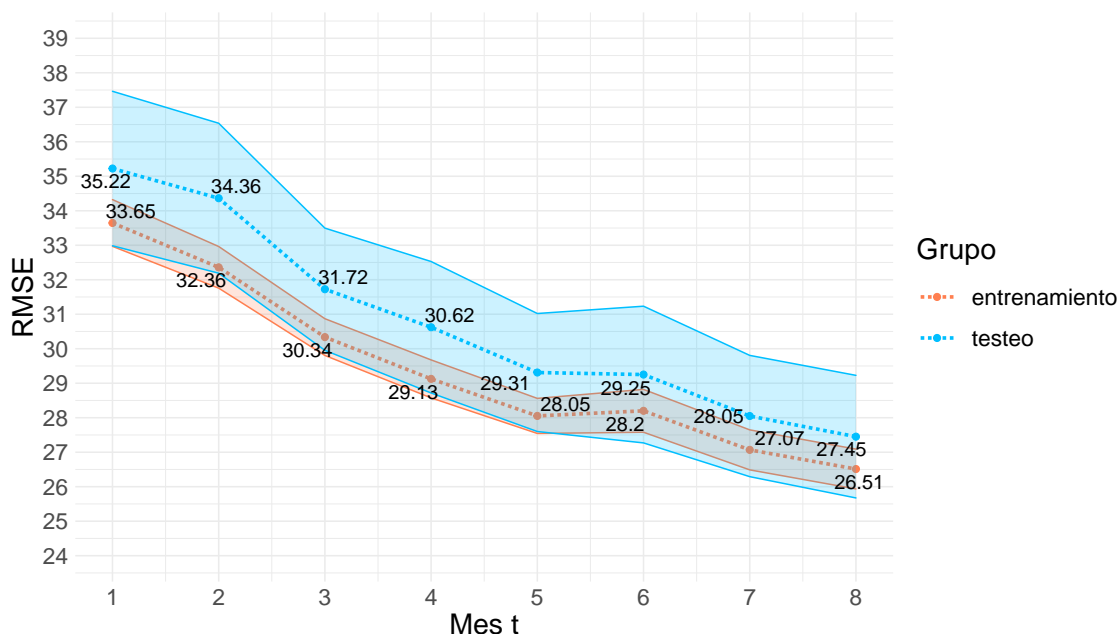


Figura 2.25: RMSE en el modelo con memoria

donde las bandas de error corresponden a una desviación estándar de los RMSE obtenidos en las 100 validaciones cruzadas. Esto se puede notar en mayor detalle observando los datos de la siguiente tabla

Meses	1	2	3	4	5	6	7	8
RMSE promedio en entrenamiento	33.65	32.36	30.34	29.13	28.05	28.20	27.07	26.51
RMSE promedio en testeo	35.22	34.36	31.72	30.62	29.31	29.25	28.05	27.45
Desviación RMSE entrenamiento	0.68	0.60	0.53	0.55	0.51	0.62	0.58	0.58
Desviación RMSE testeo	2.24	2.17	1.77	1.91	1.71	1.98	1.76	1.78

Tabla 2.16: RMSE promedio y desviación del RMSE para entrenamiento y testeo en el modelo con memoria

Hasta ahora, este modelo presenta los mejores resultados en términos de error de estimación, pues logra un RMSE bajo los 28 puntos en testeo en $t = 8$. Lo cual este modelo logra con 105 parámetros entre los 8 modelos para cada t . Al mismo tiempo, la diferencia entre entrenamiento y testeo no alcanza a ser más de un punto en el último mes (0.94 puntos), lo cual es una buena señal de la calidad de ajuste de este modelo.

Casos de entrenamiento

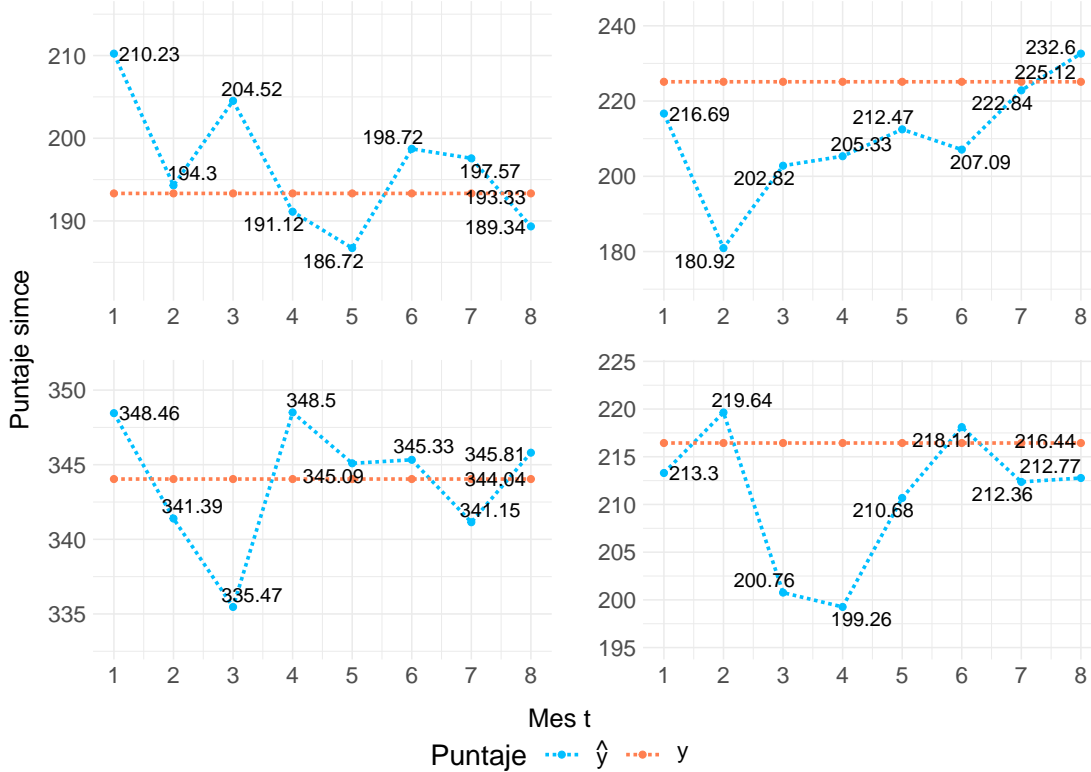


Figura 2.26: Cuatro casos en entrenamiento y sus distintas predicciones según el mes t de información usada

Al observar algunos de los casos en el grupo de entrenamiento en la figura 2.26, se sigue apreciando la misma tendencia que en los casos anteriores, pero de manera más consistente. Esto concuerda con el hecho que este modelo obtiene mejores predicciones que los anteriores. Si bien el modelo glotón tiene un error más bajo en entrenamiento, en testeo no es así, lo cual habla de la robustez de este modelo, y es esto lo que se aprecia en estos casos y otros tantos en cada iteración de entrenamiento y testeo durante la validación cruzada.

Es así como a diferencia de los casos anteriores, en los casos en testeo también se comienza a notar una tendencia a estimar el puntaje correcto, lo cual no era tan visible en los modelos ya analizados.

Casos de testeo

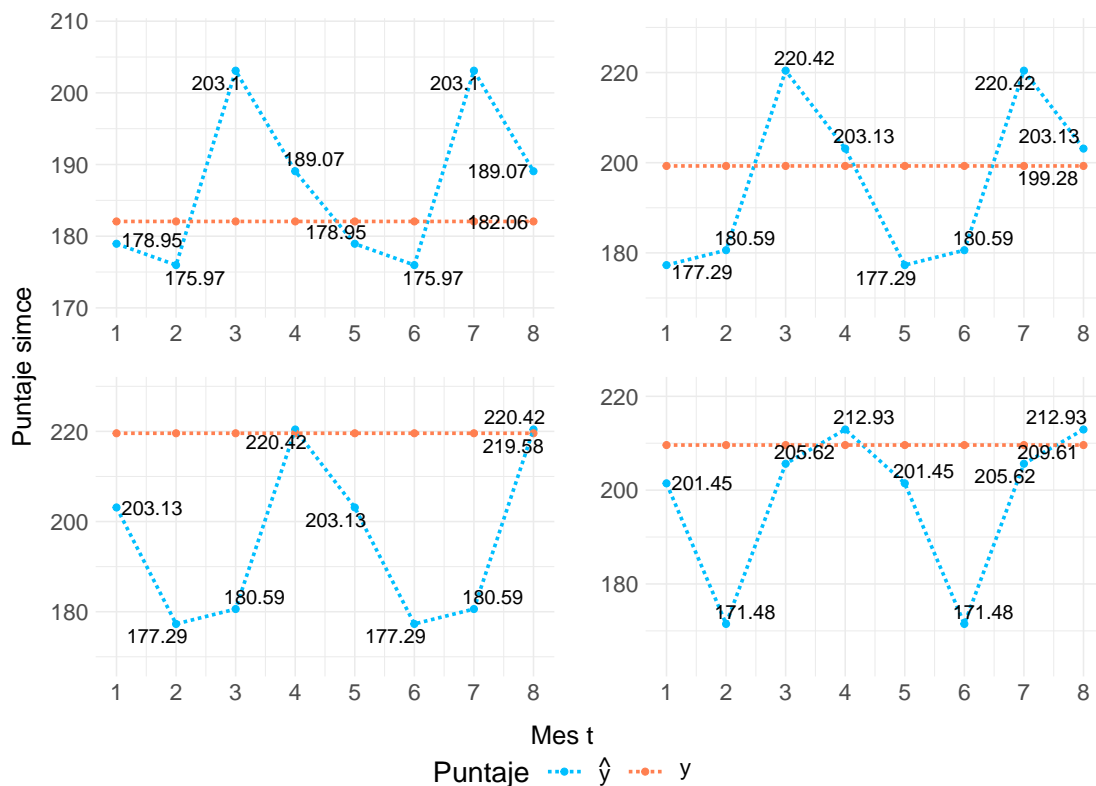


Figura 2.27: Cuatro casos en testeo y sus distintas predicciones según el mes t de información usada

Por su parte, el RMSE del promedio curso, muestra resultados no tan prometedores.

RMSE por curso en modelo con memoria

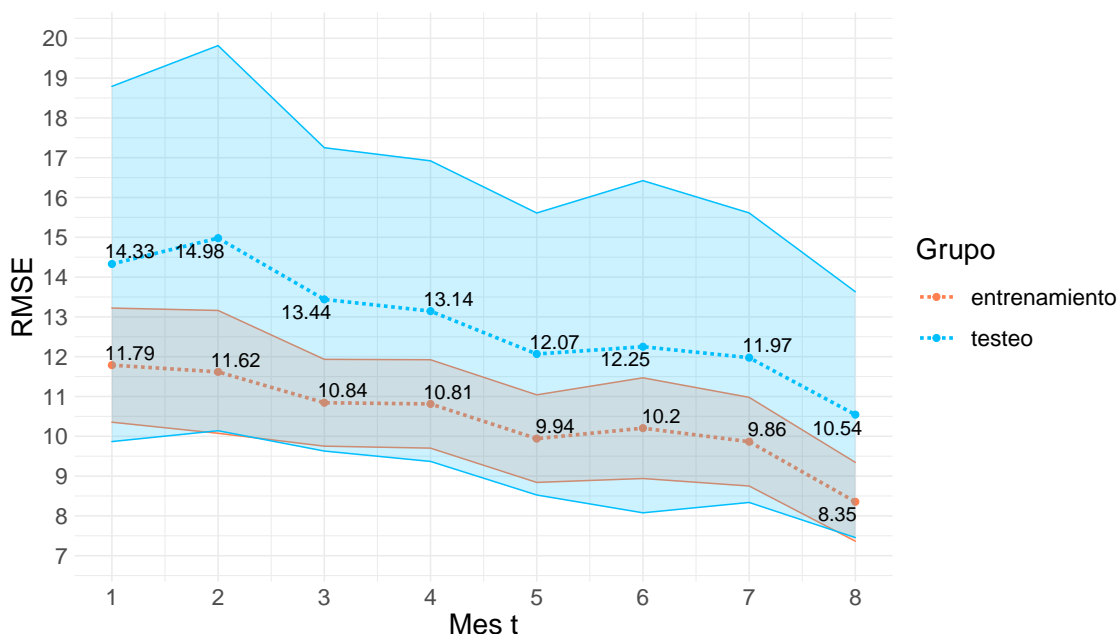


Figura 2.28: RMSE del promedio curso en el modelo con memoria

Meses	1	2	3	4	5	6	7	8
RMSE promedio en entrenamiento	11.79	11.62	10.84	10.81	9.94	10.20	9.86	8.35
RMSE promedio en testeo	14.33	14.98	13.44	13.14	12.07	12.25	11.97	10.54
Desviación RMSE entrenamiento	1.43	1.54	1.09	1.11	1.10	1.27	1.11	0.99
Desviación RMSE testeo	4.46	4.84	3.81	3.78	3.54	4.17	3.64	3.09

Tabla 2.17: RMSE promedio del promedio curso y desviación de este RMSE para entrenamiento y testeo en el modelo con memoria

En este caso en particular, notamos que el error para la estimación del puntaje promedio en testeo es mayor que para los dos modelos anteriores, y la desviación del RMSE también es mayor. Esto nos indica que este modelo logra mejores resultados a nivel de cada individuo, pero al agrupar estos resultados por curso, las estimaciones son peores que los modelos anteriores, es decir, las estimaciones realizadas con los efectos de cada parámetro, no se ajustan a nivel curso tan bien como los modelos sin memoria y glotón, siendo el RMSE 1.23 y 0.21 puntos más altos, respectivamente.

Notando esto, procedemos a observar el comportamiento de los parámetros del modelo, como lo es el efecto γ_t , el cual notamos en la figura 2.29 que se mantiene la tendencia a ser un valor decreciente en t que se ha observado antes.

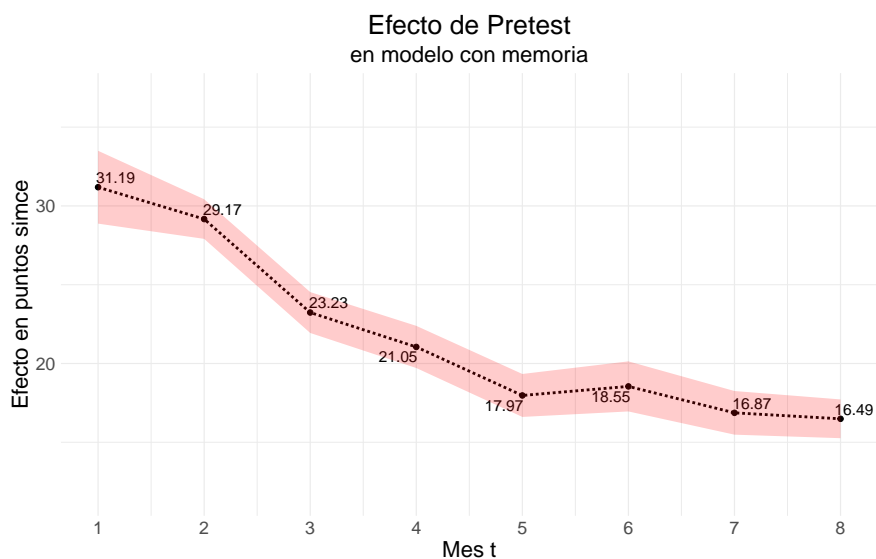


Figura 2.29: Efecto γ_t que acompaña la variable pretest, para distintos meses t hasta el cual se tiene información.

Sin necesidad de un mayor análisis, dado que la lectura es análoga a la realizada en previas instancias, es que continuamos con el análisis de las variables históricas, siendo la primera de estas el efecto asociado al puntaje simce histórico, del promedio simce del establecimiento en los años 2014 a 2016.

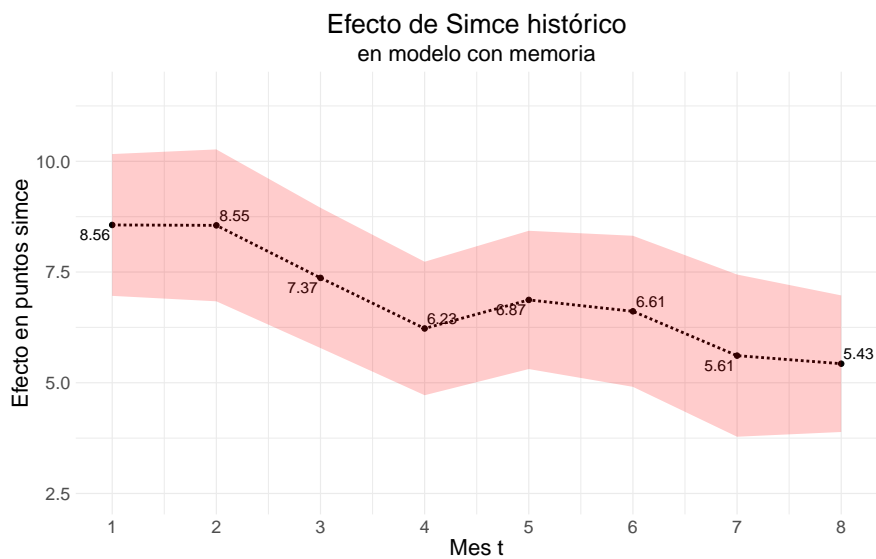


Figura 2.30: Efecto $\beta_{1,t}$ asociado a la variable simce histórico, para distintos meses t de información.

En esta ocasión, a diferencia de la figura 2.16, donde el efecto era prácticamente constante en t , en este caso se aprecia una curva decreciente, donde en los primeros meses, el efecto es de 8.5 puntos para un establecimiento con un simce histórico una desviación por sobre el resto, mientras que en $t = 8$ este efecto ha llegado a ser 5.4 puntos. Esto muestra que

al considerar la ejercitación de los alumnos de manera acumulativa, la posición relativa de un establecimiento respecto del resto como uno con mejores resultados SIMCE en los años anteriores, deja de explicar algunos puntos de desempeño al tener mayor información sobre la ejercitación realizada por los estudiantes.

En cambio, el efecto que acompaña la variable “Mis padres dicen que soy buen estudiante” presenta un comportamiento creciente, aunque también es posible especular que a partir de $t = 4$ se tiende a estabilizar o inclusive, a decrecer, como se aprecia en la siguiente figura

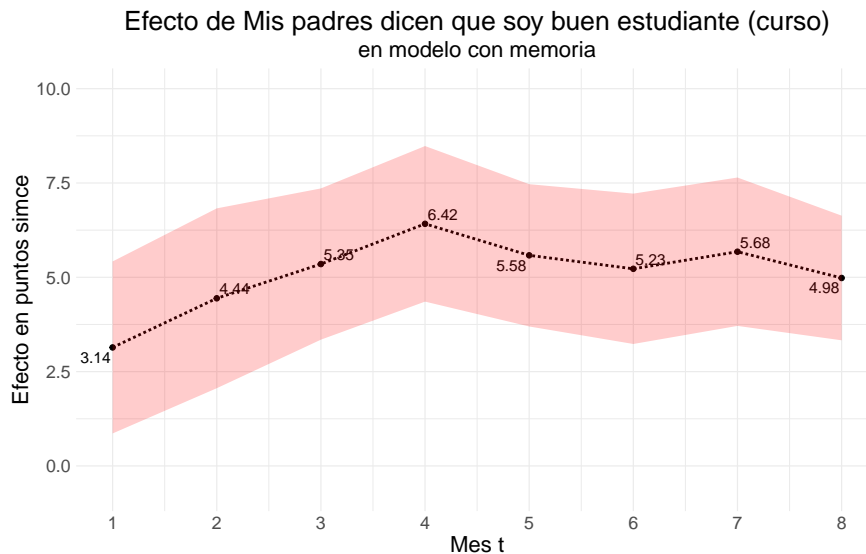


Figura 2.31: Efecto $\beta_{2,t}$ asociado a la variable “Mis padres dicen que soy buen estudiante”, para distintos meses t de información.

Esta variable, al corresponder a la media curso sobre esta afirmación, nos indica que en la medida que se tiene mayor información sobre el rendimiento que han mostrado los estudiantes en la plataforma ConectaIdeas, cobra más importancia la percepción promedio que tienen los padres de un curso sobre los alumnos y el hecho que hacen saber esto a los estudiantes. Es decir, cuanto más se sabe sobre la ejercitación de los alumnos, más importante es que la mayoría de los estudiantes tengan padres que les hagan saber que son buenos estudiantes.

El efecto de esta variable crece entonces debido a una eventual relación de dependencia con el rendimiento observado en las variables dinámicas, es decir que su capacidad como indicador del desempeño en el SIMCE, es condicionado a conocer cómo les ha ido a los alumnos en los ejercicios que han realizado hasta la fecha. Una posible explicación de esto, sería en relación con el efecto pigmalión y cómo este efecto solo se va acentuando en la medida que los alumnos muestran resultados que se alínean con la retroalimentación que reciben en sus hogares.

Es así como notamos que ahora las variables dinámicas tienen un efecto mucho mayor que en casos anteriores, esto es porque al trabajar con variables de 1 hasta un mes t , que un alumno esté una desviación por sobre la media, será una señal de un desempeño consistentemente por sobre el resto.

Efecto \ Mes	Mes							
	1	2	3	4	5	6	7	8
Núm. y op.	0.9314	9.2497	16.3451	15.4681	14.6998	12.8248	13.4239	12.1995
Pat. y álq.	0.0068	-0.5602	0.2703	2.6399	3.9970	3.5336	4.7123	4.5703
Geometría	1.9936	1.0017	-0.2236	2.6618	1.1790	0.3766	-0.2187	1.3031
Medición	0.2437	-2.5544	4.8095	4.3973	4.8527	7.7886	7.0319	5.8475
Datos y Prob.	1.9616	2.9095	0.4077	1.5725	5.4217	4.2104	6.2282	7.9620

Tabla 2.18: Efectos $\delta_{k,t}^{(1)}$ para los 5 ejes de contenidos dependiendo de la información en t .

Es así como en la tabla se nota que los efectos ahora son de varios puntos, tanto que si un alumno está una desviación por sobre la media en los cinco ejes al fines del mes de octubre, el modelo estima que ese alumno debería tener 31.88 puntos más que un alumno que haya tenido un desempeño promedio en los cinco ejes. Es decir, el equivalente a 0.67 veces la desviación del SIMCE en el 2017.

Efecto \ Mes	Mes							
	1	2	3	4	5	6	7	8
Núm. y op.	-	-	0.7811	2.6809	2.0487	1.1448	0.8421	1.0698
Pat. y álq.	-	-	-	-	-	-	-	-
Geometría	-	-	-	-	-	-	-	-
Medición	-	-	-	-	-	-	-	-
Datos y Prob.	-	-	-	-	-	-	-	-

Tabla 2.19: Efectos $\delta_{k,t}^{(2)}$ para los 5 ejes de contenidos dependiendo de la información en t .

Al igual que en el modelo glotón, como se pudo ver en la tabla 2.14, gran parte de los efectos no son estimados, pero a diferencia del caso anterior, los efectos del eje números y operaciones en los meses 6 y 7 se puede estimar, pues al trabajar con variables que contienen el acumulado hasta un mes t , a partir del mes 3 cuando ya se tiene una variable con suficiente significancia, aunque un mes no se realice ninguna pregunta abierta, esta variable seguirá siendo significativa pues contiene la información de los mese anteriores en los que ya lo era.

2.6. Modelo lineal con efecto olvido

Los tres modelos propuestos tienen cada uno sus ventajas y desventajas, siendo el que tiene un mejor desempeño el modelo con memoria, pues el error en testeo para $t = 8$ es el menor entre los 3. Aunque el modelo glotón tenga un error más bajo en entrenamiento, como se mencionó anteriormente, esto ocurre en desmedro de estimaciones fuera de la muestra, es decir, se produce sobre ajuste. En tal caso, ¿cómo podemos aprovechar la información disponible hasta un instante t como en el modelo glotón, pero sin tener sobre ajuste? Una respuesta a esta pregunta es reducir el número de parámetros, pues podemos hacer algo como el modelo con memoria, pues esto se observó ser eficiente en la tabla 2.16.

Observamos la tabla 2.13, donde en conjunto con las tendencias observadas en los gráficos 2.29 y 2.31, podemos inferir que existe cierto descuento que hace que los meses iniciales no tengan tanto peso en el modelo en la medida que la información disponibles es más cercana al horizonte de predicción. Por esto, planteamos la existencia de un efecto de olvido ρ en las variables dinámicas, por lo que la formulación de estas sería

$$\delta_t^{(m)} U_{i,j,t}^{(m)} = \begin{cases} \sum_{s=1}^t \rho^{t-s+1} \left(\sum_{k=1}^5 \delta_{k,t}^{(m)} u_{i,j,k,s}^{(m)} \right) & \text{si } u_{i,j,k,s}^{(m)} \in \{e_{i,j,k,s}, f_{i,j,k,s}, g_{i,j,k,s}, d_{i,j,k,s}, o_{i,j,k,s}, p_{i,j,k,s}\} \\ \frac{1}{t} \sum_{s=1}^t \rho^{t-s+1} \left(\sum_{k=1}^5 \delta_{k,t}^{(m)} u_{i,j,k,s}^{(m)} \right) & \text{si } u_{i,j,k,s}^{(m)} \in \{n_{i,j,s}, a_{i,j,k,s}, q_{i,j,k,s}, r_{i,j,k,s}\} \end{cases} \quad (2.11)$$

Esto básicamente hará que la incidencia de los primeros meses del año sea menor a la de los meses finales, es decir, un factor que crece en función de t , y por lo tanto, cumple con el propósito deseado. Así formulamos un cuarto modelo, pero que ya no será puramente lineal puesto que ρ le da un carácter polinomial a nuestra formula.

$$\hat{y}_{i,j,t} = \gamma_t p_i + \beta_0 + \beta_{1,t} \Theta_{i,j}^{(1)} + \beta_{2,t} \Theta_{i,j}^{(2)} + \sum_{s=1}^t \rho^{t-s+1} \left(\sum_{k=1}^5 \delta_{k,t}^{(1)} u_{i,j,k,s}^{(1)} + \sum_{k=1}^5 \delta_{k,t}^{(2)} u_{i,j,k,s}^{(2)} \right) \quad (\text{con olvido})$$

Esto permite hacer uso de toda la información hasta t , dando un distinto peso para cada mes, pero sin aumentar desmesuradamente el número de parámetros como en el modelo glotón, es más, al definir ρ como un parámetro que no depende de t para su estimación, al igual que el intercepto β_0 , será el mismo para todos los $\hat{y}_{i,j,t}$, $\forall t$, es decir, esta formulación tendrá 106 parámetros como máximo en total, uno más que el máximo de 105 en los modelos sin memoria y con memoria, y muchos menos que glotón.

2.6.1. Resultados en el Modelo con olvido

Para este modelo con efecto olvido, tenemos esta combinación de variables óptimas:

Pretest	Variables históricas	Variables dinámicas
Si	“Mi comportamiento es un problema para el profesor de matemáticas” (media curso)	Precisión de respuesta en el primer intento
	”Matemáticas es fácil para mí” (respuesta por alumno)	

Tabla 2.20: Variables que entregaron un mejor desempeño para el modelo con olvido

Es decir, si se define:

- $\Theta_{i,j}^{(1)} = com_{i,j}$, promedio curso de las respuestas sobre el nivel de acuerdo sobre “Mi comportamiento es un problema para el profesor de matemáticas”, de los alumnos del establecimiento j .

- $\Theta_{i,j}^{(2)} = m f_{i,j}$, promedio curso de las respuestas sobre el nivel de acuerdo sobre “Mate-máticas es fácil para mi”, de los alumnos del establecimiento j .
- $u_{i,j,k,t}^{(1)} = a_{i,j,k,t}$, la precisión de respuesta del alumno i en el primer intento a ejercicios del eje k en el mes t .
- $u_{i,j,k,t}^{(2)} = 0$

el modelo será de la siguiente forma

$$\hat{y}_{i,j,t} = \gamma_t p_i + \beta_0 + \beta_{1,t} com_{i,j} + \beta_{2,t} m f_{i,j} + \frac{1}{t} \sum_{s=1}^t \rho^{t-s+1} \left(\sum_{k=1}^5 \delta_{k,t}^{(1)} a_{i,j,k,s} \right) \quad (2.12)$$

Así, con esta combinación de variables y la incorporación de un efecto olvido, obtenemos los RMSE que se aprecian a continuación

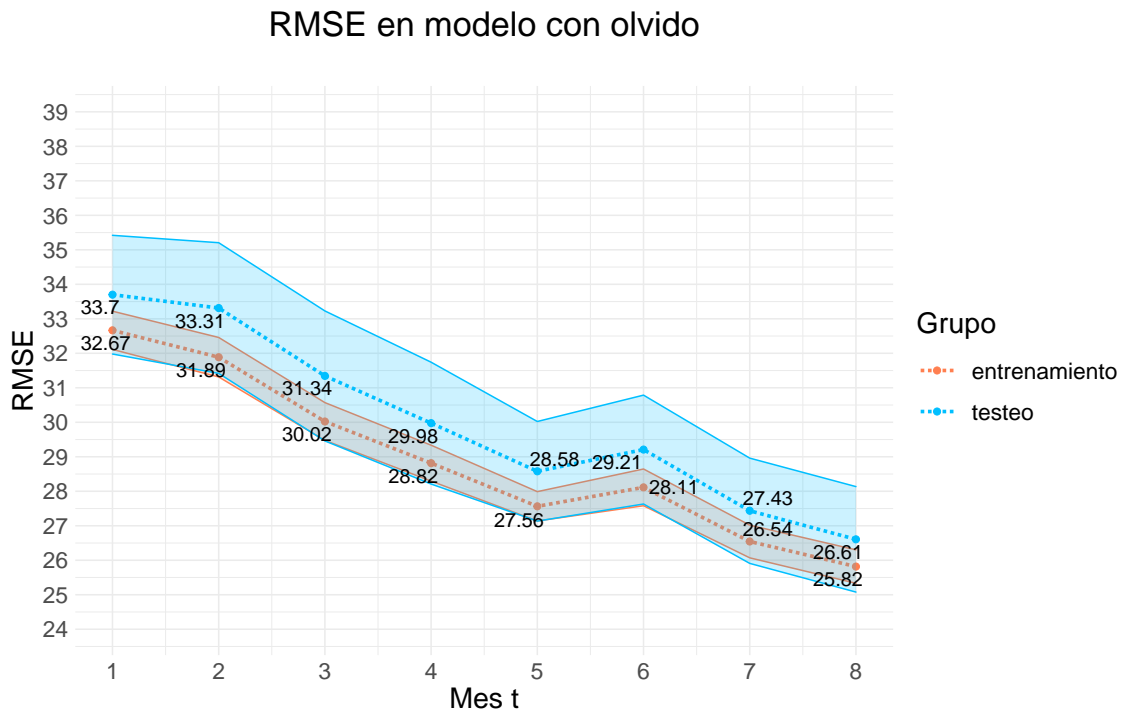


Figura 2.32: RMSE en el modelo con olvido

donde notamos que este modelo es mejor que los 3 anteriores, especialmente en los últimos dos meses donde el error en testeo está incluso por debajo de los 27 puntos, lo cual se puede apreciar en mayor detalle en la tabla.

Meses	1	2	3	4	5	6	7	8
RMSE promedio en entrenamiento	32.67	31.89	30.02	28.82	27.56	28.11	26.54	25.82
RMSE promedio en testeo	33.70	33.31	31.34	29.98	28.58	29.21	27.43	26.61
Desviación RMSE entrenamiento	0.55	0.57	0.55	0.52	0.43	0.53	0.47	0.49
Desviación RMSE testeo	1.72	1.89	1.88	1.77	1.44	1.58	1.53	1.53

Tabla 2.21: RMSE promedio y desviación del RMSE para entrenamiento y testeo en el modelo con olvido

Al mismo tiempo notamos que ya en $t = 5$ se obtiene un RMSE que en los otros modelos se obtenía en $t = 7$, lo cual hace de este modelo, uno que permite lograr una predicción tan buena como los anteriores, pero con menos información o dicho de otra forma, 2 meses antes.

Así como en los otros 3 modelos observamos el comportamiento en 4 casos elegidos al azar y notamos que ahora la tendencia es similar al modelo con memoria, lo cual es lo que se esperaba obtener dado las similitudes entre estos dos modelos.

Casos de entrenamiento

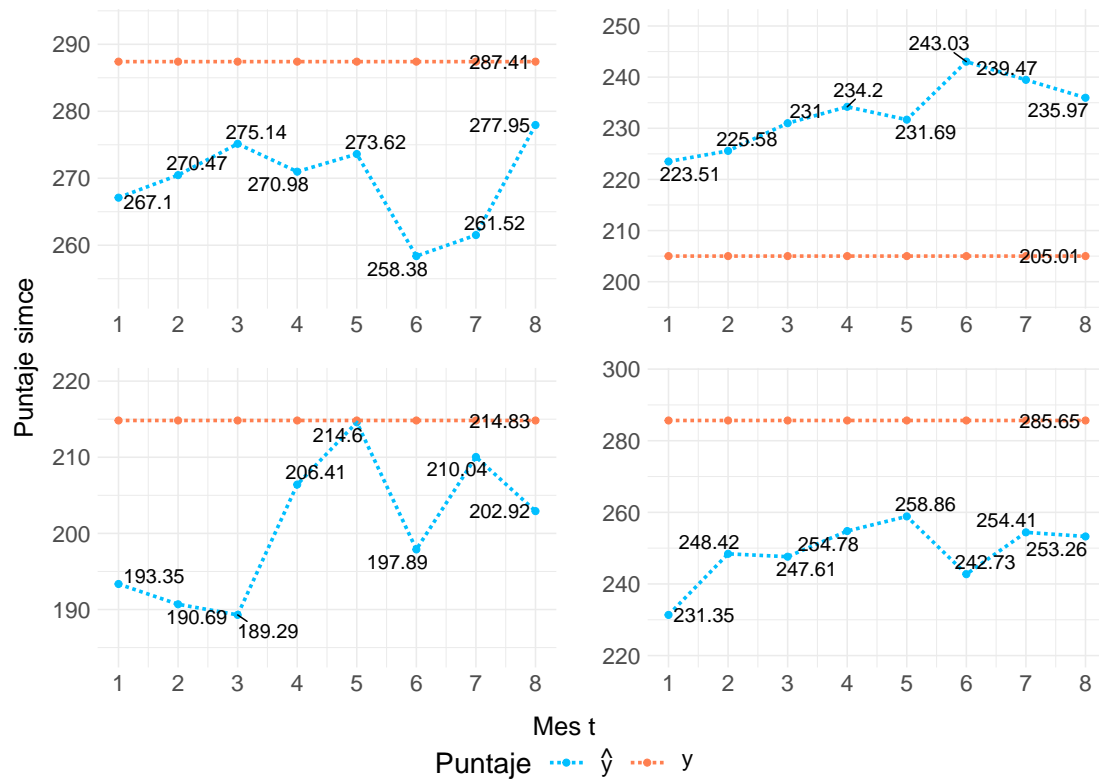


Figura 2.33: Cuatro casos en entrenamiento y sus distintas predicciones según el mes t de información usada

De igual modo en testeo, se aprecia una tendencia a obtener estimaciones más cercanas al puntaje real obtenido por cada alumno, pero uno de los problemas que sigue presente es el salto que ocurre en los meses 5 y 6.

Casos de testeo

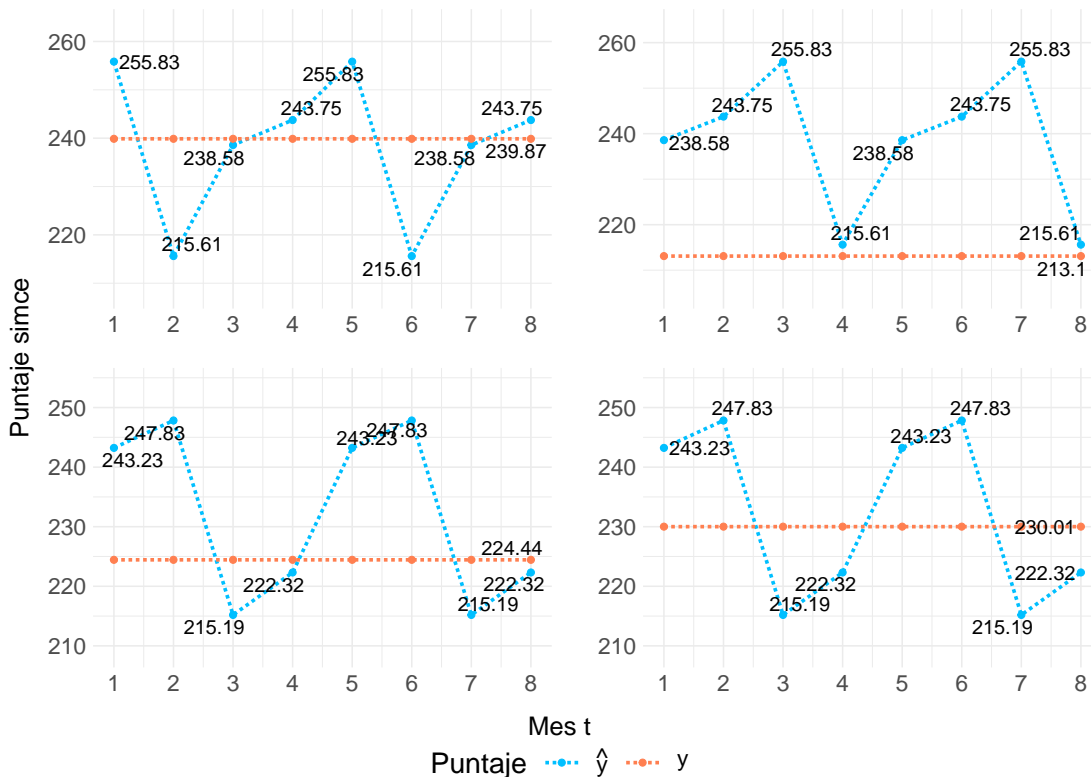


Figura 2.34: Cuatro casos en testeo y sus distintas predicciones según el mes t de información usada

Al igual que para los otros tres modelos, analizamos el error en la estimación del puntaje promedio de cada curso.

RMSE por curso en modelo con olvido

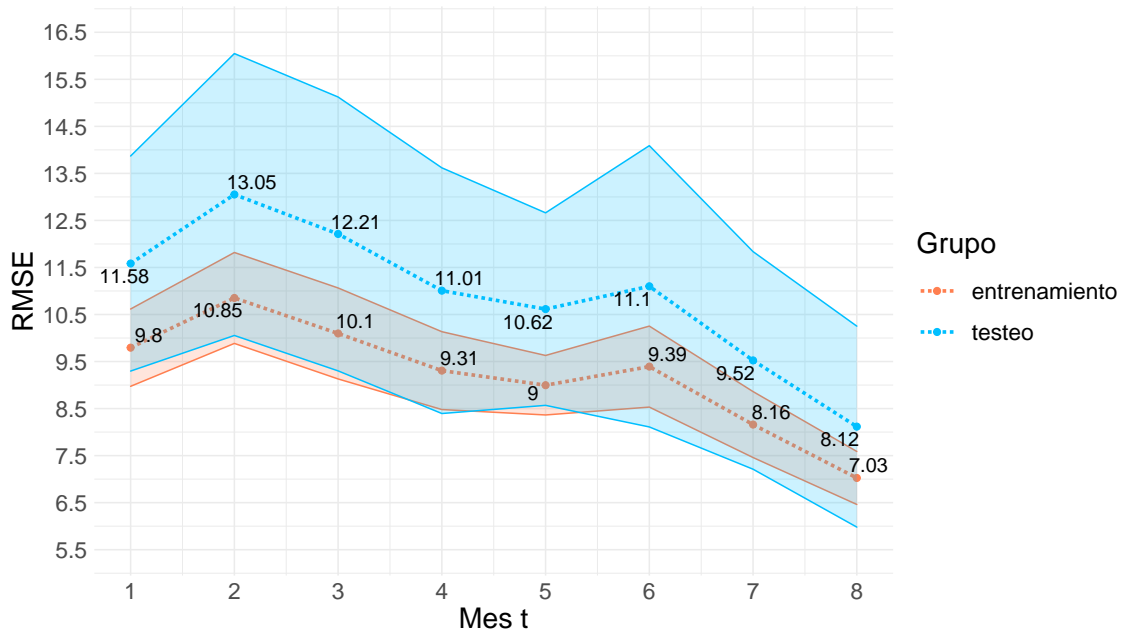


Figura 2.35: RMSE del promedio curso en el modelo con olvido

Meses	1	2	3	4	5	6	7	8
RMSE promedio en entrenamiento	9.80	10.85	10.10	9.31	9.00	9.39	8.16	7.03
RMSE promedio en testeo	11.58	13.05	12.21	11.01	10.62	11.10	9.52	8.12
Desviación RMSE entrenamiento	0.82	0.97	0.97	0.83	0.63	0.86	0.70	0.56
Desviación RMSE testeo	2.29	3.00	2.91	2.61	2.05	2.99	2.31	2.14

Tabla 2.22: RMSE promedio del promedio curso y desviación de este RMSE para entrenamiento y testeo en el modelo con olvido

Con esto, el modelo con efecto olvido (con olvido), es el que entrega los RMSE más bajos tanto para estimaciones del puntaje de un alumno, a 10.05 del error del SIMCE; así como el RMSE más bajo del curso en que este pertenece, estando a 0.86 puntos del SEM del SIMCE. Más aún, también la desviación del RMSE es más baja, una buena señal sobre la estabilidad de estos resultados.

Ahora, analicemos el comportamiento de los parámetros del modelo.

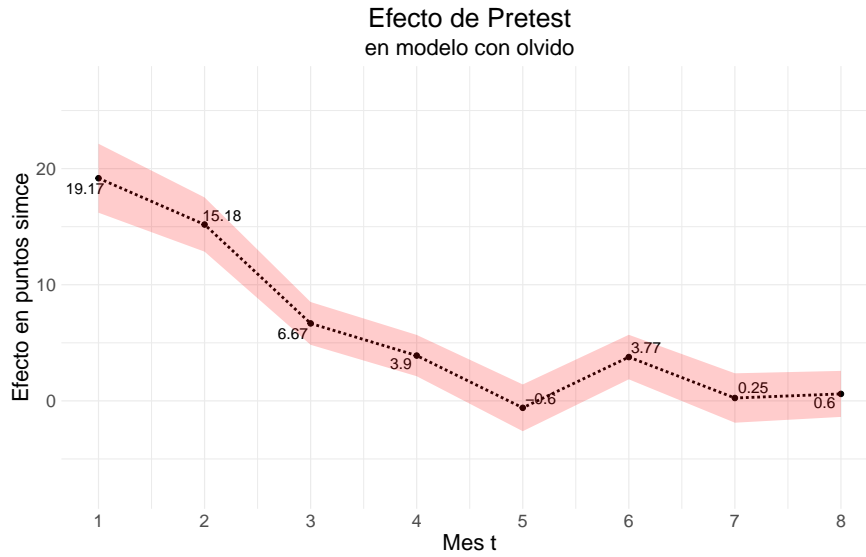


Figura 2.36: Efecto γ_t que acompaña la variable pretest, para distintos meses t hasta el cual se tiene información.

También observamos el efecto del pretest en el gráfico 2.36, donde la tendencia a ser decreciente en el tiempo es similar a los anteriores, pero con una sorpresa significativa, que se observa en el gráfico.

En este modelo, el efecto γ_1 es mucho menor que los casos anteriores y llegando a γ_8 el efecto es cercano a cero, es decir que en este modelo cuando se tiene la información del trabajo en el programa ConectaIdeas hasta justo antes del SIMCE, el efecto del desempeño en el pretest es prácticamente cero para la estimación.

Por su parte, el efecto que acompaña la variable se puede notar algo similar que en la figura 2.23

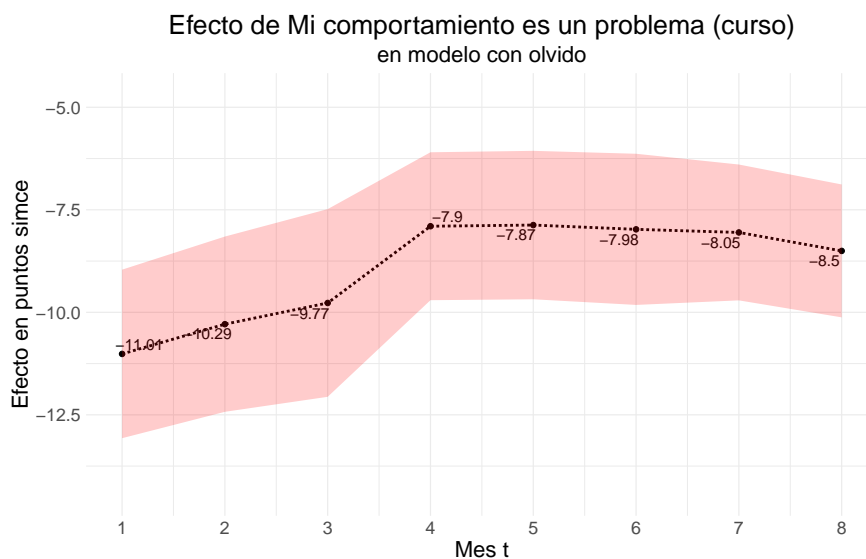


Figura 2.37: Efecto $\beta_{1,t}$ asociado a la variable “Mi problema es un comportamiento para el profesor de matemática”, para distintos meses t de información.

Aquí, el efecto es negativo y la magnitud de este decrece luego de los primeros meses para mantener un valor en torno a los -8 puntos. Entonces para este modelo, un curso cuyo promedio de acuerdo sobre la afirmación “Mi problema es un comportamiento para el profesor de matemática” esté una desviación por sobre el resto de los cursos, se estima tendrá 8 puntos menos. Esto permite inferir que un curso donde los alumnos sienten que su conducta afecta negativamente al profesor, se esperaría que tenga un puntaje más bajo que el resto. Analizando desde el punto de vista contrario, en un curso donde estos alumnos son minoría y por ende el acuerdo promedio con esta afirmación está bajo la media, este curso se esperará que tenga mejor desempeño.

Luego, analizando el efecto de la variable sobre la percepción de cada alumno si matemáticas es fácil o no para ellos, vemos que efectivamente tiene un efecto positivo, es decir, se espera que aquellos alumnos que dicen tener facilidad para matemáticas, tengan unos puntos más que el resto.

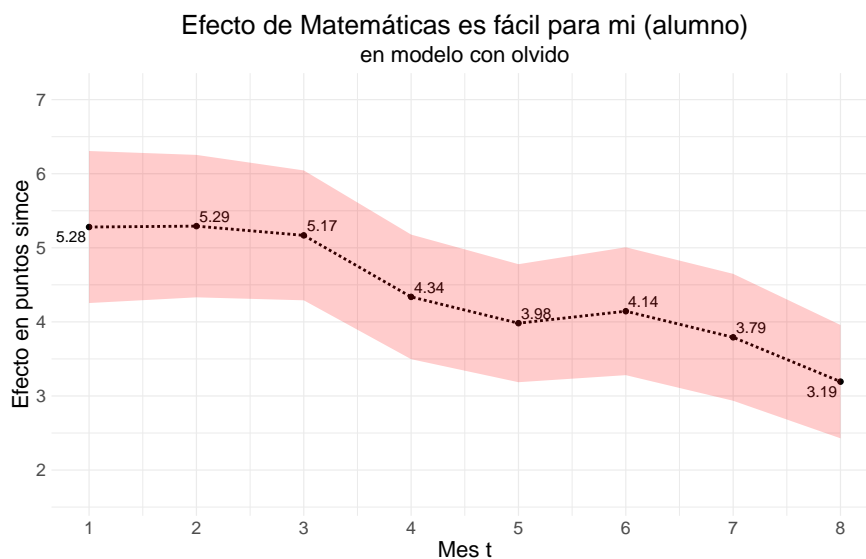


Figura 2.38: Efecto $\beta_{2,t}$ asociado a la variable “Matemáticas es fácil para mi”, para distintos meses t de información.

En el gráfico se observa como el efecto decae en t de manera similar a lo observado para el modelo sin memoria con la variable “Me va bien en matemáticas” en el gráfico 2.16. Esto nos permite hacer un paralelo con este modelo y al mismo tiempo contrastar con el caso en el modelo con memoria donde el efecto es creciente para la misma variable del modelo sin memoria.

Este modelo comparte similitudes con el modelo sin memoria en el sentido que este modelo, con el uso de un efecto ρ de olvido, entrega mayor importancia a la información del último mes t de información disponible, mientras el modelo con memoria da la misma importancia a cada mes de información entre 1 y t .

Entonces, en la medida que el modelo considera en alguna medida un efecto de pérdida de memoria, ya sea simplemente manteniendo solo el último mes de información o con el uso de un efecto de descuento para los meses anteriores, ese efecto de decrecimiento se verá reflejado en los parámetros que acompañan las variables “Me va bien en matemáticas” y “Matemáticas es fácil para mi”.

Este efecto de olvido mencionado, se puede apreciar en el siguiente gráfico, donde no habrá una variación en t puesto que este parámetro, al igual que el intercepto β_0 son estimados de manera transversal.

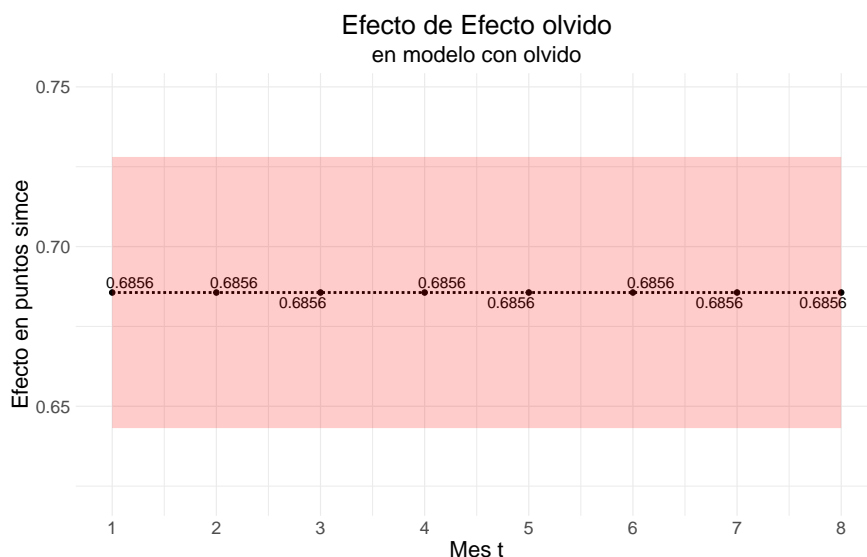


Figura 2.39: Efecto ρ de olvido.

Así, en este caso fue de 0.6856, esto significa que para cada mes t , el grado de descuento que afecta las variables dinámicas en los meses anteriores a este mes, se puede resumir en la siguiente tabla

Meses antes de t	Descuento
0	$\rho^{t-1+1} = 0.6856$
1	$\rho^{t-2+1} = 0.4700$
2	$\rho^{t-3+1} = 0.3223$
3	$\rho^{t-4+1} = 0.2209$
4	$\rho^{t-5+1} = 0.1515$
5	$\rho^{t-6+1} = 0.1039$
6	$\rho^{t-7+1} = 0.0712$
7	$\rho^{t-8+1} = 0.0488$

Tabla 2.23: Efecto descuento para cada mes

Esto será significativo a la hora de analizar los resultados de la siguiente tabla, donde se muestra los efectos de las variable dinámica de precisión de respuesta para cada t .

Efecto	Mes							
	1	2	3	4	5	6	7	8
Núm. y op.	8.8344	20.0109	38.1485	45.5507	53.4960	55.9048	68.6129	71.1036
Pat. y álg.	1.2729	3.2116	6.0180	9.7379	17.9034	17.4209	27.1154	31.6246
Geometría	-0.6988	0.5732	2.1601	9.6131	11.3839	11.7966	13.7649	15.2859
Medición	1.0463	1.5858	8.8767	16.4380	18.8832	23.7324	24.8770	21.5935
Datos y Prob.	2.8363	6.1791	7.6444	10.6768	19.7341	17.5817	26.8128	37.5778

Tabla 2.24: Efectos $\delta_{k,t}^{(1)}$ para los 5 ejes de contenidos dependiendo de la información en t .

Entonces, recordamos como es la forma de esta variable dinámica en particular.

$$\frac{1}{t} \sum_{s=1}^t \rho^{t-s+1} \left(\sum_{k=1}^5 \delta_{k,t}^{(1)} u_{i,j,k,s}^{(1)} \right)$$

así, analizando un caso extremo de un alumno que mantiene una precisión de respuesta para el eje números y operaciones siempre una desviación por sobre la media ($u_{i,j,k,s}^{(1)} = 1$, $k \in \{1, \dots, 5\}, s \in \{1, \dots, 8\}$). En este caso, para cada t tendríamos lo siguiente

$$\frac{1}{t} \sum_{s=1}^t \rho^{t-s+1} \left(\delta_{1,t}^{(1)} \cdot 1 \right) = \frac{1}{t} \sum_{s=1}^t \rho^{t-s+1} \delta_{1,t}^{(1)}$$

luego, para un t dado, se tiene

$$\frac{1}{t} \delta_{1,t}^{(1)} \left(\sum_{s=1}^t \rho^{t-s+1} \right) = \frac{1}{t} \delta_{1,t}^{(1)} \frac{p(p^t - 1)}{p - 1}$$

con esto, si vemos los efectos de la tabla 2.24, tendríamos

Mes	Efecto
$t = 1$	$\frac{1}{1} \cdot 8.8344 \left(\frac{p(p-1)}{p-1} \right) = 6.0569$
$t = 2$	$\frac{1}{2} \cdot 20.0109 \left(\frac{p(p^2-1)}{p-1} \right) = 11.5628$
$t = 3$	$\frac{1}{3} \cdot 38.1485 \left(\frac{p(p^3-1)}{p-1} \right) = 18.7934$
$t = 4$	$\frac{1}{4} \cdot 45.5507 \left(\frac{p(p^4-1)}{p-1} \right) = 19.3402$
$t = 5$	$\frac{1}{5} \cdot 53.4960 \left(\frac{p(p^5-1)}{p-1} \right) = 19.7971$
$t = 6$	$\frac{1}{6} \cdot 55.9048 \left(\frac{p(p^6-1)}{p-1} \right) = 18.2081$
$t = 7$	$\frac{1}{7} \cdot 68.6129 \left(\frac{p(p^7-1)}{p-1} \right) = 19.8526$
$t = 8$	$\frac{1}{8} \cdot 71.1036 \left(\frac{p(p^8-1)}{p-1} \right) = 18.4355$

Tabla 2.25: Ejemplo de un caso con desempeño sobre la media en todos los meses

Este pequeño ejemplo permite apreciar que a partir de $t = 3$, el efecto acumulado para el caso ejemplo se mantiene entre 18 y 20 puntos, y esto plantea una cuestión interesante sobre el efecto ρ y su interacción con los efectos $\delta_{k,t}^{(1)}$. Más aún, podemos hacer este mismo ejercicio con los demás ejes y obtener la siguiente tabla

Efecto \ Mes	1	2	3	4	5	6	7	8
Núm. y op.	6.0570	11.5631	18.7940	19.3468	19.7980	18.2090	19.8536	18.4364
Pat. y álg.	0.8727	1.8558	2.9648	4.1360	6.6258	5.6742	7.8460	8.1999
Geometría	-0.4791	0.3312	1.0642	4.0830	4.2130	3.8423	3.9830	3.9635
Medición	0.7173	0.9163	4.3731	6.9817	6.9884	7.7300	7.1983	5.5990
Datos y Prob.	1.9446	3.5705	3.7660	4.5348	7.3033	5.7266	7.7584	9.7435

Tabla 2.26: Efectos $\delta_{k,t}^{(1)}$ para los 5 ejes de contenidos dependiendo de la información en t .

los cuales serían el puntaje se estima tendría un alumno que siempre se mantiene una desviación por sobre la media en precisión de respuesta para los cinco ejes. Aquí se puede apreciar cómo en ejes como el de números y operaciones de alcanza cierta estabilidad del efecto a partir de $t = 3$, mientras para ejes como el de geometría y el de medición, esto ocurre en $t = 4$, pero para los ejes de patrones y álgebra, y de datos y probabilidad no se alcanza a notar alguna estabilidad del parámetro. Esto coincide con el orden en que se cubrieron los ejes durante el año y la prioridad que se dio a cada uno de los ejes, es decir, los ejes de contenidos que se comenzaron a ver antes, el efecto en el modelo se estabiliza antes que aquellos ejes que comenzaron a abordarse más adelante en el año académico.

2.7. Resumen

Cada uno de los cuatro modelos permite comprender las dinámicas entre las variables históricas y dinámicas gracias a las diferentes formas que cada modelo considera las variables dinámicas. El modelo que logra el RMSE más bajo en las predicciones, es el modelo con olvido, con 26.61 puntos, lo cual lo ubica a 10.05 puntos del SEM de la prueba SIMCE, reduciendo en un 67.8% el error explicable. Otros, como el modelo glotón permite una comprensión más descriptiva de las interacciones entre los 5 ejes de la variable dinámica de precisión de respuesta en el primer intento. Los modelos sin memoria y con memoria permiten comprender cómo la inclusión o no inclusión de los meses anteriores al mes en que se está realizando una predicción pueden afectar para bien o para mal la estimación de puntaje, y esto reportando RMSE de 28.27 y 27.43, respectivamente para cada modelo en el grupo de testeo.

Capítulo 3

Modelos espacio estado para la estimación de desempeño

En este capítulo se presenta los modelos espacio estado, su implementación y los resultados obtenidos. Estos modelos presentan una alternativa que usa 17 parámetros en total y reportamos un RMSE que llega a los 27.52 puntos para estimaciones en el grupo de testeo. Esto se logra haciendo uso de del pretest, 2 variables históricas y 2 variables dinámicas como en los modelos anteriores. Este modelo se plantea como una alternativa flexible para estimaciones fuera de la muestra y usando menos parámetros que los requeridos para los modelos anteriores.

3.1. Modelo Espacio Estado

En el capítulo anterior se hizo uso de modelos que dependían del instante t hasta el cual se tenía información disponible, y entre estos, el que mostró un mejor desempeño, fue aquel modelo que incluyó un efecto ρ de descuento. Esta estructura se asimila bastante a la que uno puede apreciar en los modelos de **espacio estado**, los cuales constan de tres componentes: $x(t) \in \mathbb{R}^n$ el estado latente u oculto, $y(t) \in \mathbb{R}^m$ el estado observable, medible o de salida, y $u(t) \in \mathbb{R}^p$ la variable de control o de entrada, entonces dado un instante $t \geq 0$ se formula el siguiente modelo

$$\begin{aligned}x(t+1) &= A(t)x(t) + B(t)u(t) \\y(t) &= C(t)x(t) + D(t)u(t) \\x(0) &= x_0\end{aligned}\tag{3.1}$$

donde $A(t) \in \mathbb{R}^{n \times n}$, $B(t) \in \mathbb{R}^{n \times p}$, $C(t) \in \mathbb{R}^{m \times n}$ y $D(t) \in \mathbb{R}^{m \times p}$, $t \geq 0$. Luego, se puede considerar un ruido asociado al estado latente, $w(t) \sim \mathcal{N}(0, Q)$ y un ruido de medición, asociado al estado observable, $v(t) \sim \mathcal{N}(0, R)$, independientes entre si; con lo que el modelo será

$$\begin{aligned}
x(t+1) &= A(t)x(t) + B(t)u(t) + w(t) & w(t) &\sim \mathcal{N}(0, Q) \\
y(t) &= C(t)x(t) + D(t)u(t) + v(t) & w(t) &\sim \mathcal{N}(0, R) \\
x(0) &= x_0
\end{aligned} \tag{3.2}$$

donde $Q \in \mathbb{R}^{n \times n}$ y $R \in \mathbb{R}^{m \times m}$, $t \geq 0$.

Ahora, para el caso particular con que se está trabajando, nos interesa un sistema autónomo, es decir, donde las matrices $A(t)$, $B(t)$, $C(t)$ y $D(t)$ no dependen del tiempo. Por lo tanto, usaremos el modelo discreto autónomo, esto es

$$\begin{aligned}
x(t+1) &= Ax(t) + Bu(t) + w(t) & w(t) &\sim \mathcal{N}(0, Q) \\
y(t) &= Cx(t) + Du(t) + v(t) & w(t) &\sim \mathcal{N}(0, R) \\
x(0) &= x_0
\end{aligned} \tag{3.3}$$

donde $t \in \{0, \dots, 9\}$.

3.2. Estados y variable de control

Debemos ahora considerar las dimensiones del modelo lineal que hemos usado y las variables disponibles. Lo primero será determinar el estado observable, que teniendo el objetivo de estimar el puntaje SIMCE, es natural establecer que éste sea el estado $y(t)$, pero en tal caso no contamos con un puntaje para cada $t \in \{0, \dots, 9\}$, sino que tan solo para $t = 9$, pero una de las variables que si tenemos para cada t , es la calificación $n_{i,j,k,t}$ por eje asignada por el sistema, la cual promediamos para cada mes. De igual modo, de la tabla 1.1, conocemos qué porcentaje q_k de las preguntas del SIMCE corresponde a cada eje $k \in \{1, \dots, 5\}$, por lo que podemos tomar el promedio ponderado de las notas para cada eje y obtener una calificación que pondera estos de igual modo que el SIMCE. Entonces definimos:

$$y_i(t) = \begin{cases} \sum_{k=1}^5 q_k n_{i,j,k,t} & t \in \{1, \dots, 8\} \\ y_i & t = 9 \end{cases} \tag{3.4}$$

donde las notas han sido previamente normalizadas y reescaladas con la media y desviación del SIMCE.

Luego, para determinar $x(t)$ simplemente es necesario establecer la dimensión de este, puesto que al ser el estado no observable, no tenemos mediciones de este. Podríamos considerar que es de dimensión $n = 5$, considerando los 5 ejes de contenidos, pero por simpleza trabajaremos con $n = 1$ para que la dimensión sea la misma que el estado observable.

El siguiente estado a definir será $u(t)$, donde deberemos incluir aquellas variables que eran los regresores en el modelo lineal con olvido, esto es, el pretest p_i , las variables históricas $\Theta_{i,j}^{(m)}$ y variables dinámicas $u_{i,j,1,t}^{(m)}$, $m \in \{1, 2\}$. Como el pretest es unidimensional, al igual que cada variable histórica, entre estos son 3 dimensiones, y las variables dinámicas están separadas por ejes, por lo que entre las dos variables dinámicas constituyen 10 dimensiones. Luego, la estructura de $u(t)$ será algo como:

$$u_i(t) = \begin{pmatrix} u_{i,j,1,t}^{(1)} \\ \vdots \\ u_{i,j,5,t}^{(1)} \\ u_{i,j,1,t}^{(2)} \\ \vdots \\ u_{i,j,5,t}^{(2)} \\ p_i \\ \Theta_{i,j}^{(1)} \\ \Theta_{i,j}^{(2)} \end{pmatrix} \quad (3.5)$$

Primero, están las variables dinámicas, es decir variables provenientes de la plataforma ConectIdeas, producto de la ejercitación hecha por los estudiantes, en el instante $t = 0$ son todas cero, pues son inexistentes; pero a partir del primer mes ($t = 1$), ya se puede tener información en estas. Entonces para las primeras 10 coordenadas del vector $u(t)$ ($u^{(\ell)}(t), \ell \in \{1, \dots, 10\}$), que corresponden a las variables dinámicas, serán tales que:

$$u_i^{(\ell)}(t) = \begin{cases} u_{i,j,\ell,t}^{(1)} & \ell \in \{1, \dots, 5\} \\ u_{i,j,\ell-5,t}^{(2)} & \ell \in \{6, \dots, 10\} \end{cases}, \forall t > 0 \quad (3.6)$$

Por su parte, las variables históricas son invariables en el tiempo, pero sabemos que el efecto de estas puede variar en el tiempo, al igual que el pretest, como se pudo ver en las figura 2.15, 2.22, 2.29 y 2.36 (siendo de particular interés la última de estas), es decreciente en la medida que se avanza en el tiempo en cuanto a la información disponible.

Debido a esto, consideramos que en las coordenadas 11 a 13 del vector $u(t)$ ($u^{(\ell)}(t), \ell \in \{11, 12, 13\}$) tendremos que considerar algunas modificaciones si queremos emular estos comportamiento.

En modelos espacio estado, el factor principal de decrecimiento o crecimiento estará dado por A , y haremos uso de este parámetro para adaptar la forma del efecto de B en relación con las variables históricas y el pretest. Para esto se considerará 6 posibles comportamientos en base a las tendencias observadas de los efectos de variables históricas en los modelos lineales.

1. Constante.
2. Decreciente.
3. Creciente.

4. Decreciente hasta un instante t y luego constante.
5. Creciente hasta un instante t y luego constante.
6. Decreciente hasta un instante t y luego creciente.

En los 6 casos consideramos que el factor de crecimiento o decrecimiento siempre será producto del efecto A , el cual estará acotado entre 0 y 2; lo que significará que nuestra estructura deberá ser distinta si $A \leq 1$ ó $A > 1$. Con esto podremos hacer un análisis similar al realizado, y para simpleza de notación trabajaremos siempre con $u(t)$ unidimensional para esta pequeña sección.

Para cada caso, consideraremos el caso simplificado en que $x(t)$ y $u(t)$ son unidimensional, puesto que solo nos interesa el efecto asociado a una dimensión específica $\ell \in \{11, 12, 13\}$. Sea entonces $A \leq 1, B, x(0) = x_0$ y $u(0) = u_0$ dados. Notando de antemano que si $A > 1$, tan solo se invierte el comportamiento de cuando es decreciente por cuando es creciente y viceversa.

3.2.1. Efecto $Bu(t)$ constante.

Para mantener siempre el mismo efecto que en el instante inicial, de Bu_0 , se impone en el primer instante lo siguiente

$$x(1) = Ax_0 + \underbrace{Bu_0}_{\text{efecto en } t=0} \implies x(2) = A^2x_0 + \underbrace{ABu_0 + Bu(1)}_{\text{efecto en } t=1}$$

Entonces queremos imponer lo siguiente

$$ABu_0 + Bu(1) = Bu_0 \implies u(1) = (1 - A)u_0$$

y repitiendo este procedimiento se puede mantener el efecto $Bu(t)$ constante en t . Es decir, $u(t)$ quedará definido como

$$u(t) = \begin{cases} u_0 & t = 0 \\ (1 - A)u_0 & t > 0 \end{cases} \quad (3.7)$$

3.2.2. Efecto $Bu(t)$ decreciente.

Como $A \leq 1$, el simple efecto acumulado de este parámetro genera un decaimiento del efecto de $Bu(t)$.

$$\begin{aligned} x(1) &= Ax(0) + Bu(0) = Ax_0 + Bu_0 \\ \implies x(2) &= Ax(1) + \underbrace{Bu(1)}_{=0} = A(Ax_0 + Bu_0) = A^2x_0 + \underbrace{ABu_0}_{\text{efecto en } t=2} \\ \implies x(3) &= Ax(2) + \underbrace{Bu(2)}_{=0} = A^3x_0 + \underbrace{A^2Bu_0}_{\text{efecto en } t=3} \\ &\vdots \\ \implies x(t) &= A^tx_0 + \underbrace{A^{t-1}Bu_0}_{\text{efecto en } t} \end{aligned}$$

Luego, el valor del efecto mes a mes termina siendo la secuencia $(A^s Bu_0)_s = 0^t$, y como hemos considerado $A \leq 1$, es decreciente. Entonces la estructura es la siguiente.

$$u(t) = \begin{cases} u_0 & t = 0 \\ 0 & t > 0 \end{cases} \quad (3.8)$$

3.2.3. Efecto $Bu(t)$ creciente.

Este caso es lo opuesto al anterior, si el efecto olvido nos genera un decrecimiento de un $(1 - A)\%$ al multiplicar por A , un crecimiento del mismo orden de porcentaje, será de un $(1 - A)\%$, es decir, multiplicar por $[1 + (1 - A)] = (2 - A)$, es decir

$$\begin{aligned} ABu_0 & \text{ es un } (1 - A)\% \text{ menor que } Bu_0 \\ (2 - A)Bu_0 & \text{ es un } (1 - A)\% \text{ mayor que } Bu_0 \end{aligned}$$

Entonces, igual que el caso anterior, tomamos el efecto en el primer instante

$$x(1) = Ax_0 + Bu_0 \implies x(2) = A^2x_0 + ABu_0 + Bu(1)$$

Luego, se impone que el efecto en $t = 1$ (que se ve en la ecuación de $x(2)$) debe ser un $(1 - A)\%$ mayor que el instante anterior, esto es

$$ABu_0 + Bu(1) = (2 - A)Bu_0 \implies u(1) = 2(1 - A)u_0$$

y al instante siguiente

$$A \underbrace{((2 - A)Bu_0)}_{\text{efecto en } t=1} + Bu(2) = (2 - A) \underbrace{((2 - A)Bu_0)}_{\text{efecto en } t=1} \implies u(2) = 2(1 - A)u_0(2 - A)$$

y repitiendo este procedimiento se deduce que para obtener un efecto creciente es necesario imponer

$$u(t) = \begin{cases} u_0 & t = 0 \\ 2(1 - A)u_0(2 - A)^{t-1} & t > 0 \end{cases} \quad (3.9)$$

3.2.4. Efecto $Bu(t)$ decreciente y constante.

Es una mezcla de los casos 3.8 y 3.7; esto es, que hasta un instante \bar{t} el efecto será decreciente como en el segundo caso considerado, para luego ser constante, es decir que cada mes se sumará un mismo efecto tamaño del efecto asociado.

$$u(t) = \begin{cases} u_0 & t = 0 \\ 0 & 0 < t \leq \bar{t} \end{cases}$$

Luego, para que el efecto que se aprecia cada mes en el modelo espacio estado sea el mismo, anteriormente se había considerado $u(t) = (1 - A)u_0$, pero esto es en el caso que el efecto sea constante desde el instante 1, si es a partir del instante \bar{t} , ya se habrá experimentado un decaimiento de $A^{\bar{t}}$. Por lo tanto, agregando este decaimiento se ajusta el valor para $u(t)$ de modo que el efecto se mantenga constante a partir de \bar{t} .

$$u(t) = \begin{cases} u_0 & t = 0 \\ 0 & 0 < t \leq \bar{t} \\ A^{\bar{t}}(1 - A)u_0 & t > \bar{t} \end{cases} \quad (3.10)$$

3.2.5. Efecto $Bu(t)$ creciente y constante.

Ahora, combinando los casos 3.9 y 3.7, similar al caso anterior, se usa los resultados del caso creciente hasta \bar{t} , es decir

$$u(t) = \begin{cases} u_0 & t = 0 \\ 2(1 - A)u_0(2 - A)^{t-1} & 0 < t \leq \bar{t} \end{cases}$$

Luego, se considera el efecto del crecimiento hasta \bar{t} y la compensación necesaria para mantener constante este efecto. Es decir, una compensación de $(2 - A)^{\bar{t}}$ producto del crecimiento. Con esto, se obtiene lo siguiente:

$$u(t) = \begin{cases} u_0 & t = 0 \\ 2(1 - A)u_0(2 - A)^{t-1} & 0 < t \leq \bar{t} \\ (1 - A)u_0(2 - A)^{\bar{t}} & t > \bar{t} \end{cases} \quad (3.11)$$

3.2.6. Efecto $Bu(t)$ decreciente y creciente.

Finalmente, para este caso unimos 3.8 y 3.9. En este caso, tenemos un decaimiento hasta un \bar{t} dado para luego considerar el crecimiento, por lo que la estructura hasta ese instante será similar al caso 3.11, pero a partir de ahí se produce la diferencia.

La diferencia en este caso se observará en la expresión:

$$u(t) = 2(1 - A)u_0A^{\bar{t}}(2 - A)^{t-1}$$

Esta fórmula para el efecto creciente, es considerando a partir de $t = 1$, por lo que habrá que considera el descuento experimentado hasta \bar{t} , es decir $A^{\bar{t}}$ y al mismo tiempo considerar que el crecimiento ocurre a partir de \bar{t} , por lo que la expresión $(1 - A)^{t-1}$ deberá experimentar un desfase, es decir pasará a ser $(2 - A)^{t-\bar{t}-1}$. Con esto, tenemos que:

$$u(t) = \begin{cases} u_0 & t = 0 \\ 0 & 0 < t \leq \bar{t} \\ 2(1 - A)u_0A^{\bar{t}}(2 - A)^{t-\bar{t}-1} & t > \bar{t} \end{cases} \quad (3.12)$$

3.2.7. Resumen

Con los 6 casos considerados, cada variable histórica será asignada dos de estas opciones en base a los resultados observados en los modelos lineales. La opción por defecto será el modelo con efecto decreciente (o creciente si $A > 1$) y el segundo será aquel que se ajuste mejor a las curvas obtenidas entre los 20 mejores resultados en los modelos 2.6 y con olvido.

Con todo esto, hemos determinado que $x(t) \in \mathbb{R}$, $y(t) \in \mathbb{R}$ y $u(t) \in \mathbb{R}^{13}$. Así, esto nos determina las dimensiones de cada una de las matrices, por lo que $A \in \mathbb{R}$, $B \in \mathbb{R}^{1 \times 13}$, $C \in \mathbb{R}$, $D \in \mathbb{R}^{1 \times 13}$, $Q \in \mathbb{R}$, $R \in \mathbb{R}$ y el estado inicial $x_0 \in \mathbb{R}$. Todas estas matrices y el estado inicial serán parámetros a estimar, pero primeramente notamos que como el estado latente y el estado oculto son unidimensionales, esto hace que la matriz C también lo sea, entonces para términos prácticos lo único que esta hará será reescalar el valor de $x(t)$ para obtener el

valor $y(t)$ en la escala SIMCE, por lo que podemos sin pérdida de generalidad asumir que ambos estados están en la misma escala y en tal caso $C = 1$. Por otro lado, no interesa cómo las variables históricas y dinámicas inciden en el desempeño en instantes futuros, en especial en el instante $t = 9$, por lo que para los objetivos de esta investigación no es de interés considerar el efecto que estas tienen sobre la medición, es decir, podemos considerar $D = \vec{0}$. Con esto, el sistema queda de la siguiente forma

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) + w(t) & w(t) &\sim \mathcal{N}(0, Q) \\ y(t) &= x(t) + v(t) & v(t) &\sim \mathcal{N}(0, R) \quad (\text{espacio estado}) \\ x(0) &= x_0 \end{aligned}$$

Así, será necesario estimar $A \in \mathbb{R}, B \in \mathbb{R}^{1 \times 13}, Q \in \mathbb{R}, R \in \mathbb{R}$ y $x_0 \in \mathbb{R}$, es decir, 17 parámetros.

3.3. Estimación de parámetros

Para la estimación de parámetros, podríamos simplemente usar la información disponible en $t \in \{0, \dots, 8\}$ y minimizar el error de estimación en $t = 9$, pero de igual manera que con los modelos lineales, nos interesa definir un modelo que no solo realice buenas estimaciones del SIMCE cuando se tiene toda la información, sino también cuando se tiene información de forma parcial, es decir, para todos los $t \leq 8$, puesto que será de interés hacer estimaciones en la medida que se va avanzando en el programa ConectaIdeas. Esto, plantea una situación, dado que a diferencia del modelo lineal, donde uno simplemente puede agregar o reducir las variables en la medida que aumentan los meses, en este caso el modelo necesitará alguna entrada para $u(t)$ en los instantes t de los cuales no se tiene información.

Sea t el instante hasta el cual se tiene información disponible, entonces para todo $s > t$ será necesario construir una estimación o emulación de $u(s)$, que llamaremos $\hat{u}_t(s)$. Esta estimación dependerá también de la coordenada $\ell \in \{1, \dots, 10\}$ que son las coordenadas que corresponden con las variables dinámicas. Para realizar esta simulación consideramos tres alternativas

$$\hat{u}_t^{(\ell)}(s) = \frac{1}{t} \sum_{r=1}^t u(r) \quad \forall s > t, \ell \in \{1, \dots, 8\} \quad (3.13a)$$

$$\hat{u}_t^{(\ell)}(s) = \begin{cases} \left(\frac{1}{t} \sum_{r=1}^t \sum_{k=1}^5 u^{(k)}(r) \right) M_{k,t} & \ell \in \{1, \dots, 5\} \\ \frac{1}{t} \sum_{r=1}^t u^{(\ell)}(r) & \ell \in \{6, \dots, 10\} \end{cases} \quad \forall s > t, \ell \in \{1, \dots, 8\} \quad (3.13b)$$

$$\hat{u}_t^{(\ell)}(s) = u(t)^{(\ell)} \quad \forall s > t, \ell \in \{1, \dots, 8\} \quad (3.13c)$$

donde $M_{k,t}$ indica la proporción promedio de ejercicios del eje k que se llevaron a cabo en el mes t . Entonces, dados A_0, B_0, Q_0, R_0 y \hat{x}_0 valores iniciales de estimación para A, B, Q, R y x_0 , se calcula la matriz de ganancia de Kalman[20], la cual es:

$$K = [A\bar{P}C^T + Cov(w(t), v(t))] [C\bar{P}C^T + R]^{-1} \quad (3.14)$$

donde \bar{P} se obtiene como una solución semi definida positiva a la ecuación estacionaria de Riccati (DARE por Discrete time Algebraic Riccati Equation)[7]:

$$\bar{P} = A\bar{P}A^T + Q - [A\bar{P}C^T + Cov(w(t), v(t))] [C\bar{P}C^T + R]^{-1} [A\bar{P}C^T + Cov(w(t), v(t))]^T \quad (3.15)$$

Luego, dado que en el caso que estamos trabajando consideramos que $C = 1$ y $w(t), v(t)$ son independientes entre si, por lo tanto $Cov(w(t), v(t)) = 0$. Con esto, la ecuación de Riccati se simplifica a lo siguiente:

$$\bar{P} = A\bar{P}A^T + Q - [A\bar{P}] [\bar{P} + R]^{-1} [A\bar{P}]^T \quad (3.16)$$

y de igual modo la ecuación 3.14 se convierte en:

$$K = [A\bar{P}] [\bar{P} + R]^{-1} \quad (3.17)$$

Teniendo la matriz de Kalman K , se plantea la estimación $\hat{x}_t(s)$ del estado latente para todo $s \leq t$

$$\hat{x}_t(s+1) = A\hat{x}_t(s) + Bu(s) + K(y(s) - \hat{x}_t(s)), \quad \forall s \in \{0, \dots, t\} \quad (3.18)$$

Luego, para $s > t$ se construye una estimación $\hat{\hat{x}}(s)$ dada por

$$\begin{aligned} \hat{\hat{x}}_t(s+1) &= A\hat{\hat{x}}_t(s) + B\hat{u}(s) \\ \hat{\hat{x}}_t(t+1) &= \hat{x}_t(t+1) \\ \hat{y}(s) &= \hat{\hat{x}}_t(s) \end{aligned} \quad \forall s \in \{t+1, \dots, 9\} \quad (3.19)$$

Así se obtiene una estimación de $y(s=9)$, que será $\hat{y}_t(s=9)$ que usa información disponible hasta t . Con esto se obtiene la raíz del error cuadrático medio (rsme por Root Square Mean Error) con datos hasta t dado por

$$rmse_t = \sqrt{\frac{1}{N} \sum_{i=1}^N (y(9) - \hat{y}_t(9))^2} \quad (3.20)$$

donde N será el número de alumnos en la base de datos de entrenamiento. Este proceso se lleva a cabo para $t \in \{1, \dots, 8\}$ y se calcula el promedio de estos casos para obtener el error del modelo

$$error = \frac{1}{8} \sum_{t=1}^8 rmse_t = \frac{1}{8} \sum_{t=1}^8 \sqrt{\frac{1}{N} \sum_{i=1}^N (y(9) - \hat{y}_t(9))^2} \quad (3.21)$$

Entonces se plantea el problema de optimización

$$\begin{aligned} \min_{A, B, Q, R, x_0} & \quad \frac{1}{8} \sum_{t=1}^8 \sqrt{\frac{1}{N} \sum_{i=1}^N (y(9) - \hat{y}_t(9))^2} \\ \text{s.a:} & \quad \hat{y}_t(9) \text{ dado por (3.18) - (3.19)} \\ & \quad 0 \leq A \leq 2 \\ & \quad -200 \leq B^{(\ell)} \leq 200 \quad , \ell \in \{1, \dots, 13\} \\ & \quad 0 \leq Q \leq 200 \\ & \quad 0 \leq R \leq 200 \\ & \quad 200 \leq x_0 \leq 300 \end{aligned} \quad (3.22)$$

donde las restricciones fueron derivadas de los resultados obtenidos en los modelos lineales.

El problema es planteado y se resuelve usando un algoritmo De punto interior mediante la función `fmincon` de MATLAB, del paquete de optimización[23].

Al igual que para los modelos lineales, la estimación de los parámetros del modelo, en este caso 17, se obtiene buscando el mejor ajuste en un grupo de entrenamiento (18 establecimientos, 75% del total), y después se comprueba si el desempeño es similar en los 6 establecimientos de testeo. Este procedimiento se repite realizando validación cruzada 100 veces para obtener un desempeño promedio del modelo y también una estimación promedio de los parámetros.

Aprovechando los resultados obtenidos al realizar las validaciones cruzadas con los 4 modelos del capítulo anterior, se selecciona los 20 con mejor desempeño en cada modelo y se prueba las variables de cada uno de estos con el modelo espacio estado propuesto. A priori, esto significaría realizar 100 validaciones de 80 modelos, sin embargo, existe algunos modelos que comparten el conjunto de variables históricas y dinámicas que emplean, por lo que se trabajó con una lista de 71 combinaciones de variables para probar.

Habiendo probado las combinaciones, se selecciona los 5 mejores de este grupo (tanto viendo el desempeño promedio como en el último mes), y se observa que existe 3 puntos en común en estos modelos:

1. Consideran el pretest como parte del modelo.
2. Incluyen 2 variables históricas
3. Una de las variables históricas es $com_{i,j}$, el promedio curso de las respuestas sobre

el nivel de acuerdo sobre “Mi comportamiento es un problema para el profesor de matemáticas”, de los alumnos del establecimiento j .

4. Una de las variables dinámicas es $a_{i,j,k,t}$, la precisión de respuesta del alumno i en el primer intento a ejercicios del eje k en el mes t .

Con esto, los mejores modelos tienen varias similitudes y más aún, si se observa hasta los mejores 8, sin considerar la segunda variable dinámica, hay solo 4 conjuntos de variables distintos; es decir los mejores 8 son en realidad 4 pares que difieren tan solo por una variable dinámica. Considerando esto, se procede a elegir estas 4 combinaciones de 2 variables históricas (siendo una $com_{i,j}$) y la variable dinámica $a_{i,j,k,t}$, para luego probar con otra de las 8 variables dinámicas disponibles (pues de las 11, ya estamos usando la precisión de respuesta como variable dinámica, la nota promedio como output y y las componentes principales solo se usan solas por colinealidad).

Es decir, las primeras 5 coordenadas $u_{i,j,t}^{(k)}$, $k \in \{1, \dots, 5\}$ podrán tomar una de las siguientes opciones:

1. $e_{i,j,k,t}$, el número de ejercicios resueltos por el alumno i del eje k resueltos en el mes t .
2. $f_{i,j,k,t}$, el número de ejercicios resueltos por el alumno i del eje k resueltos en el primer intento en el mes t .
3. $g_{i,j,k,t}$, el número de ejercicios resueltos por el alumno i del eje k resueltos en más de un intento en el mes t .
4. $d_{i,j,k,t} = f_{i,j,k,t} - g_{i,j,k,t}$, la diferencia en el número de ejercicios resueltos en el primer intento y los resueltos en más de un intento por el alumno i del eje k , en el mes t .
5. $o_{i,j,k,t}$: Número de preguntas abiertas contestadas.
6. $p_{i,j,k,t}$: Total de palabras contestadas en preguntas abiertas.
7. $q_{i,j,k,t}$: Promedio de palabras por pregunta abierta contestada.
8. $r_{i,j,k,t}$: Calificación promedio en preguntas abiertas.

Así, se obtiene 8 variantes del modelo espacio estado dependiendo de qué variable dinámica ocupa las primeras 5 coordenadas de $u(t)$. Con esto, tendremos 32 conjuntos de variables para el modelo donde seleccionaremos aquella que presente el mejor desempeño entre todas.

3.4. Resultados

Manteniendo consistencia con los resultados anteriores, se realizaron 100 validaciones cruzadas para cada instancia del modelo, con lo cual se puede obtener un desempeño promedio en entrenamiento y en testeo, lo que permitirá comparar con las formulaciones lineales 2.6 y pseudo lineal con olvido.

Luego del proceso descrito anteriormente, el conjunto de variables seleccionadas es el siguiente

Pretest	Variables históricas	Variables dinámicas
Si	“Mi comportamiento es un problema para el profesor de matemáticas” (media curso)	Precisión de respuesta en el primer intento
	“Matemáticas es fácil para mí” (respuesta por alumno)	Diferencia entre numero de preguntas resueltas en un intento y más intentos

Tabla 3.1: Variables que entregaron un mejor desempeño para el modelo espacio estado

Ahora, antes de detallar el RMSE asociado al modelo espacio estado con las variables de la tabla, es importante recalcar las otras variantes en consideradas para este modelo, esto es el tipo de estimación que se está haciendo en $\hat{u}(t)$ cuando la información no está disponible, y si se está realizando algún ajuste en las dimensiones asociadas a las variables históricas.

Primero, de las tres estrategias planteadas para la estimación de las coordenadas de $u(t)$ asociadas a las variables dinámicas, aquella que entregó un mejor desempeño fue la tercera (3.13c), que corresponde a estimar replicando el desempeño en el último mes con información disponible.

Respecto el eventual ajuste en las coordenadas 11, 12 y 13 de $u(t)$, el modelo que hizo uso de las curvas modificadas tuvo un error promedio menor y las curvas para cada coordenadas son:

11. El pretest, se trabaja con el ajuste por defecto 3.8, es decir, un efecto cuyo decrecimiento o crecimiento depende directamente del efecto A .
12. La variable histórica sobre el comportamiento, se ajusta con 3.10 con $t = 4$, donde si $A < 1$, será decreciente y luego constante, si $A \geq 1$, será creciente y luego ligeramente creciente.
13. La segunda variable histórica respecto de la apreciación de los alumnos si matemáticas es fácil para ellos, tiene el mismo ajuste que el pretest, esto es 3.8.

y con estas, se obtiene un desempeño promedio mejor que los primeros dos modelos lineales, que es algo significativo, considerando que este modelo tiene 17 parámetros en comparación con los anteriores (ver tabla 2.4).

Entonces, usando la notación del modelo con olvido, la estructura de $u(t)$ será:

$$u_i(t) = \begin{cases} (0, \dots, 0, 0, \dots, 0, p_i, com_{i,j}, mf_{i,j})^T, & t = 0 \\ (d_{i,j,1,t}, \dots, d_{i,j,5,t}, a_{i,j,1,t}, \dots, a_{i,j,5,t}, 0, 0, 0)^T, & t \in \{1, 2, 3\} \\ (d_{i,j,5,t}, \dots, d_{i,j,5,t}, a_{i,j,1,t}, \dots, a_{i,j,5,t}, 0, 0, A^{\bar{t}}(1-A)mf_{i,j})^T, & t \geq 4 \end{cases} \quad (3.23)$$

donde recordamos que:

- $com_{i,j}$, promedio curso de las respuestas sobre el nivel de acuerdo sobre “Mi comportamiento es un problema para el profesor de matemáticas”, de los alumnos del establecimiento j .
- $mf_{i,j}$, Respuestas de cada alumno i del establecimiento j sobre el nivel de acuerdo sobre “Matemática es fácil para mi”.
- p_i , el puntaje en pretest del alumno i .
- $d_{i,j,k,t} = f_{i,j,k,t} - g_{i,j,k,t}$, la diferencia en el número de ejercicios resueltos en el primer intento y los resueltos en más de un intento por el alumno i del eje k , en el mes t .
- $a_{i,j,k,t}$, la precisión de respuesta del alumno i en el primer intento a ejercicios del eje k en el mes t .

En el siguiente gráfico y tabla se detalla el error promedio obtenido para cada mes.

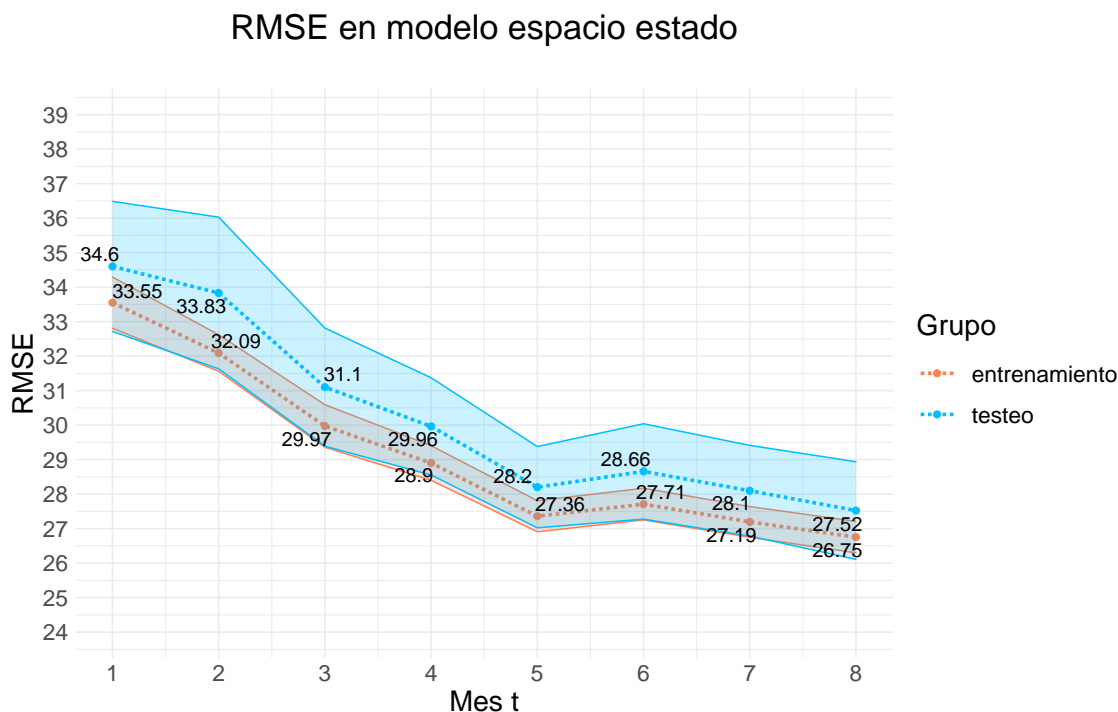


Figura 3.1: RMSE en el modelo espacio estado

Meses	1	2	3	4	5	6	7	8
RMSE promedio en entrenamiento	33.55	32.09	29.97	28.90	27.36	27.71	27.19	26.75
RMSE promedio en testeo	34.60	33.83	31.10	29.96	28.20	28.66	28.10	27.52
Desviación RMSE entrenamiento	0.74	0.53	0.62	0.50	0.45	0.46	0.45	0.46
Desviación RMSE testeo	1.89	2.20	1.71	1.41	1.18	1.38	1.31	1.42

Tabla 3.2: RMSE promedio y desviación del RMSE para entrenamiento y testeo en el modelo espacio estado

Estos resultados son bastante prometedores porque en orden de magnitud, son tan buenos como el modelo con memoria, pero con la diferencia que el RMSE baja antes que ese modelo, es decir, con menos meses de información logra resultados similares y llegando a $t = 8$ se acercan más entre sí.

Esto nos da la ventaja de un modelo que logra buenos desempeños antes en el año. Por ejemplo, en $t = 4$ el RMSE se encuentra bajo los 30 puntos, mientras que esto ocurre en $t = 5$ para el modelo con memoria.

Al ver algunos casos ejemplos, se aprecia cómo estos resultados no solo ocurren de forma agregada (en el RMSE), sino también en las estimaciones para cada individuo.

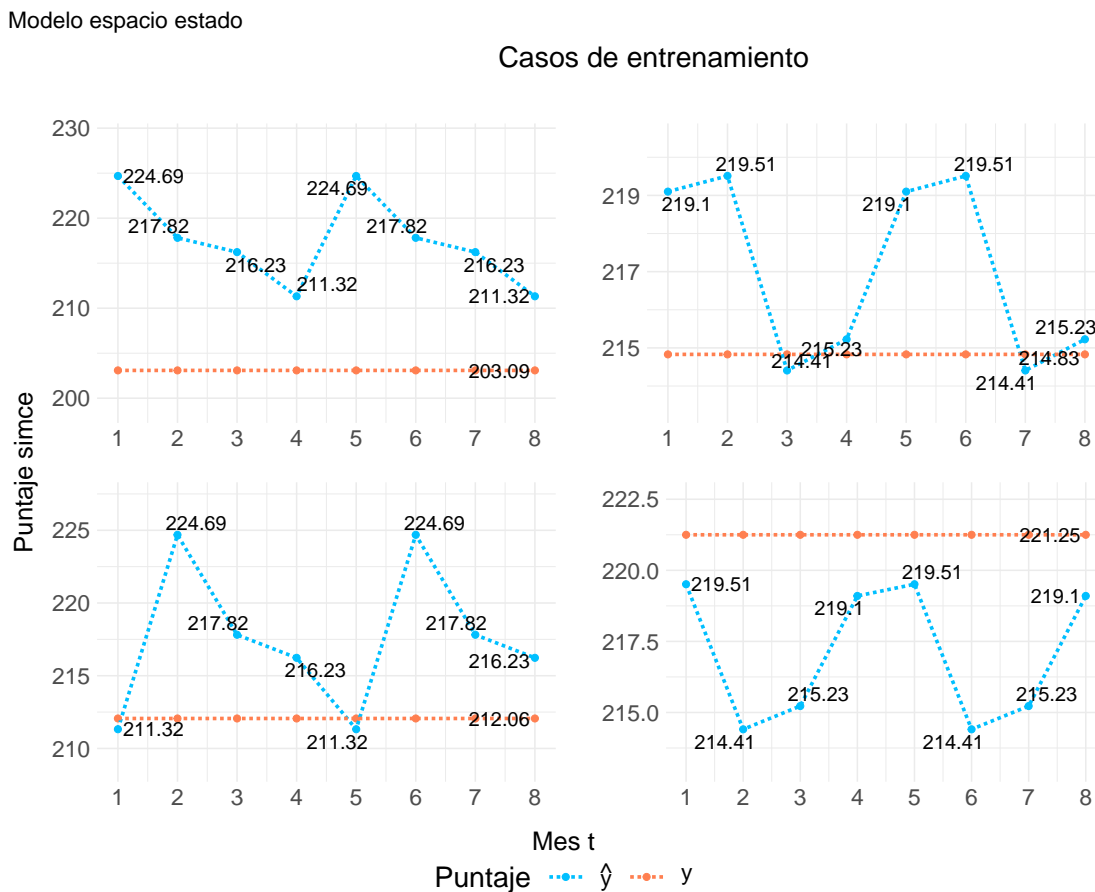


Figura 3.2: Cuatro casos en entrenamiento y sus distintas predicciones según el mes t de información usada

Se puede ver la tendencia del modelo a acercarse al puntaje real en pocos meses. Los casos de entrenamiento por su parte, muestran lo siguiente.

Casos de testeo

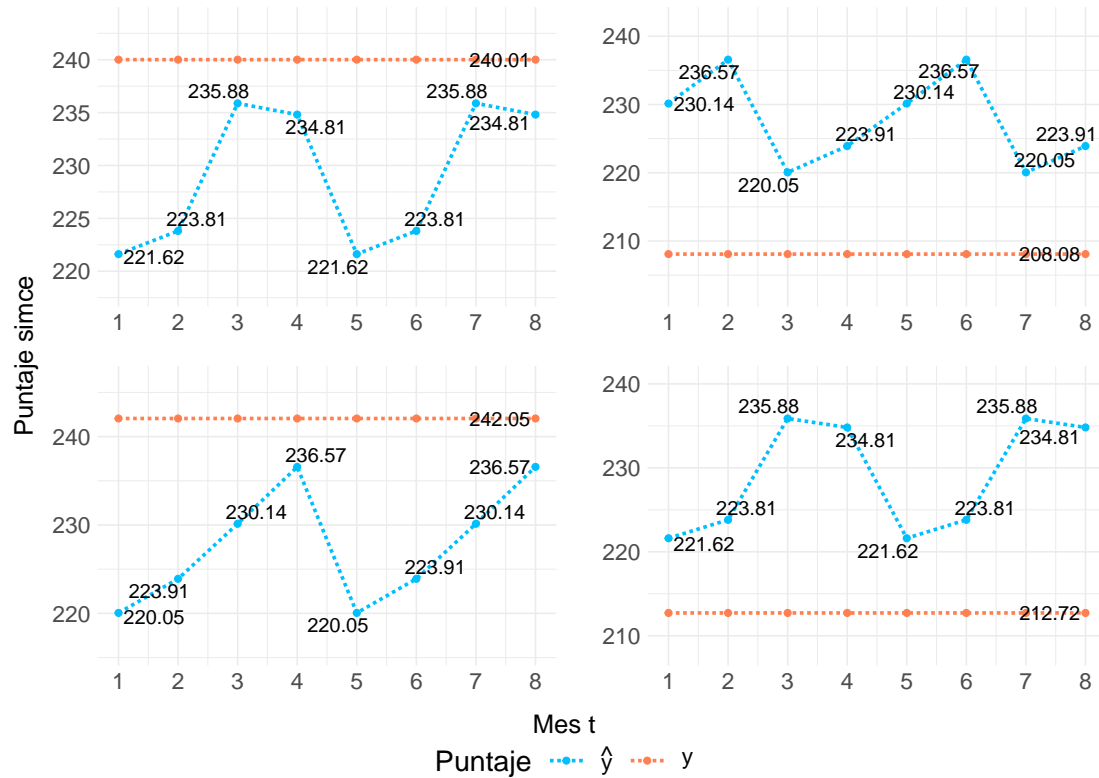


Figura 3.3: Cuatro casos en testeo y sus distintas predicciones según el mes t de información usada

Es decir, que también en los casos en entrenamiento se puede apreciar una tendencia más marcada de una estimación más cercana en la medida que se dispone más meses de información, en lugar de las oscilaciones que se apreciaron en modelos lineales. Esto es, que el error de estimación para cada individuo decrece en t casi de forma monótona.

Siguiendo la misma estructura que para los modelos lineales, observamos el RMSE para la estimación de promedios curso.

RMSE por curso en modelo espacio estado

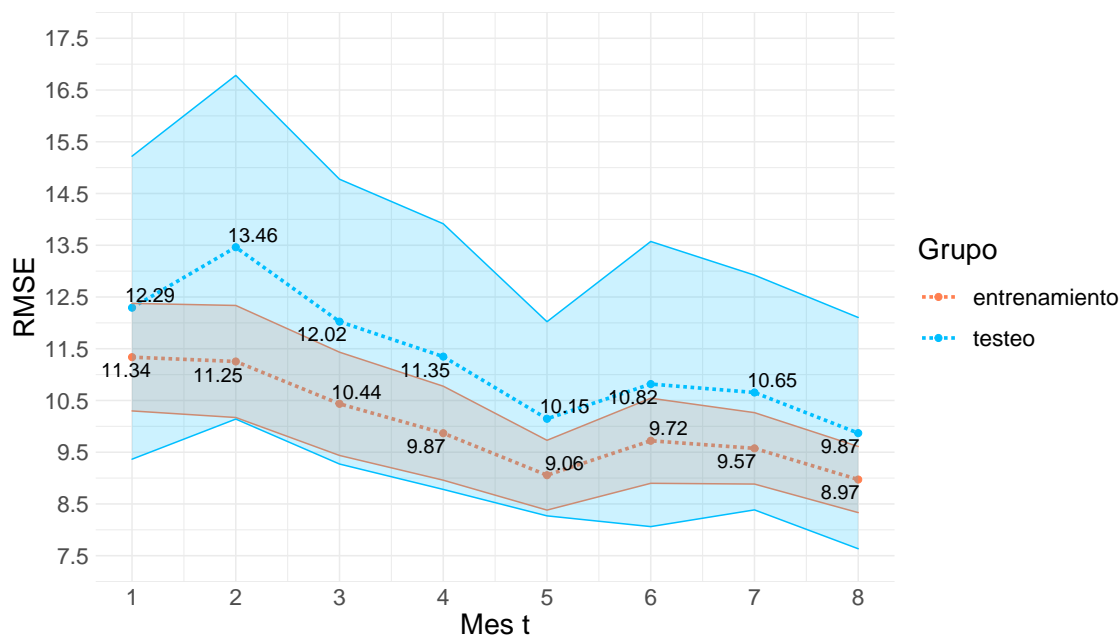


Figura 3.4: RMSE del promedio curso en el modelo espacio estado

Meses	1	2	3	4	5	6	7	8
RMSE promedio en entrenamiento	11.34	11.25	10.44	9.87	9.06	9.72	9.57	8.97
RMSE promedio en testeo	12.29	13.46	12.02	11.35	10.15	10.82	10.65	9.87
Desviación RMSE entrenamiento	1.04	1.08	1.00	0.91	0.67	0.82	0.69	0.64
Desviación RMSE testeo	2.93	3.32	2.75	2.57	1.88	2.76	2.27	2.23

Tabla 3.3: RMSE promedio del promedio curso y desviación de este RMSE para entrenamiento y testeo en el modelo espacio estado

Estos resultados, si bien no son los mejores entre los 5 modelos, deja al modelo espacio estado bien posicionado respecto del resto, puesto que el único modelo que con RMSE menor de forma consistente es el modelo con efecto olvido. Es importante mencionar que pese a que el primer modelo (sin memoria) tiene un RMSE más bajo en $t = 8$ en testeo, el modelo espacio estado tiene un RMSE menor por un punto que este, entre los meses 4 y 7.

A continuación se muestra un resumen del promedio entre las validaciones cruzadas de los valores estimado para cada uno de los 17 parámetros, e indicando la desviación de estos.

Parámetro	Valor	Desviación	Efecto o variable asociada
A	1.0004	0.0002	Olvido
B_1	1.0559	0.1737	$d_{i,j,1,t}$
B_2	0.5090	0.1005	$d_{i,j,2,t}$
B_3	-0.2504	0.1165	$d_{i,j,3,t}$
B_4	0.5377	0.1229	$d_{i,j,4,t}$
B_5	0.0209	0.1198	$d_{i,j,5,t}$
B_6	0.6039	0.2087	$a_{i,j,1,t}$
B_7	0.3370	0.1210	$a_{i,j,2,t}$
B_8	0.2528	0.1571	$a_{i,j,3,t}$
B_9	0.1628	0.1657	$a_{i,j,4,t}$
B_{10}	0.5129	0.1195	$a_{i,j,5,t}$
B_{11}	26.2960	1.3009	p_i
B_{12}	-13.2074	2.1581	$com_{i,j}$
B_{13}	3.3171	2.8892	$mf_{i,j}$
Q	28.0428	35.1162	Varianza de estado latente $x(t)$
R	69.3524	55.2830	Varianza de estado observable $y(t)$
x_0	261.9954	3.1589	Estado inicial

Tabla 3.4: Parámetros del modelo espacio estado

donde entre el modelo con un crecimiento para B_{11} , B_{12} y B_{13}

Si bien, el modelo espacio estado, a diferencia de los modelos anteriores estima un solo valor para los parámetros en lugar de obtener estimaciones para cada mes, debido a la estructura de este modelo, la forma en que estos efectos interactúan en el modelo en cada mes, no es siempre la misma, en especial para los caso en que se ajustó la curva de manera artificial.

3.4.1. Análisis de resultados

El primer parámetro, A , pese a lo que se podría esperar de los resultados con los modelos 2.6 y con olvido, resultó ser un valor > 1 , y una de las primeras consecuencias que esto conlleva es que en lugar de funcionar como un efecto de olvido, funciona como un efecto de maduración de contenidos, en el sentido de que los conocimientos de instantes anteriores se van acumulando y su incidencia crece con el tiempo mes a mes.

Considerando el estado inicial x_0 , que cumple el rol de intercepto en este modelo, aislamos su efecto mes a mes para ver como incide en la dinámica

$$\begin{aligned}
 x(1) &= Ax(0) + Bu(0) = \underbrace{Ax_0}_{\text{efecto } x_0 \text{ en } 0} + Bu(0) \\
 \implies x(2) &= \underbrace{A^2x_0}_{\text{efecto } x_0 \text{ en } 1} + Bu(1) + ABu(0) \\
 \dots \implies &\underbrace{A^{(t+1)}x_0}_{\text{efecto } x_0 \text{ en } t}
 \end{aligned}$$

Luego, el efecto que tiene el parámetro x_0 en la estimación del puntaje, será este con $t = 8$, pues cuando el estado llega a $t = 9$, se usa la información de los meses anteriores, ya sea con valores $u(t)$ o valores $\hat{u}(t)$ si se dispone de toda la información o se dispone solo hasta un t dado, según corresponda.

Por lo tanto el efecto concreto del parámetro x_0 sobre la estimación $\hat{x}(9)$ será de $A^9 x_0$, es decir de 262.9401.

Algo similar ocurre con los parámetro en las coordenadas 11 y 13 del vector $u(t)$, que corresponden al pretest y la variable del acuerdo de cada alumno con la afirmación “Matemáticas es fácil para mí”, pues estas solo tienen un valor distinto de 0 en $t = 0$, por lo que el efecto que interactúa en la estimación final de un parámetro B_j , con $j \in \{11, 13\}$ es

$$\begin{aligned} x(1) &= Ax(0) + Bu(0) = Ax(0) + \sum_{k=1}^{13} B_k u^{(k)}(0) = Ax_0 + \underbrace{B_j u^{(j)}(0)}_{\text{efecto } B_j \text{ en } 0} + \sum_{\substack{k=1 \\ k \neq j}}^{13} B_k u^{(k)}(0) \\ &\implies A^t \underbrace{B_j u^{(j)}(0)}_{\text{efecto } B_j \text{ en } t+1} \end{aligned}$$

Así el efecto del pretest en la estimación final es de $A^8 B_{11} u^{(11)}(0) = 1.0032 \cdot 26.2960 \cdot u^{(11)}(0) = 26.3803 \cdot p_i$, valor que está dentro del rango de valores observados para este efecto en los modelos lineales. Por otro lado, para la variable “Matemáticas es fácil para mí”, el efecto final será de $B_{13} u^{(13)}(0) = 1.0032 \cdot 3.3171 \cdot u^{(13)}(0) = 3.3277 \cdot m f_{i,j}$, valor que también es similar al obtenido para esta variable en el modelo con efecto olvido (con olvido).

El efecto de la variable “Mi comportamiento es un problema para el profesor de matemáticas” (media curso), requiere un análisis teniendo en cuenta el ajuste usado, que fue con la fórmula 3.10, que para el caso $A < 1$ corresponde a una curva de efecto decreciente hasta un instante $\bar{t} = 4$ y luego constante, pero en este caso, como $A > 1$, es una curva creciente hasta $\bar{t} = 4$ y luego con un crecimiento constante.

Esto nos indica que el efecto que acompaña $u^{(12)}(t)$ es

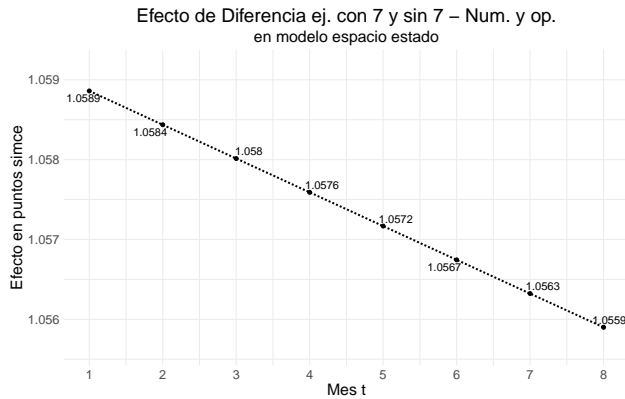
$$\begin{aligned} t = 1 &\implies B_{12} u^{(12)}(0) \\ &\vdots \\ t = 5 &\implies A^4 B_{12} u^{(12)}(0) \\ t = 6 &\implies A^5 B_{12} u^{(12)}(0) + B_{12} (A^4 (1 - A) u^{(12)}(0)) = A^4 B_{12} u^{(12)}(0) \end{aligned}$$

por lo tanto el efecto sobre la estimación en $t = 9$ será de $A^4 B_{12} u^{(12)}(0) = 1.0016 \cdot -13.2074 u^{(12)}(0) = -13.2285 \cdot com_{ij}$ y al igual que los efectos anteriores, este valor tampoco escapa de los esperado y observado en los modelos glotón y con olvido que comparten esta variable.

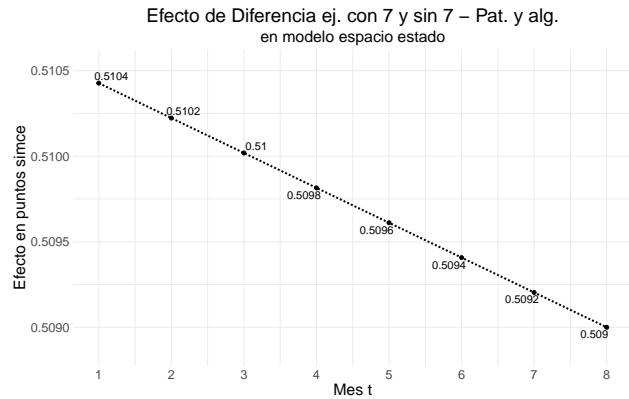
Las variables dinámicas, al ser variables que aparecen cada mes, generan un efecto acumulado, y cada una de estas tendrá un efecto de

$$\begin{aligned}
\sum_{t=1}^8 A^{8-t} B_j u^{(j)}(t) &= \underbrace{A^7 B_j u^{(j)}(1)}_{\text{efecto de } t=1} + \underbrace{A^6 B_j u^{(j)}(2)}_{\text{efecto de } t=2} + \underbrace{A^5 B_j u^{(j)}(3)}_{\text{efecto de } t=3} + \underbrace{A^4 B_j u^{(j)}(4)}_{\text{efecto de } t=4} \\
&+ \underbrace{A^3 B_j u^{(j)}(5)}_{\text{efecto de } t=5} + \underbrace{A^2 B_j u^{(j)}(6)}_{\text{efecto de } t=6} + \underbrace{A B_j u^{(j)}(7)}_{\text{efecto de } t=7} + \underbrace{B_j u^{(j)}(8)}_{\text{efecto de } t=8}
\end{aligned}$$

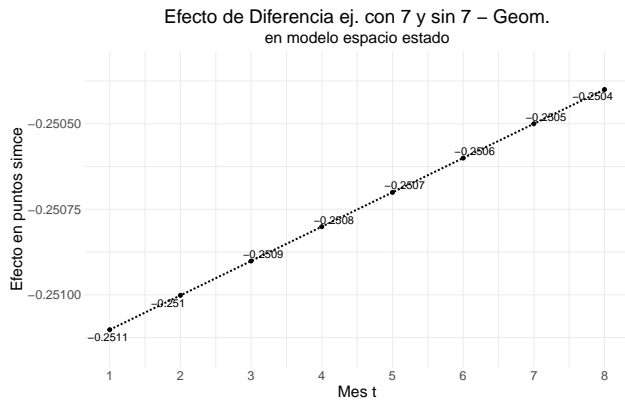
Con esto se podría graficar cómo la diferencia entre el número de ejercicios resueltos en el primer intento y los que no en cada mes tiene una incidencia en la estimación final de puntaje, como se muestra a continuación.



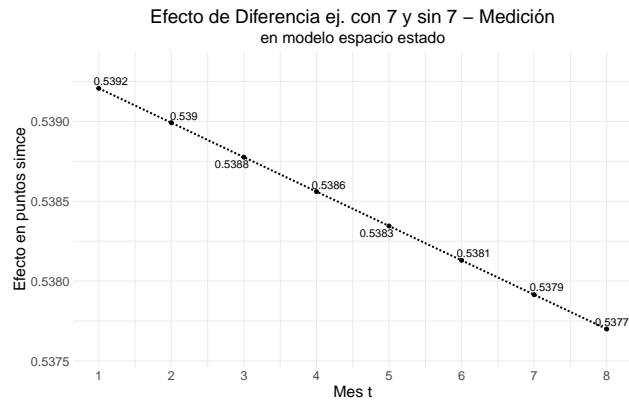
(a) Efecto $A^{8-t}B_1$



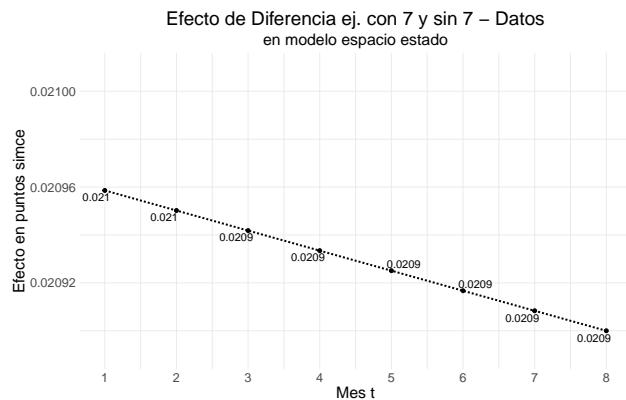
(b) Efecto $A^{8-t}B_2$



(c) Efecto $A^{8-t}B_3$



(d) Efecto $A^{8-t}B_4$



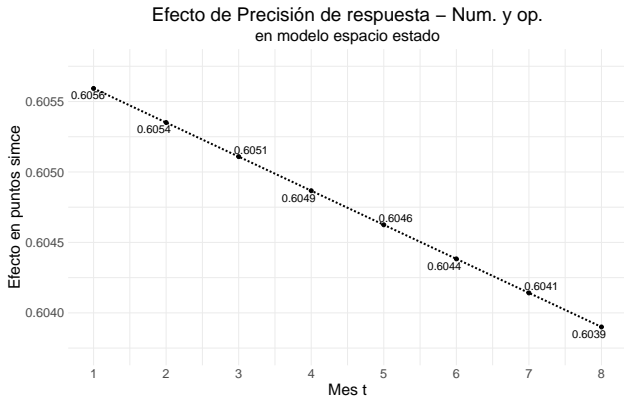
(e) Efecto $A^{8-t}B_5$

donde las curvas son decrecientes en valor absoluto pues al ser $A > 1$, los primeros meses del año acumularán un mayor efecto en la medida que el sistema va avanzando en t . Es decir, los primeros meses tendrán un efecto más positivo o más negativo dependiendo del signo.

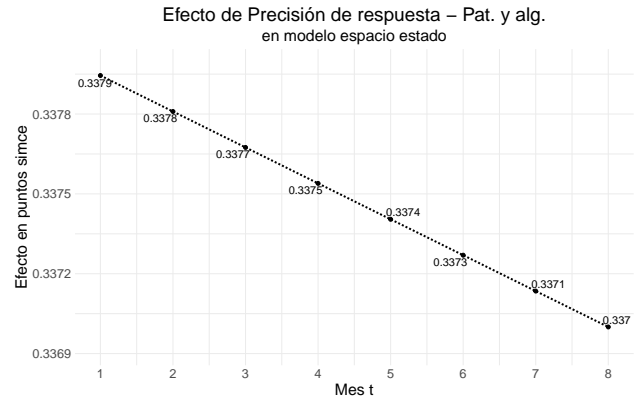
Este caso es el completo opuesto del modelo olvido (con olvido), en este caso se podría considerar el parámetro A como un efecto de maduración en lugar de un efecto olvido, es decir que indica cuánto va aumentando la comprensión de los contenidos en la medida que pasa el tiempo, y por esto el efecto de los primeros instantes va aumentando en la medida que aumenta la cantidad de periodos entre la ejercitación (input $u(t)$) y la medición (otuput

$y(t)$.

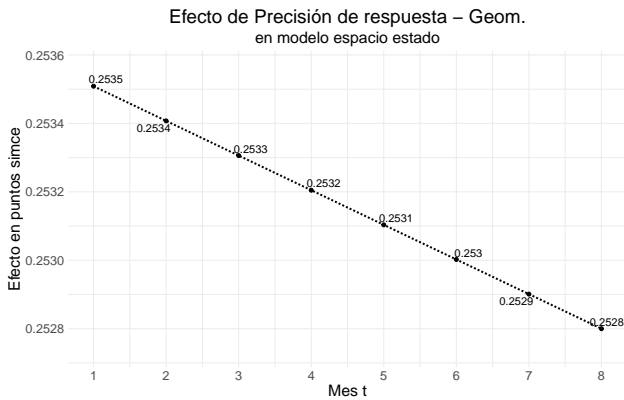
Similarmente, si observamos los efecto de la precisión de respuesta obtenemos lo siguiente



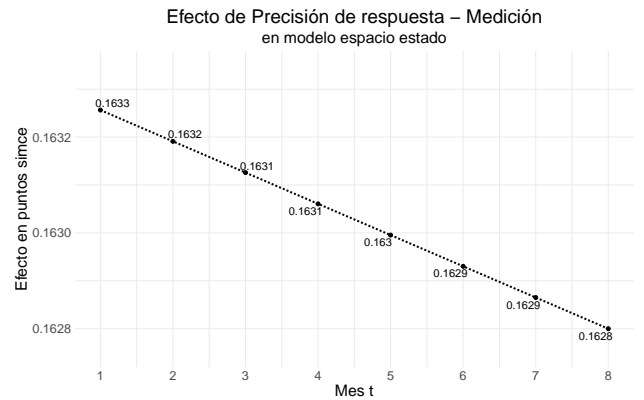
(f) Efecto $A^{8-t}B_6$



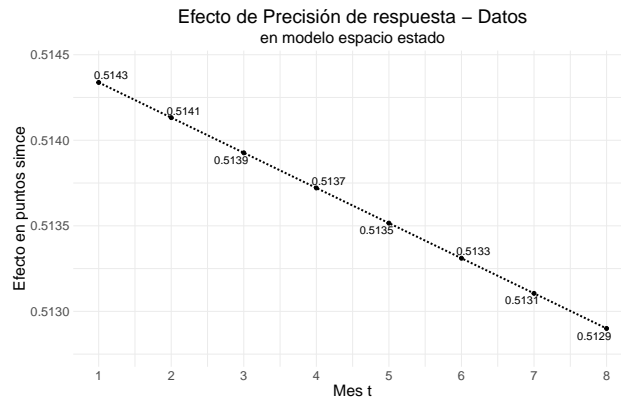
(g) Efecto $A^{8-t}B_7$



(h) Efecto $A^{8-t}B_8$



(i) Efecto $A^{8-t}B_9$



(j) Efecto $A^{8-t}B_{10}$

3.5. Resumen

El modelo espacio estado, con 17 parámetros alcanza un RMSE de 27.52, esto es, a 10.96 puntos del SEM del SIMCE, es decir un modelo que reduce un 64.9% del error explicable.

Usando el pretest, las variables históricas del promedio curso del nivel de acuerdo con la afirmación “Mi comportamiento es un problema para el profesor de matemática” y el nivel de acuerdo de cada alumno con la afirmación “Matemáticas es fácil para mí”. Por su parte, las variables dinámicas usadas son la precisión de respuesta en el primer intento y la diferencia entre el número de ejercicios resueltos en el primer intento y los resueltos en más de un intento.

Capítulo 4

Comparación de los distintos modelos

Cada uno de los modelos presentados en esta tesis tiene sus ventajas y desventajas. En este capítulo analizaremos los principales aspectos de cada uno de estos modelos. Recordemos que estos aspectos son:

1. RMSE en entrenamiento y testeo, como una medida de la capacidad predictiva de cada modelo.
2. Flexibilidad de cada modelo para realizar estimaciones fuera de la muestra.
3. Explicabilidad de cada modelo.
4. Costo de predicción, en términos de tiempo y recursos computacionales.

Antes de analizar estos aspectos, recordamos el número de parámetros que usa cada modelo, las variables históricas y dinámicas que usan, junto con mencionar algunos de los aspectos importantes de cada modelo.

Modelo	Núm. de parám.	Variables Históricas	Variables Dinámicas
sin memoria	65	Simce histórico - media curso	Componentes Principales
		“Me va bien en matemáticas” (respuesta por alumno)	
glotón	385	“Mi comportamiento es un problema para el profesor de matemáticas” (media curso)	Precisión de respuesta en el primer intento
		Nota a torneos espiral (respuesta por alumno)	Promedio de notas en preguntas abiertas
con memoria	105	Simce histórico - media curso	Precisión de respuesta en el primer intento
		“Mis padres dicen que soy buen estudiante” - media curso	Promedio de notas en preguntas abiertas
con olvido	66	“Mi comportamiento es un problema para el profesor de matemáticas” (media curso)	Precisión de respuesta en el primer intento
		”Matemáticas es fácil para mi” (respuesta por alumno)	
espacio estado	17	“Mi comportamiento es un problema para el profesor de matemáticas” (media curso)	Precisión de respuesta en el primer intento
		”Matemáticas es fácil para mi” (respuesta por alumno)	Dif. entre n. de preg. resueltas en un intento y más de un intento

Tabla 4.1: Resumen de los modelos

4.1. RMSE en entrenamiento y testeo

Un aspecto primordial para evaluar cada modelo usado para estimar puntajes en la prueba SIMCE es cuánto se acercan sus predicciones al puntaje real obtenido por los alumnos. En el capítulo anterior, analizamos dicha capacidad y los resultados se pueden revisar en las tablas 2.6, 2.10, 2.21, 2.21 y 3.2. Allí es fácilmente notable que el modelo con efecto olvido (con olvido) tiene el RMSE más bajo de todos; sin embargo, este modelo requiere de 66 parámetros.

El segundo lugar lo disputan el modelo lineal con variables de 1 hasta t (con memoria) y el modelo espacio estado (espacio estado). El primero obtiene el RMSE más bajo para el caso de información hasta $t = 8$. Por su parte, el segundo, si bien está un par de centésimas por encima, obtiene mejores predicciones en instantes anteriores, pues tiene un RMSE de menos de 30 puntos en testeo a partir de $t = 4$, mientras que para el modelo lineal esto ocurre en $t = 5$. En cuanto al número de parámetros, el modelo con memoria hace uso de 105 parámetros; en cambio, el modelo espacio estado requiere solo 17.

Luego siguen los modelos sin memoria y glotón. No obstante, el RMSE de estos en $t = 8$ para testeo no baja de los 28 puntos. Además, tienen un número igual o mayor de parámetros que los modelos anteriores, lo cual los hace menos convenientes.

Esto se puede apreciar en las siguientes figuras del RMSE de estimaciones del puntaje de cada alumno y el RMSE para estimaciones del puntaje promedio del curso.

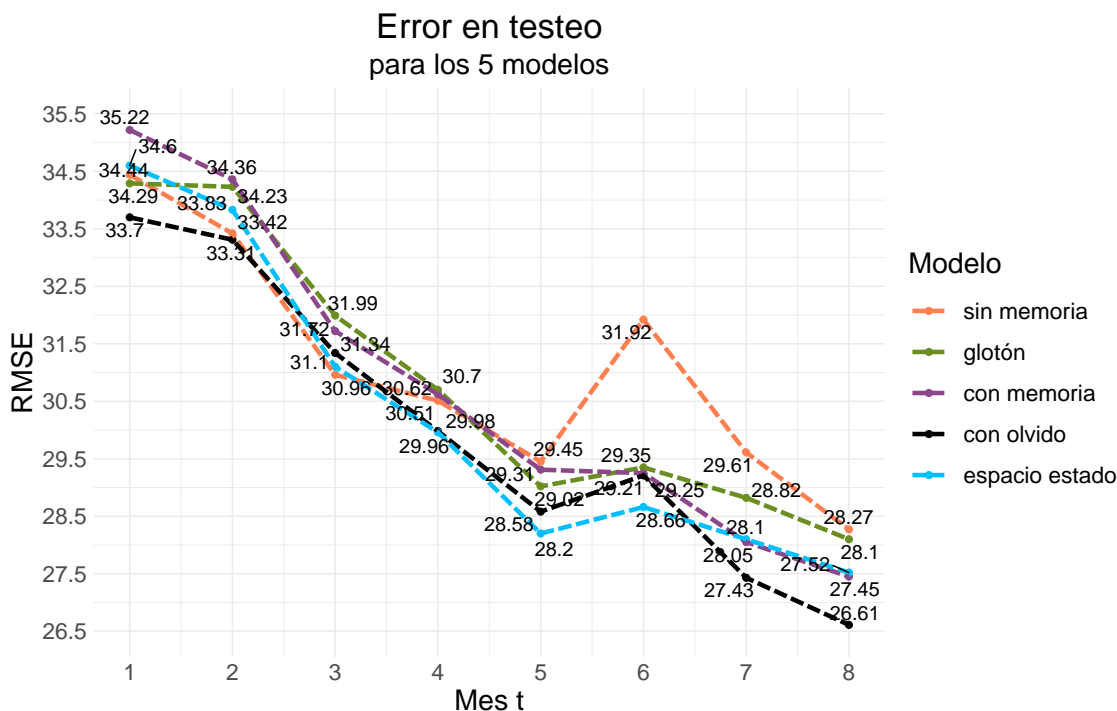


Figura 4.1: RMSE para estimaciones de puntaje para los 5 modelos en el grupo de testeo

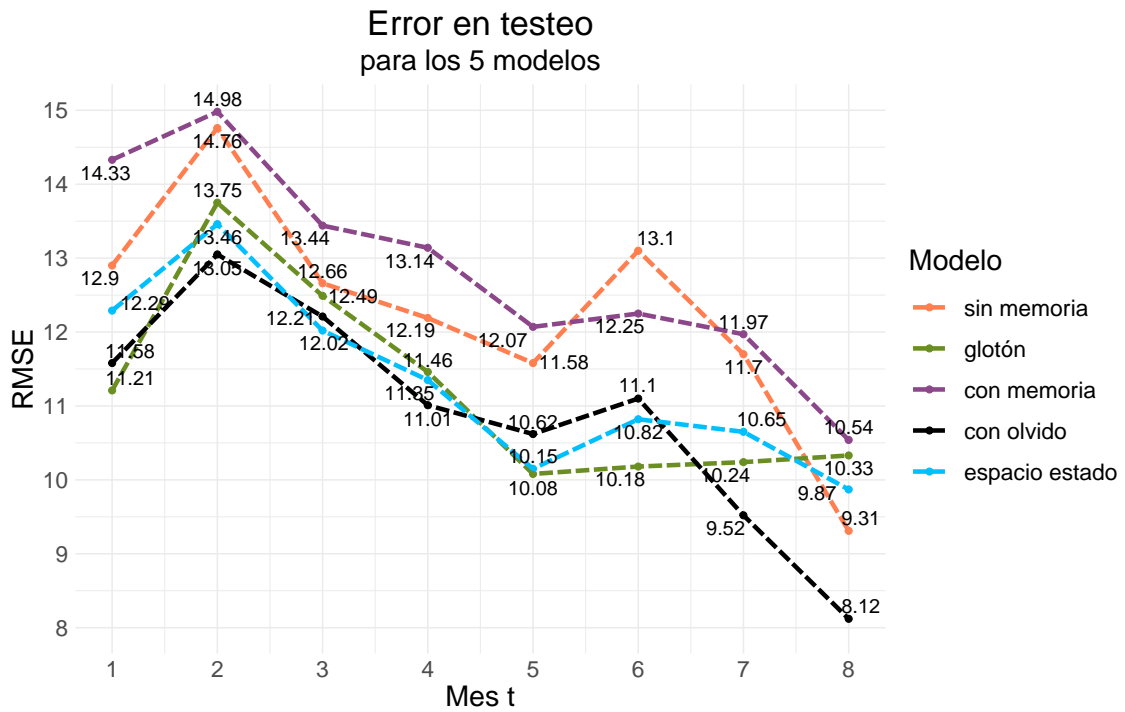


Figura 4.2: RMSE para estimaciones del promedio curso de puntaje para los 5 modelos en el grupo de testeo

Más aún, si descontamos las cotas del SEM para las estimaciones por alumno, que usaremos 16.56 según el cálculo en 1.1 y para la estimaciones en el puntaje promedio del curso, descontamos un SEM de 7.26 en base a lo calculado en 1.2.

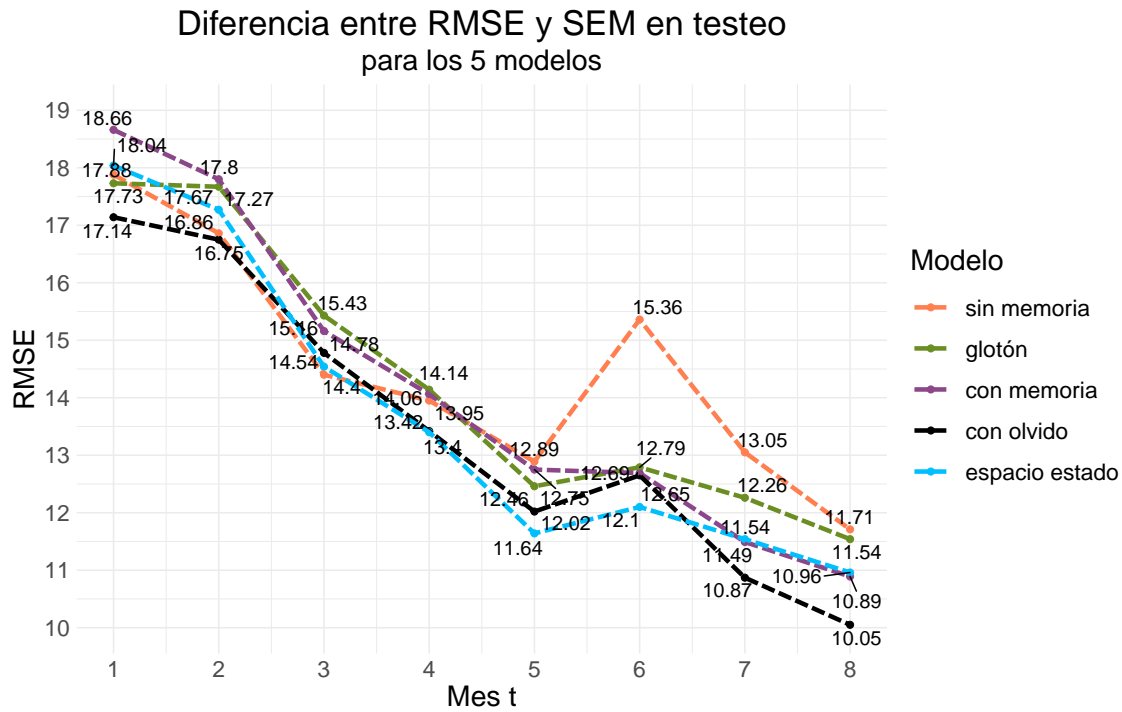


Figura 4.3: Diferencia del RMSE y SEM para estimaciones de puntaje para los 5 modelos en el grupo de testeo

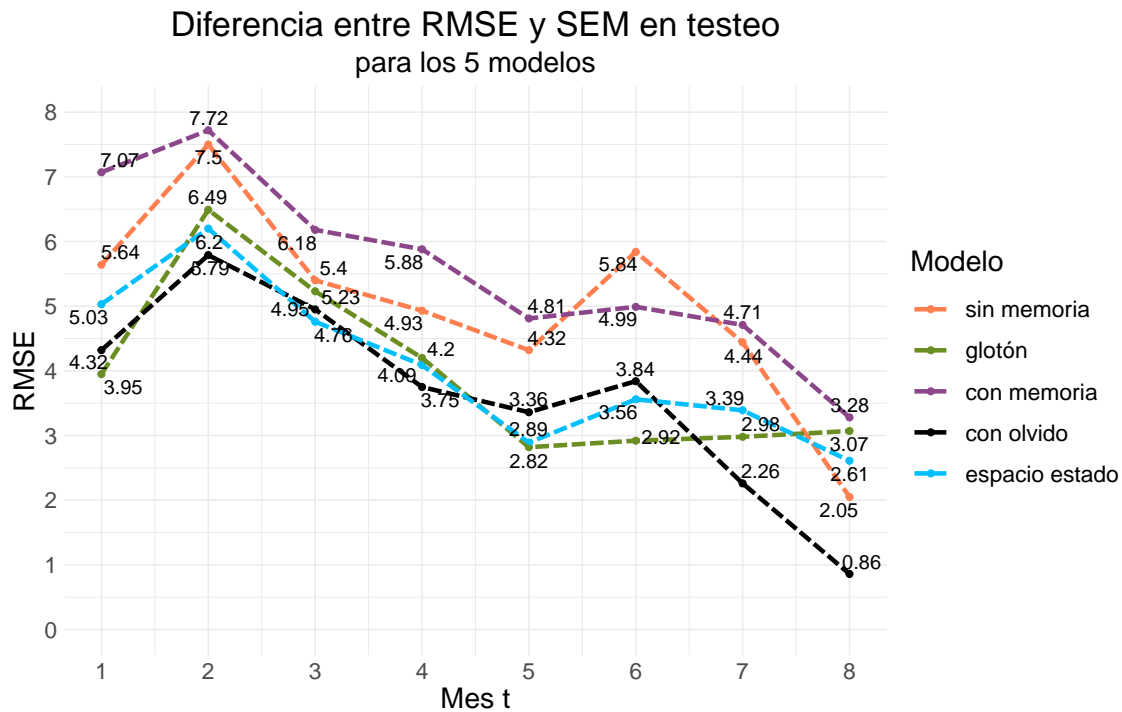
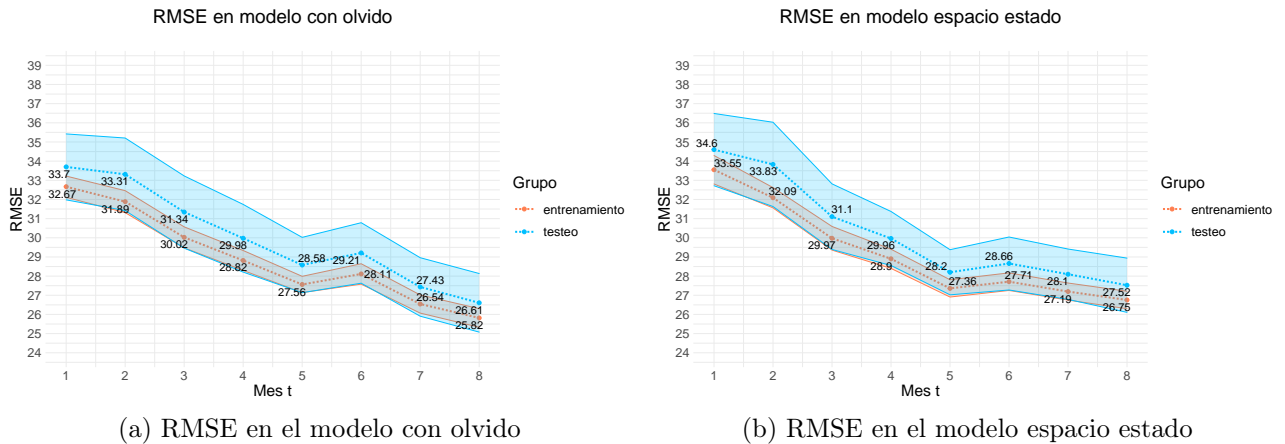


Figura 4.4: Diferencia del RMSE y SEM para estimaciones del promedio curso de puntaje para los 5 modelos en el grupo de testeo

En resumen, el modelo con mejor desempeño fue el modelo con olvido. Sin embargo, si se

considera la cantidad de parámetros que se requieren, resulta más práctico el modelo espacio estado, que necesita 89 parámetros menos que el modelo con olvido. Pese a ello, el RMSE del modelo espacio estado, en los últimos 2 meses, está por encima del RMSE del modelo con olvido por poco menos de 1 punto. Esto se puede apreciar en las figuras 2.32 y 3.1, que se muestran una vez más a continuación:



También se puede apreciar esta ligera diferencia en la siguiente tabla en la que se compara el RMSE en testeo de estos dos modelos.

Meses	1	2	3	4	5	6	7	8
RMSE promedio modelo con olvido	33.70	33.31	31.34	29.98	28.58	29.21	27.43	26.61
RMSE promedio modelo espacio estado	34.60	33.83	31.10	29.96	28.20	28.66	28.10	27.52
Diferencia entre modelos	-0.90	-0.52	0.24	0.02	0.38	0.55	-0.67	-0.91
Desviación RMSE modelo con olvido	1.72	1.89	1.88	1.77	1.44	1.58	1.53	1.53
Desviación RMSE modelo espacio estado	1.89	2.20	1.71	1.41	1.18	1.38	1.31	1.42

Tabla 4.2: RMSE promedio y desviación del RMSE para testeo en modelos con olvido y espacio estado

Por lo tanto, según sea el caso, se podría optar por cualquiera de estos modelos. El primero ofrece un menor RMSE y el segundo necesita menos parámetros y es casi tan bueno como el primero.

4.2. Flexibilidad

En este trabajo, con flexibilidad nos referimos a la capacidad de un modelo para funcionar en nuevas instancias o variaciones del programa.

Los 5 modelos propuestos deberían ser aplicables a datos fuera de la muestra, puesto que no se trabajó con efectos fijos o variables que estén limitadas a la instancia en particular del programa realizado en 2017; la única excepción es el *pretest* que se realizó ese año y que fue usado como regresor en todos los modelos. Por esto, se podría aplicar cualquiera de los modelos a nuevas instancias, como las de 2018 en adelante, siempre que exista un *pretest*.

En cuanto a la conveniencia de usar cada modelo en futuras instancias, no es igual para todos los casos. Es evidente que los modelos 2.6 y con olvido tienen muchos parámetros que dependen de cómo se desarrolló la ejercitación por eje mes a mes. En consecuencia, si ocurren cambios en el orden en que se estudian los ejes de los contenidos o la cantidad de ejercicios que se realizan mes a mes, es posible que los errores en estimaciones fuera de la muestra se vean acentuados. Si bien la ejercitación entre un curso y otro fue diferente en 2017, sí existió una tendencia general a priorizar unos ejes por sobre otros y también estos se abordaron en el mismo orden durante el año, según las directrices del programa que entregó el Ministerio de Educación.

Por su parte, el modelo espacio estado espacio estado, al no tener parámetros que dependan del mes, permite un mayor grado de flexibilidad; por lo que tendría un mejor desempeño que otros modelos en situaciones en las que el cronograma de contenidos cambie. Esto también se debe a la menor cantidad de parámetros que tiene este modelo; en el caso de los otros modelos, al emplear más parámetros, aumenta la probabilidad de que uno de ellos no se ajuste bien en años posteriores.

Un aspecto que no se ha mencionado es la ventaja del modelo espacio estado como una estructura de control óptimo. Esta ventaja se manifiesta en que, una vez identificados los parámetros de este modelo, se puede plantear un problema de control óptimo y estudiar una estructura óptima o ideal de ejercitación con base en los valores que tendría un $u(t)$ óptimo.

Por todo esto, el modelo que ofrece un mayor grado de flexibilidad es el modelo espacio estado, seguido por el modelo con olvido y luego el resto. El menos flexible es el modelo glotón, pues mostró un mayor grado de sobre ajuste y emplea muchos más parámetros que los demás.

Una característica de flexibilidad que comparten los 5 modelos es que no dependen del conjunto de ejercicios realizados por los alumnos, es decir, los modelos son ciegos a qué preguntas respondieron los alumnos. En ese sentido, para el modelo, dos alumnos que hayan respondido ejercicios diferentes, pero hayan contestado la misma cantidad de ejercicios de cada eje y mostrado el mismo rendimiento en estos, son indistintos. En primera instancia, esto puede considerarse una ventaja: la única información que los modelos requieren de cada pregunta es el eje de contenidos asociado y la nota asignada por el sistema (si se conoce la nota, también se sabe si fue resuelto en uno o más intentos). Por esta razón, es posible implementar los modelos para predecir el rendimiento en otros años en que se lleve a cabo el programa, aún cuando las preguntas sean distintas.

Por otro lado, esta característica puede también considerarse un campo idóneo para buscar modelos que arrojen mejores predicciones. Una forma en la que esto se podría hacer implicaría tomar en cuenta la dificultad de cada pregunta en función de las respuestas recopiladas en la plataforma, ya sea que simplemente se ponderen las preguntas de forma distinta, según el

porcentaje de veces que se contestó correctamente en el primer intento, o que se usen técnicas más complejas, como modelos IRT.

4.3. Explicabilidad

El concepto de explicabilidad suele ser obviado en muchos estudios y, en especial, debido a la prevalencia de modelos de *machine learning* con estructuras de caja negra y/o caja gris (que solo una parte del modelo es la que actúa como caja negra). Estos modelos suelen entregar buenas predicciones y esto los ha hecho sumamente populares, pero tienen la característica de ser mucho más complejos que, por ejemplo, un modelo lineal, que es bastante fácil de entender sin necesidad de conocimientos especializados en el área de probabilidades y estadística. Los modelos analizados en esta tesis se plantearon con el objetivo de construir modelos con un poder predictivo que supere los *baselines* lineales habituales (2.7) sin sacrificar la simpleza de un modelo lineal.

Los 5 modelos propuestos son, *a priori*, un poco más complejos que los *baselines*. Los primeros 4 incluyen una estructura de sumatoria, pero esta estructura se puede reescribir como un modelo lineal multivariado en su forma estándar $y = \beta X + \varepsilon$. Sin embargo, este planteamiento oculta las variables que están interactuando en el modelo y cómo estas se relacionan entre sí. Por esto, a pesar de que visualmente parezcan más complejos que un modelo lineal habitual, consideramos que no lo son; se encuentran en el mismo nivel puesto que la supuesta complejidad se debe simplemente a la notación usada.

En cuanto al quinto modelo, el de espacio estado, este es más complejo que los modelos lineales en solo dos aspectos. En primer lugar, este modelo tiene una estructura de serie temporal, pero se puede visualizar como un modelo lineal entre dos meses y por lo tanto, para un lector que ya está familiarizado con modelos lineales, esta estructura no resulta ajena y debería ser fácil de incorporar y entender. En segundo lugar, hay que considerar la existencia de dos estados, uno latente u oculto y uno observable, y la forma en que estos interactúan. Por fortuna, esto se simplifica al momento de considerar que las matrices C y D del modelo espacio estado son I (la identidad) y 0 respectivamente. De ese modo, el estado observable se convierte una medición con ruido del estado latente, por lo que tampoco debería reflejar un mayor desafío para su fácil interpretación.

4.4. Costo de predicción

Con miras a aplicar estos modelos a gran escala, vale la pena analizar el costo computacional de cada uno de ellos. Cuando el número de alumnos n es bastante grande, el uso de herramientas más sofisticadas aumenta los costos de manera exponencial, pues requieren más tiempo y unidades de procesamiento más potentes.

Para los fines de esta comparación, se obviarán cálculos que son necesarios para todos los modelos por igual, por ejemplo, la normalización de los datos. Nos centraremos en aquellas operaciones de mayor complejidad o que requieren un mayor espacio en la memoria. La memoria la medimos en el número de entradas de datos requeridas (siendo cada entrada un valor flotante, es decir, 4 bytes). Aclarado esto, procedemos a analizar cada uno de los 5

modelos estudiados en este trabajo.

4.4.1. Modelo sin memoria

Debido a que este modelo realiza predicciones mensuales, necesita re-escalar las 10 variables dinámicas para hacer uso de las componentes principales (ver tabla 2.5). Si bien, la estructura de estas componentes principales se considera parte de la construcción del modelo, es necesario transformar los nuevos datos para la predicción de acuerdo con la estructura de las 5 componentes a utilizar.

En términos de memoria, para cada mes t se requiere del *pretest*, las 2 variables históricas, y las 10 variables dinámicas en sus 5 ejes, es decir, una matriz de $n \times 53$.

Por su parte, en términos de cómputo, será necesario calcular las 5 componentes, lo cual significa multiplicar las 50 variables dinámicas por la matriz de rotación del ACP (de 50×5). Entonces, para esta operación se necesita llevar a cabo $n \cdot 5 \cdot 50 = 250n$ multiplicaciones y $n \cdot 5 \cdot 49 = 245n$ adiciones. Luego, para cada alumno se realiza el producto de los efectos por cada una de las variables y luego se suman, esto es $8n$ multiplicaciones y $8n$ sumas.

Entonces como esto se repite para cada mes t en 1 a T meses, este modelo requiere:

	en t	total de 1 a T
Memoria	$53n$	$53nT$
Sumas	$253n$	$253nT$
Multiplicaciones	$258n$	$258nT$

Tabla 4.3: Costos de memoria y cómputo de modelo sin memoria

Esto nos dice que el orden de operaciones y memoria requerido para este modelo son $O(n)$ para cada mes, y $O(nT)$ para todos los periodos.

4.4.2. Modelo glotón

El segundo modelo hace uso del *pretest*, 2 variables históricas y 2 dinámicas, pero para cada mes, aumenta el número de parámetros usados y, al mismo tiempo, el número de variables usadas. Por lo tanto, es importante considerar la dependencia del número de meses.

Para un mes t dado, se necesita tener $3n + 10nt$ datos: $3n$ por las variables históricas y el *pretest*, mientras que por las variables dinámicas en sus cinco ejes, se necesita $10n$ datos para cada mes t .

En cuanto a operaciones, es necesario llevar a cabo $3n + 10nt$ multiplicaciones por el mismo motivo explicado antes. Asimismo, la cantidad de sumas requeridas es del mismo valor. Entonces para un total de T meses, sumamos para todos los instantes t .

$$\sum_{t=1}^T 3n + 10nt = 3nT + 10n \left(\sum_{t=1}^T t \right) = 3nT + 5nT(T + 1) = 5nT^2 + 8nT$$

Con esto, se necesita:

	en t	total de 1 a T
Memoria	$3n + 10nt$	$5nT^2 + 8nT$
Sumas	$3n + 10nt$	$5nT^2 + 8nT$
Multiplicaciones	$3n + 10nt$	$5nT^2 + 8nT$

Tabla 4.4: Costos de memoria y cómputo de modelo glotón

Luego, tanto para memoria como para operaciones de adición y producto, para cada mes t , el orden requerido es $O(nt)$, mientras que para operar todo el modelo se requiere un total de orden $O(nT^2)$.

4.4.3. Modelo con memoria

Este modelo también tiene dos variables dinámicas, y al considerar parámetros de 1 hasta t reduce la cantidad de parámetros usados en los casos anteriores, pero para cada mes de todos modos se requerirá de $3n + 10nt$ datos, pues es necesario llamar todos los datos desde el instante 1 hasta t y luego operar sobre estos para reducirlos a las variables en su forma acumulada.

Las operaciones necesarias son la suma o promedio hasta t para cada variable dinámica y para cada eje. Es decir que para un mes t es necesario realizar $10n(t - 1)$ sumas. En este caso al ser ambas variables que requieren calcular un promedio, esto significa un producto por $\frac{1}{t}$, es decir $10n$ productos para todo $t > 1$. Luego, puesto que se usan estas variables junto con las variables históricas para la estimación en el modelo lineal, se realizan $13n$ multiplicaciones y $13n$ sumas.

Entonces para un total de T meses se tiene:

$$\begin{aligned} \sum_{t=1}^T 3n + 10nt &= 5nT^2 + 8nT \\ \sum_{t=1}^T 10n(t - 1) + 13n &= \sum_{t=1}^T 3n + 10nt = 5nT^2 + 8nT \\ \sum_{t=1}^T 10n\mathbb{1}_{t>1} + 13n &= 10n(T - 1) + 13n = 23nT - 10n \end{aligned}$$

En resumen, para hacer las estimaciones entre los meses 1 y T es necesario

	en t	total de 1 a T
Memoria	$3n + 10nt$	$5nT^2 + 8nT$
Sumas	$10n(t - 1) + 13n$	$5nT^2 + 8nT$
Multiplicaciones	$10n\mathbb{1}_{t>1} + 13n$	$23nT - 10n$

Tabla 4.5: Costos de memoria y cómputo de modelo con memoria

En este caso, tanto para memoria como para adiciones en t , se requiere un orden $O(nt)$, mientras para multiplicaciones, el orden es $O(n)$, por esto, el total de operaciones de suma y memoria son de orden $O(nT^2)$ y para productos, son $O(nT)$.

4.4.4. Modelo con olvido

El modelo con efecto de olvido ρ es similar al anterior, en el sentido en que se requiere la misma cantidad de datos para las variables. Sin embargo, este modelo usa solo una variable dinámica mientras que el anterior usa dos; por eso, para un mes t usa $3n + 5nt$ datos.

Respecto a las operaciones realizadas, se emplea la misma cantidad de sumas, pero aumenta la cantidad de multiplicaciones, pues ahora se tiene una variable dinámica en lugar de dos. De este modo, este modelo requiere en un mes t , $5n(t - 1) + 8n$ sumas, en las que se suman elementos ponderados por distintas potencias de ρ , lo que significa que se realizan $5nt + 8n$ multiplicaciones. Luego al promediar, se realiza $5n$ multiplicaciones más para todo $t > 1$.

Con esto, para T se tiene:

$$\begin{aligned} \sum_{t=1}^T 3n + 5nt &= 3nT + \frac{5nT(T+1)}{2} = \frac{5nT^2 + 11nT}{2} \\ \sum_{t=1}^T 5n(t-1) + 8n &= \sum_{t=1}^T 3n + 5nt = \frac{5nT^2 + 11nT}{2} \\ \sum_{t=1}^T 5n\mathbb{1}_{t>1} + 5nt + 8n &= 5n(T-1) + \frac{5nT(T+1)}{2} + 8nT = \frac{5nT^2 + 31nT - 10n}{2} \end{aligned}$$

Entonces, este modelo en T meses usa

	en t	total de 1 a T
Memoria	$3n + 5nt$	$\frac{5nT^2 + 11nT}{2}$
Sumas	$5n(t-1) + 8n$	$\frac{5nT^2 + 11nT}{2}$
Multiplicaciones	$5n\mathbf{1}_{t>1} + 5nt + 8n$	$\frac{5nT^2 + 31nT - 10n}{2}$

Tabla 4.6: Costos de memoria y cómputo de modelo con olvido

4.4.5. Modelo espacio estado

Por último, el modelo espacio estado requiere de las variables históricas y el *pretest* tan solo en la predicción del primer mes ($t = 1$). Por lo tanto, para una estimación cuando $t = 1$, se usan $3n$ entradas por estas variables y $10n$ por las variables dinámicas del primer mes. En cambio, para $t > 1$, solo importará estas $10n$ entradas y n entradas de $x(t-1)$.

En cuanto a operaciones, se necesita de $3n$ multiplicaciones para las variables históricas cuando $t = 1$ y $10n$ para las variables dinámicas en cada t . Además, consideramos el producto de $Ax(t-1)$ en cada caso, esto es, n productos más. Las sumas realizadas son $3n$ para las variables históricas y x_0 cuando $t = 1$, y son $10n$ para las variables dinámicas en $Ax(t) + Bu(t)$.

El total de memoria y operaciones requeridos desde el inicio hasta un instante T es:

$$\begin{aligned} \sum_{t=1}^T 11n + 2n\mathbf{1}_{t=1} &= 11nT + 2n = 11nT + 2n \\ \sum_{t=1}^T 10n + 3n\mathbf{1}_{t=1} &= 10nT + 3n = 10nT + 3n \\ \sum_{t=1}^T 11n + 4n\mathbf{1}_{t=1} &= 11nT + 4n = 11nT + 4n \end{aligned}$$

En resumen, un modelo que usa información hasta T empleará:

	en t	total de 1 a T
Memoria	$11n + 2n\mathbf{1}_{t=1}$	$11nT + 2n$
Sumas	$10n + 3n\mathbf{1}_{t=1}$	$10nT + 3n$
Multiplicaciones	$11n + 4n\mathbf{1}_{t=1}$	$11nT + 4n$

Tabla 4.7: Costos de memoria y cómputo de modelo espacio estado

Con esto, todas las operaciones en t son de orden $O(n)$ y para el total hasta T , son de orden $O(nT)$.

4.4.6. Resumen

En resumen, en la siguiente tabla se indica la memoria y operaciones necesarias para cada modelo para un t dado

Modelo	sin memoria	glotón	con memoria	con olvido	espacio estado
Memoria	$53nT$	$3n + 10nt$	$3n + 10nt$	$3n + 5nt$	$11n + 2n\mathbb{1}_{t=1}$
Multiplicaciones	$258nT$	$3n + 10nt$	$3n + 10nt$	$3n + 5nt$	$10n + 3n\mathbb{1}_{t=1}$
Sumas	$253nT$	$3n + 10nt$	$13n + 10n\mathbb{1}_{t>1}$	$8n + 5nt + 5n\mathbb{1}_{t>1}$	$11n + 4n\mathbb{1}_{t=1}$

Tabla 4.8: Costo en memoria y número de operaciones requeridas para cada modelo en t

Esto, en términos del orden de estos valores, es

Modelo	sin memoria	glotón	con memoria	con olvido	espacio estado
Memoria	$O(n)$	$O(nt)$	$O(nt)$	$O(nt)$	$O(n)$
Multiplicaciones	$O(n)$	$O(nt)$	$O(nt)$	$O(nt)$	$O(n)$
Sumas	$O(n)$	$O(nt)$	$O(n)$	$O(nt)$	$O(n)$

Tabla 4.9: Orden de memoria y número de operaciones requeridas para cada modelo en t

y el total de memoria y operaciones para T periodos es:

Modelo	sin memoria	glotón	con memoria	con olvido	espacio estado
Memoria	$53nT$	$5nT^2 + 8nT$	$5nT^2 + 8nT$	$\frac{5nT^2 + 11nT}{2}$	$11nT + 2n$
Sumas	$258nT$	$5nT^2 + 8nT$	$5nT^2 + 8nT$	$\frac{5nT^2 + 11nT}{2}$	$10nT + 3n$
Multiplicaciones	$253nT$	$5nT^2 + 8nT$	$23nT - 10n$	$\frac{5nT^2 + 31nT - 10n}{2}$	$11nT + 4n$

Tabla 4.10: Costo en memoria y número de operaciones requeridas para cada modelo hasta T

lo que en términos de orden, es:

Modelo	sin memoria	glotón	con memoria	con olvido	espacio estado
Memoria	$O(nT)$	$O(nT^2)$	$O(nT^2)$	$O(nT^2)$	$O(nT)$
Multiplicaciones	$O(nT)$	$O(nT^2)$	$O(nT^2)$	$O(nT^2)$	$O(nT)$
Sumas	$O(nT)$	$O(nT^2)$	$O(nT)$	$O(nT^2)$	$O(nT)$

Tabla 4.11: Orden de memoria y número de operaciones requeridas para cada modelo hasta T

Para ejemplificar esto mejor, consideraremos cómo la memoria y número de operaciones aumenta al aumentar el número de alumnos a los que se busca estimar el puntaje en un mes t . Para esto, consideraremos $t = 8$, es decir con la información hasta octubre, justo antes del SIMCE.

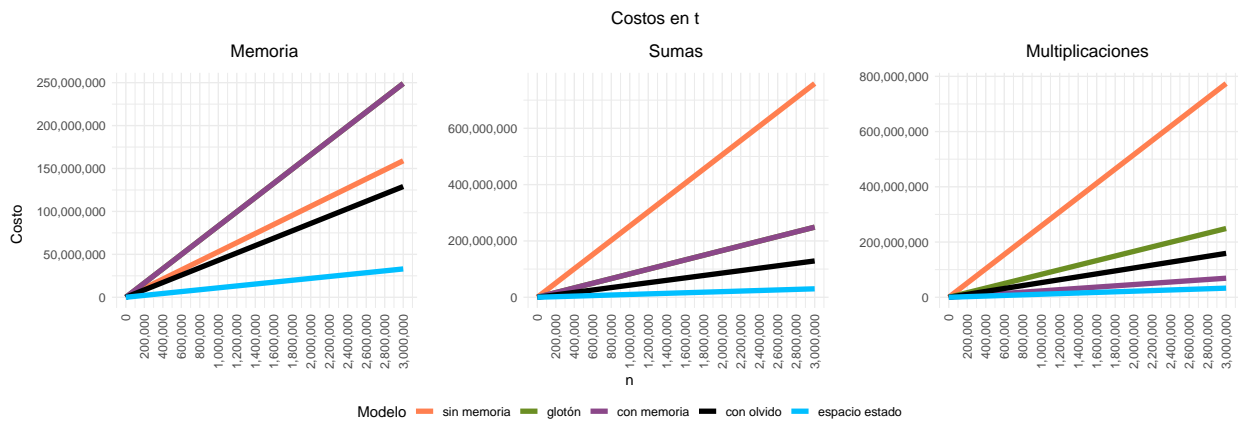


Figura 4.5: Costos en $t = 8$ para distinto número n de alumnos

y para el total acumulado entre los meses 1 y $T = 8$, se tiene algo similar

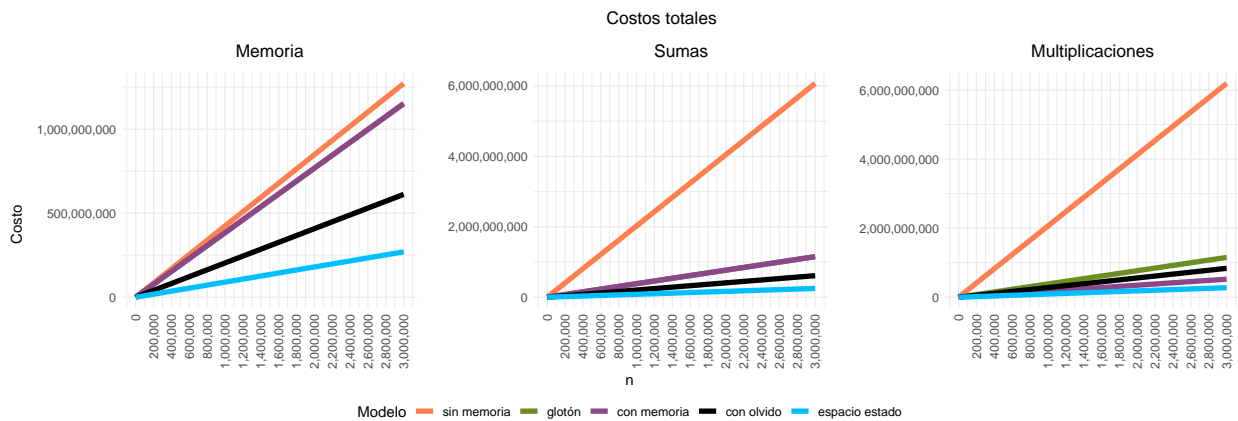


Figura 4.6: Costos totales a $T = 8$ para distinto número n de alumnos

De igual modo se puede hacer el ejercicio de ver qué ocurre cuando cambia el número de periodos con los que estos modelos trabajan. Por ejemplo, si quisiéramos trabajar con un modelo semana a semana, en el periodo de marzo a octubre en que se realizó el programa, hay alrededor de 30 semanas. Esta pequeña variación generará grandes diferencias entre los modelos, puesto que muchas operaciones y memoria dependen cuadráticamente del número de instantes T .

Para graficar esto consideraremos t el número de semanas, que irá de 1 a 30, y usaremos $n = 100,000$ fijo.

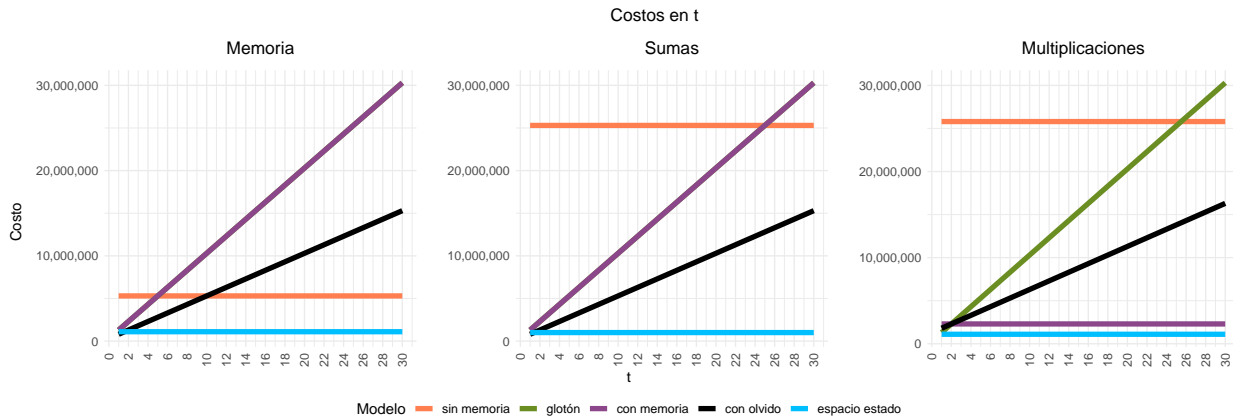


Figura 4.7: Costos en t para $n = 100,000$ de alumnos

y de igual modo se puede ver qué ocurre al aumentar el total de instantes T :

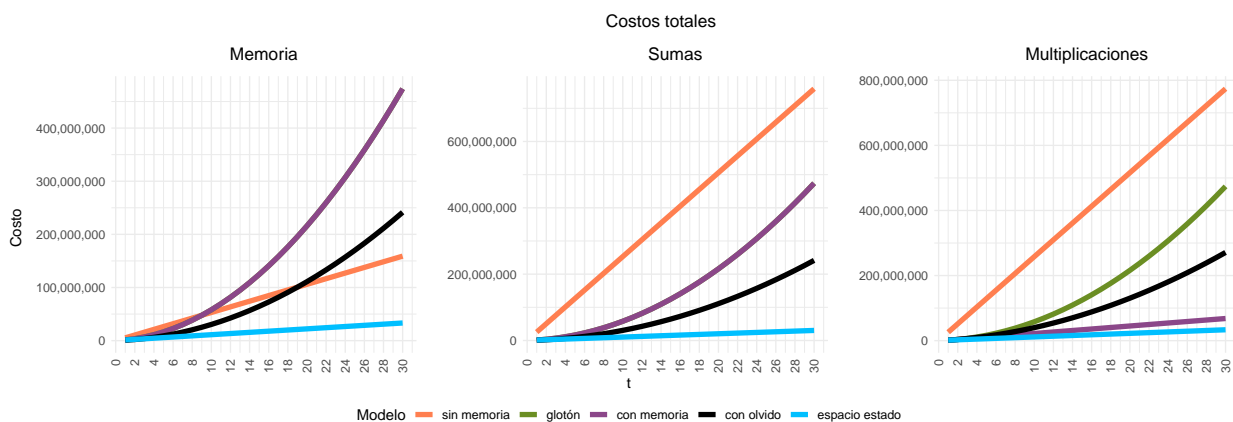


Figura 4.8: Costos totales a T para $n = 100,000$ de alumnos

Entonces, ya sea si aumenta la cantidad n de alumnos o el número de instantes T con los cuales se desea realizar la estimación, el modelo espacio estado (espacio estado) requiere una cantidad mucho menor de memoria y requiere realizar menos operaciones para su cómputo.

Conclusión

El modelo de mayor interés de este trabajo, el modelo espacio estado, ha mostrado un desempeño favorable, a pesar de que las referencias revisadas lo anunciaban como un modelo inferior comparado con otros modelos de predicción. Esta discrepancia es posible explicarla si consideramos que los casos en que se habían usado modelos espacio estado, trabajaban con estados observables con valores binarios o discretos, en lugar de usar una escala continua de valores como se hizo en este trabajo. Otra razón para la valoración negativa que estos estudios dieron al modelo espacio estado se debe a que se le comparaba con modelos de redes neuronales (NN) o de *Bayesian Knowledge Tracing* (BKT), pero obviando las ventajas y desventajas que estos modelos tienen con respecto al modelo de espacio estado.

Los resultados obtenidos en esta tesis abren posibilidades para futuros análisis de esta estructura y para mejorar su poder predictivo, puesto que se usó una forma bastante simple del modelo y se trabajó con pocos instantes temporales. Existen dos aspectos fácilmente identificables en los que se puede obtener mejores resultados con esta estructura. El primero es considerar estados latentes y/o observables de mayor dimensión, para los que se necesitaría un mayor número de parámetros a estimar, pero permitirían un modelo más flexible en cuanto a cómo se modela el aprendizaje de cada alumno. El segundo aspecto es desglosar la información en un mayor número de instantes temporales –tal vez pasar de meses a quincenas o semanas, inclusive a estados diarios–, pero esto llevaría a la aparición de datos faltantes, dado que no se realizan clases ni ejercitaciones todos los días de la semana.

Una de las ventajas que comparten los modelos presentados en esta tesis en comparación con otros que usan factorización matricial, NN o BKT, es la independencia con respecto a las preguntas realizadas, pues en ninguno se usó algún parámetro que requiera tener más información sobre las preguntas que el eje de contenidos al que pertenecen. Es común ver estudios donde los estimadores están fuertemente condicionados al conjunto de preguntas utilizadas, por lo que modificar ese conjunto puede inutilizar al modelo como herramienta para la estimación fuera de la muestra. Este problema no se presenta con los modelos trabajados.

En cuanto a los resultados obtenidos con este modelo, se pudo concluir que es casi tan bueno como los modelos lineales con más de 100 parámetros, con la diferencia de que este usa solo 17. Además, la elección de estos parámetros nos permite vislumbrar qué aspectos son buenos indicadores del desempeño de los alumnos. En ese sentido, cabe resaltar dos variables que estuvieron presentes en varios modelos y que, al analizar cada modelo individualmente, estuvieron dentro de las 10 mejores combinaciones de variables para cada modelo. Estas variables son: la variable histórica del promedio curso con respecto a la afirmación “Mi com-

portamiento es un problema para el profesor de matemáticas”, la cual se detectó como un indicador negativo del desempeño de los alumnos; y la variable dinámica de precisión de respuesta en cada eje, la cual tiene una incidencia positiva en el puntaje de cada alumno.

Que estas variables sean buenos regresores nos permite entender la importancia de al menos dos aspectos en el aprendizaje de las matemáticas. La primera nos muestra cuánto y cómo afecta el clima de un curso al desempeño. Cabría entonces prestar especial atención a cómo se construyen ambientes de aprendizaje óptimos para los estudiantes. En cuanto a la segunda, es interesante notar cómo la precisión de respuesta de los alumnos nos da una mejor visión del desempeño de los mismos, aún más que el número de ejercicios realizados, inclusive si solo se cuentan los ejercicios que el alumno realizó con máxima precisión. Esto podría traducirse a que es más importante que los alumnos respondan correctamente los ejercicios que hacen a que hagan muchos. Es decir, si bien el número de ejercicios respondidos es un buen indicador del desempeño, no es tan bueno como la precisión.

De forma general, el desempeño de cada modelo, en particular del modelo con efecto olvido, permitió obtener una estimación del puntaje SIMCE, cuyo error se acercó al error de medición de este instrumento por 10.05 puntos en cuanto a estimaciones a nivel estudiante. Como estimador del puntaje curso, la estimación es aún más cercana puesto que el RMSE se obtuvo en órdenes de magnitud de 0.86 de diferencia con el SEM del SIMCE. Cuando las predicciones de desempeño se acercan al SIMCE y, al mismo tiempo, tienen la flexibilidad de ser obtenidas mes a mes (o en el caso del modelo espacio estado, de manera aún más frecuente sin necesidad de aumentar el número de parámetros), estamos frente a un modelo que nos entrega una señal del desempeño de los alumnos en el instante presente y sin la necesidad de una gran evaluación sumativa. Esto se consigue mediante una evaluación formativa que tiene en cuenta el proceso de aprendizaje que ha experimentado el alumno.

Los estados $x(t)$ e $y(t)$ del modelo espacio estado pueden ser una herramienta para los profesores, la cual les permitiría medir el grado de habilidad o aprendizaje de los estudiantes en matemáticas. Por la forma en que se construyen estos valores, los profesores no necesitarían de una evaluación que requiriese de una logística compleja o un costo elevado, como ocurre con la evaluación MideUC, cuyo *baseline* fue superado por este modelo. Esto liberaría al docente de la carga psicológica y el estrés innecesario que suelen implicar estas evaluaciones. Así, un profesor menos agobiado podría construir un ambiente más ameno en el aula y, como ya se mencionó anteriormente, este es otro aspecto importante para el aprendizaje.

El costo que implica cada modelo en términos de memoria y operaciones requeridas, muestra cómo el modelo espacio estado es significativamente más económico, en especial si se desea trabajar con modelos que permitan obtener estimaciones de desempeño de forma más frecuente, dado que los otros modelos dependen cuadráticamente del número de instantes en que se realiza las estimaciones de desempeño. Solo el primer modelo y el espacio estado son lineales en este parámetro y también en el número de estudiantes, pero el primero requiere de aproximadamente el doble de memoria y 10 veces más operaciones que el modelo espacio estado.

Estos resultados presentan diversas oportunidades en el área de educación, siendo una de estas, la posibilidad de implementar un sistema de evaluación estandarizada que haga uso de evaluaciones formativas, entregando indicadores de desempeño mensuales, semanales

o inclusive con mayor frecuencia si se hacen los ajustes necesarios. Esto representa un sistema de evaluación con mayor frecuencia de medición, menores costos y más dinámico. Al mismo tiempo esto permitiría a los docentes adaptar la ejercitación y los contenidos acorde a la situación actual de aprendizaje de los alumnos. Existe otras posibles aplicaciones para estos modelos en educación, que pueden ser consideradas agregando mayor complejidad al modelo, como sería agregar más dimensiones a los estados, considerando las distintas áreas de contenido que aprenden los estudiantes durante el año, y por la estructura de los modelos espacio estado, esto consideraría la interacción entre las distintas disciplinas que se incluyen en el currículo escolar.

Bibliografía

- [1] Rahaf Alamri y Basma Alharbi. «Explainable Student Performance Prediction Models: A Systematic Review». En: *IEEE Access* 9 (feb. de 2021), págs. 33132-33143.
- [2] Zachary Pardos Anirudhan Badrinath Frederic Wang. «pyBKT: An Accessible Python Library of Bayesian Knowledge Tracing Models». En: *arXiv:2105.00385v1 [cs.MS]* (mayo de 2021).
- [3] Roberto Araya y col. «Does Gamification in Education Work? Experimental Evidence from Chile». En: *IDB Working Paper Series* 982 (jul. de 2019).
- [4] David Ardia, Juan Ospina Arango y Norman Giraldo Gomez. «Jump-Diffusion Calibration using Differential Evolution». En: *Wilmott Magazine* 55 (2011), págs. 76-79. URL: <http://www.wilmott.com/>.
- [5] David Ardia y col. *DEoptim: Differential Evolution in R*. version 2.2-5. 2020. URL: <https://CRAN.R-project.org/package=DEoptim>.
- [6] David Ardia y col. «Differential Evolution with DEoptim: An Application to Non-Convex Portfolio Optimization». En: *The R Journal* 3.1 (2011), págs. 27-34. URL: https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Ardia~et~al.pdf.
- [7] Chen B.M. *Solutions to Discrete-time Riccati Equations*. Natural Computing. Springer, London, 2000, págs. 95, 96.
- [8] James Dean Brown. «Standard error vs. Standard error of measurement». En: *Shiken: JALT Testing & Evaluation SIG Newsletter* 3.1 (abr. de 1999), págs. 20-25.
- [9] Investigación y Estrategia Brújula. *IX Encuesta de Acceso y Usos de Internet - Informe Final*. Inf. téc. Subsecretaría de Telecomunicaciones de Chile, dic. de 2017.
- [10] Agencia de la Calidad de la Educación. «¿Quiénes somos?» En: <http://www.agenciaeducacion.cl/nosotros/quienes-somos/> (2016).
- [11] Educación. *¿Qué hace el BID en educación?* Inf. téc. Banco Interamericano del Desarrollo, 2014.
- [12] División Estudios. *Informe Técnico Simce 2012*. Inf. téc. Agencia de Calidad de la Educación, 2013.
- [13] División Estudios. *Informe Técnico Simce 2013*. Inf. téc. Agencia de Calidad de la Educación, ene. de 2014.
- [14] División Estudios. *Informe Técnico Simce 2014*. Inf. téc. Agencia de Calidad de la Educación, 2015.
- [15] División Estudios. *Informe Técnico Simce 2017*. Inf. téc. Agencia de Calidad de la Educación, 2018.
- [16] División de Evaluación de Logros de Aprendizaje. *Informe Técnico Simce 2015*. Inf. téc. Agencia de Calidad de la Educación, 2016.

- [17] Hashmia Hamsa, Simi Indiradevi y Jubilant Kizhakkethottam. «Student Academic Performance Prediction Model Using Decision Tree and Fuzzy Genetic Algorithm». En: *Procedia Technology* 25 (dic. de 2016), págs. 326-332. DOI: 10.1016/j.protcy.2016.08.114.
- [18] Raza Hasan y col. «Student Academic Performance Prediction by using Decision Tree Algorithm». En: *2018 4th International Conference on Computer and Information Sciences (ICCOINS)*. 2018, págs. 1-5. DOI: 10.1109/ICCOINS.2018.8510600.
- [19] Carolyn J. Hill y col. «Empirical Benchmarks for Interpreting Effect Sizes in Research». En: *Child Development Perspectives* 2.3 (nov. de 2008), págs. 172-177.
- [20] Lennart Ljung. *System Identification: Theory for User*. 2.^a ed. Prentice Hall, 1999, págs. 86-87. URL: <http://gen.lib.rus.ec/book/index.php?md5=3465e4ec555ff2b40402aed92>
- [21] Hayan Bai Mansureh Kebritchi Atsusi Hirumi. «The effects of modern mathematics compure games on mathematics archievement and class motivation». En: *Computers & Education* 22.2 (sep. de 2010), págs. 427-443.
- [22] George Chrysogonidis Marina Delianidi Kostas Diamantaras y Vasileios Nikiforidis. «Student Performance Prediction Using Dynamic Neural Models». En: *Educational Data Mining 2021* (mayo de 2021).
- [23] *MATLAB Optimization Toolbox*. The MathWorks, Natick, MA, USA. 2020.
- [24] Mausumi Islam Mau Md. Ahsan Arif Asma Jahan y Razia Tummarzia. «An Improved Prediction System of Students' Performance Using Classification model and Feature Selection Algorithm». En: *International Journal of Advances in Soft Computing and its Applications* 13.1 (mar. de 2021), págs. 162-177.
- [25] Fei Mi y Dit-Yan Yeung. «Temporal Models for Predicting Student Dropout in Massive Open Online Courses». En: *IEEE International Conference on Data Mininig Workshop* (2015), págs. 256-263.
- [26] Vincent Verdult Michel Verhaegen. *Filtering and system identification: a least squares approach*. 1.^a ed. Cambridge University Press, 2007. URL: <http://gen.lib.rus.ec/book/index.php?md5=5719be329ecc48f4b77a4949a0c17487>.
- [27] Johan Van der Molen Moris. «Minería de datos educacionales: Modelos de predicción del desempeño escolar en alumnos de enseñanza básica». En: *Universidad de Chile* (2013).
- [28] Katharine Mullen y col. «DEoptim: An R Package for Global Optimization by Differential Evolution». En: *Journal of Statistical Software* 40.6 (2011), págs. 1-26. URL: <http://www.jstatsoft.org/v40/i06/>.
- [29] Kenneth V. Price, Rainer M. Storn y Jouni A. Lampinen. *Differential Evolution - A Practical Approach to Global Optimization*. Natural Computing. ISBN 540209506. Springer-Verlag, ene. de 2006.
- [30] Xin Ma Quing Li. «A Meta-analysis of the Effects of Computer Technology on School Students' Mathematics Learning». En: *Educational Psychology Review* 22 (abr. de 2010), págs. 215-243.
- [31] Hariharan Swaminathan (auth.) Ronald K. Hambleton. *Item Response Theory: Principles and Applications*. 1.^a ed. Springer Netherlands, 1985. URL: <http://gen.lib.rus.ec/book/index.php?md5=1d0db35621afc71ba06ae30605ff0ed8>.
- [32] Carlotta Schatten y Lars Schmidt-Thieme. «Hybrid Matrix Factorization Update for Progress Modeling in Interlligent Tutoring Systems». En: *International Conference on Computer Supported Education* (2016), págs. 49-70.

- [33] John A. Self. «Student Models in Computer-aided Instruction». En: *International Journal of Man-Machine Studies* 6.2 (mar. de 1974), págs. 261-276.
- [34] Manie Tadayon y Greg Pottie. «Predicting Student Performance in an Educational Game Using a Hidden Markov Model». En: *IEE Transactions On Education* 63 (2020), págs. 299-304.
- [35] Howard Wainer. «Adjusting for Differential Base Rates: Lord's Paradox Again». En: *Psychological Bulletin* 109.1 (1991), págs. 147-151.
- [36] Feng Wang y Li Chen. «A Nonlinear State Space Model for Identifying At-Risk Students in Open Online Courses». En: *EDM* (2016).
- [37] Xinhua Wang y col. «Student Performance Prediction with Short-Term Sequential Campus Behaviors». En: *Information* 11.4 (abr. de 2020). ISSN: 2078-2489. DOI: 10.3390/info11040201. URL: <https://www.mdpi.com/2078-2489/11/4/201>.
- [38] M. V. Aditya Nag Y. Sri Lalitha Y. Gayathri y Sk. Althaf Hussain Basha. «Student Performance Prediction - A Data Science Approach». En: *Modern Approaches in Machine Learning and Cognitive Science: A Walkthrough* 2 (mayo de 2021), págs. 115-125.
- [39] Tsung-Yen Yang y col. «Behavior-Based Grade Prediction for MOOCs Via Time Series Neural Networks». En: *IEEE Journal of Selected Topics in Signal Processing* 11.5 (ago. de 2017).

Anexo

En este anexo se detallará las variables empleadas, sus valores originales y cómo estos fueron codificados.

El capítulo 1, en la sección 1.3 se menciona los distintos cuestionarios usados y los ítem que estos poseen. A continuación se detallará por cada uno de estos cuestionarios y se indicará las preguntas, los ítem respectivos a esa pregunta y las posibles respuestas para cada ítem.

Cuestionario SIMCE de profesores

De este cuestionario, se usó las siguientes preguntas y sus respectivos ítem.

Pregunta	Ítems	Respuestas
¿Posee título de profesor otorgado por alguna de las siguientes instituciones?	1) Si, otorgado por una escuela normal. 2) Si, otorgado por una universidad. 3) Si, otorgado por un IP o CFT. 4) No poseo título de profesor.	1: Marca. 0: Vacío.
Indique cuál es el título que obtuvo.	1) EGB sin mención. 2) EGB con mención en Lenguaje. 3) EGB con mención en Matemática. 4) EGB con mención en Cs. Naturales. 5) EGB con mención en Cs. Sociales. 6) EGB con mención en otra mención. 7) EM con esp. en Lenguaje. 8) EM con esp. en Matemática. 9) EM con esp. en Biología. 10) EM con esp. en Química. 11) EM con esp. en Física. 12) EM con esp. en Cs Sociales. 13) EM con esp. en otra esp.	1: Marca. 0: Vacío.
¿Cuántos años de experiencia tiene usted como docente?		Numérico en años.
¿Qué asignatura enseña usted en este curso?	Lenguaje y Comunicación. Matemática. Ciencias Naturales. Historia, Geografía y Cs. Sociales. Artes Visuales. Música. Educación Tecnológica. Educación Física. Inglés.	1: Marca. 0: Vacío.
Pensando en el futuro, ¿cuál cree usted que es el nivel educacional más alto que completará la mayoría de los estudiantes del curso evaluado?		1: La mayoría no terminará la enseñanza media. 2: La mayoría completará EM técnico profesional. 3: La mayoría completará la EM científico humanista. 4: La mayoría completará una carrera en un IP o CFT. 5: La mayoría completará una carrera en una univ. 6: La mayoría completará un programa de postgrado.

Con estas 5 preguntas y 28 ítem se elaboran las siguientes variables:

Variable	Tipo de variable	Valores
Título del profesor	Categórica	Escuela normal. Universidad. IP o CFT. No posee título.
Mención en matemática	Binaria	1: Si 0: No
Mención en otra área	Categórica	Matemática. Otro. Ninguna.
Número de menciones.	Numérica	Entero positivo.
Tiene mención en EGB o EM	Categórica	Básica. Media.
Años como docente.	Numérica	Entero Positivo
Enseña la asignatura de matemática al curso	Binaria	1: Si. 0: No.
Expectativas del curso en años	Numérica.	9: EM Incompleta. 12: EM Técnico Profesional. 12: EM Científico Humanista. 14: CFT o IP. 17: Universidad. 19: Postgrado.
Expectativas del curso.	Categórica.	EM Incompleta. EM Técnico Profesional. EM Científico Humanista. CFT o IP. Universidad. Postgrado.

Cuestionario SIMCE de padres

En este cuestionario las preguntas usadas fueron:

Pregunta	Items	Respuestas
¿Hasta qué nivel educacional llegó el padre (o pareja de la madre) del estudiante?		1: No estudió. 2: 1er año de EB. 3: 2do año de EB. 4: 3er año de EB. 5: 4to año de EB. 6: 5to año de EB. 7: 6to año de EB. 8: 7mo año de EB. 9: 8vo año de EB. 10: I año de EM. 11: II año de EM. 12: III año de EM.
¿Hasta qué nivel educacional llegó la madre (o pareja del padre) del estudiante?		13: IV año de EM científico humanista. 14: IV o V año de EM técnico profesional o vocacional. 15: Educación incompleta en un CFT o IP. 16: Educación completa en un CFT o IP. 17: Educación incompleta en una universidad. 18: Educación completa en una universidad. 19: Grado de magíster universitario. 20: Grado de doctor universitario. 21: No sabe o no recuerda. 0: Vacío. 99: Doble marca.
Pensando en el futuro, ¿cuál cree usted que es el nivel educacional más alto que el estudiante completará?		1: No creo que complete IV año de EM. 2: IV año de EM técnico profesional. 3: IV año de EM científico humanista. 4: Una carrera en un IP o CFT. 5: Una carrera en una universidad. 6: Estudios de postgrado. 0: Vacío. 99: Doble marca.

Con estas 3 preguntas se elaboraron las variables:

Variable	Tipo de variable	Valores
Educación del padre (o pareja de la madre)	Categórica	Mismas alternativas que la pregunta del cuestionario.
Educación del padre (o pareja de la madre)	Categórica	
Expectativa de los padres del nivel máximo que el estudiante completará	Categórica	
Años de educación del padre (o pareja de la madre)	Numérica	1-8: Según niveles de 1ro a 8vo año EB. 9-12: Según niveles de EM 12.5: EM Técnico Profesional. 13.5: CFT o IP incompleta. 14.5: CFT o IP completa. 15: Universitaria incompleta. 17: Universitaria completa. 19: Grado de Magíster. 21: Grado de Doctor.
Años de educación de la madre (o pareja del padre)	Numérica	
Expectativa en años de los padres del nivel máximo que el estudiante completará	Numérica	9: EM Incompleta. 12: EM Técnico Profesional. 12: EM Científico Humanista. 14: CFT o IP. 17: Universidad. 19: Postgrado.

Cuestionario SIMCE de estudiantes

De el cuestionario de estudiantes de seleccionó las siguientes preguntas:

Pregunta	ítem	Respuestas
Pensando en un día normal al que asistes a clases, ¿cuántas horas aproximadamente duermes durante la noche?		0: Vacío. 1: 5 horas o menos. 2: 6 horas. 3: 7 horas. 4: 8 horas. 5: 9 horas o más. 99: Doble marca.
En cada oración, marca la respuesta que mejor describe lo que piensas de Matemática.	Me va bien en las pruebas de matemáticas. Matemática es fácil para mi. Siento temor de que las pruebas de matemática sean difíciles para mi. Me preocupa tener malas notas en matemática. Me pongo nervioso antes de las pruebas de matemática. Mi profesor me dice que soy bueno para matemática. Me pongo nervioso cuando no entiendo una tarea de matemática	0: Vacío. 1: Muy en desacuerdo. 2: En desacuerdo. 3: De acuerdo. 4: Muy de acuerdo. 99: Doble marca.
¿Te dicen tus padres y tus profesores las siguientes cosas?	Mis padres me dicen que soy buen estudiante. Mis padres me dicen que soy capaz de aprender.	0: Vacío. 1: Muy en desacuerdo. 2: En desacuerdo. 3: De acuerdo. 4: Muy de acuerdo. 99: Doble marca.

Usando estas preguntas se construyen las siguientes variables.

Variable	Tipo de variable	Valores
Horas de sueño	Numérica	5: 5 horas o menos. 6: 6 horas. 7: 7 horas. 8: 8 horas. 9: 9 horas o más.
Me va bien en las pruebas de matemáticas	Numérica	1: Muy en desacuerdo. 2: En desacuerdo. 3: De acuerdo. 4: Muy de acuerdo.
Matemáticas es fácil para mí	Numérica	
Siento temor que las pruebas de matemáticas sean difíciles para mí	Numérica	
Me preocupa tener malas notas en matemática	Numérica	
Me pongo nervioso antes de las pruebas de matemática	Numérica	
Mi profesor dice que soy bueno para matemáticas	Numérica	
Me pongo nervioso cuando no entiendo una tarea de matemática	Numérica	
Mis padres dicen que soy buen estudiante.	Numérica	
Mis padres dicen que soy capaz de aprender	Numérica	

Cuestionario ConectaIdeas

De las preguntas del cuestionario realizado como parte del programa, se usaron las siguientes:

Pregunta	ítem	Respuestas
Pensando en aprendizaje matemáticas	<p>Me gusta estudiar matemáticas. Me gustaría no tener que estudiar matemáticas. Matemáticas es aburrida. Aprendo muchas cosas interesantes cuando estudio matemáticas. Me gusta matemáticas. Me gusta hacer tareas que tienen números. Me gusta resolver problemas matemáticos. Tengo ganas de tener otra clase de matemáticas. Matemáticas es mi materia favorita.</p>	<p>1: Muy de acuerdo 2: De acuerdo 3: En desacuerdo 4: Muy en desacuerdo 8: No contesta 9: Contesta mal.</p>
Pensando en aprendizaje matemáticas	<p>La inteligencia es algo que no se puede cambiar mucho. Se puede aprender cosas nuevas pero no se puede cambiar la inteligencia.</p>	
Pensando en clases de matemáticas	<p>Me siento feliz en esta clase de matemáticas. Estar en clase de matemáticas me hace sentir triste o enojado. Las cosas que hicimos este año en esta clase son interesantes. Gracias a este profesor estoy aprendiendo a amar las matemáticas. Me gusta la clase de matemáticas de este curso.</p>	
Pensando en clases de matemáticas	<p>Me comporto bien en matemáticas. Mi comportamiento a veces molesta al profesor. Mi comportamiento es un problema para el profesor de matemáticas.</p>	
Pensando en matemáticas	<p>Normalmente me va bien en matemáticas. Para mi matemáticas es más difícil que para muchos otros estudiantes en mi clase. No soy bueno en matemáticas.</p>	

Pregunta	Items	Respuestas
Pensando en lo que sucede en tus clases de matemáticas	Hacen ruido y desorden Los estudiantes escuchan lo que dice el profesor. Profesores tienen que esperar largo rato a que alumnos nos tranquilicemos.	1: Muy de acuerdo 2: De acuerdo 3: En desacuerdo 4: Muy en desacuerdo 8: No contesta 9: Contesta mal.
Cuánto de acuerdo estás con las siguientes afirmaciones	Me gusta leer. Leo en mi casa todos los días. Hago tareas de lenguaje todos los días en casa. Me gusta la clase de lenguaje. Hago tareas todos los días en casa de matemáticas. Me gusta la clase de matemáticas Uso PC en mi casa para practicar matemáticas.	
Preferencia por tipo de clases	¿Prefieres clase matemáticas en sala o computación?	1: Sala de clases 2: Laboratorio 8: No contesta 9: Contesta mal
¿Qué tan de acuerdo estás con las siguientes afirmaciones?	Soy bueno para escuchar a los demás. Disfruto viendo a mis compañeros tener éxito. Tomo en cuenta lo que les interesa a los demás. Disfruto teniendo en cuenta diferentes perspectivas.	1: Muy de acuerdo 2: De acuerdo 3: En desacuerdo 4: Muy en desacuerdo 8: No contesta 9: Contesta mal.
¿Qué tan de acuerdo estás con las siguientes afirmaciones?	Prefiero trabajar como parte de un equipo a trabajar solo. Encuentro que los equipos toman mejores decisiones que las personas. Encuentro que el trabajo en equipo aumenta mi propia eficiencia. Me gusta cooperar con mis compañeros.	

y con estas preguntas, cada ítem conforma una nueva variable como se detalla a continuación.

Variable	Tipo de variable	Valores
Me gusta estudiar matemáticas	Numérica	1: Muy en desacuerdo. 2: En desacuerdo. 3: De acuerdo. 4: Muy de acuerdo.
Me gustaría no tener que estudiar matemáticas		
Matemáticas es aburrida		
Aprendo muchas cosas interesantes cuando estudio matemáticas		
Me gusta matemáticas		
Me gusta hacer tareas que tienen números		
Me gusta resolver problemas matemáticas		
Tengo ganas de tener otra clase de matemáticas		
Matemáticas es mi materia favorita		
La inteligencia es algo que no se puede cambiar mucho		
Se puede aprender cosas nuevas pero no se puede cambiar la inteligencia		
Me siento feliz en esta clase de matemáticas		
Estar en clase de matemáticas me hace sentir triste o enojado		
Las cosas que hicimos este año en esta clase son interesantes		
Gracias a este profesor estoy aprendiendo a amar las matemáticas		
Me gusta las clases de matemáticas de este curso		
Me comporto bien en matemáticas		
mi comportamiento a veces molesta al profesor		
mi comportamiento es un problema para el profesor en matemáticas		
Normalmente me va bien en matemáticas		
Para mi matemáticas es más difícil que para muchos otros estudiantes en mi clase		
No soy bueno en matemáticas		
Hay ruido y desorden		
Los estudiantes escuchan lo que dice el profesor		

Variable	Tipo de variable	Valores
Profesores tienen que esperar largo rato a que alumnos nos tranquilicemos	Numérica	1: Muy en desacuerdo. 2: En desacuerdo. 3: De acuerdo. 4: Muy de acuerdo.
Me gusta leer		
Leo en mi casa todos los días		
Hago tareas de lenguaje todos los días en casa		
Me gusta la clase de lenguaje		
Hago tareas todos los días en casa de matemáticas		
Me gusta la clase de matemáticas		
Uso PC en mi casa para practicar matemáticas		
¿Prefieres clase matemáticas en sala o computación?	Categórica	Sala Laboratorio
Soy bueno para escuchar a los demás.	Numérica	1: Muy en desacuerdo. 2: En desacuerdo. 3: De acuerdo. 4: Muy de acuerdo.
Disfruto viendo a mis compañeros tener éxito.		
Tomo en cuenta lo que les interesa a los demás.		
Disfruto teniendo en cuenta diferentes perspectivas.		
Prefiero trabajar como parte de un equipo a trabajar solo.		
Encuentro que los equipos toman mejores decisiones que las personas.		
Encuentro que el trabajo en equipo aumenta mi propia eficiencia.		
Me gusta cooperar con mis compañeros.		