



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

ESTUDIO DE LA RELACIÓN ENTRE LA ACTIVIDAD ONLINE Y EL
RENDIMIENTO ACADÉMICO DE ESTUDIANTES DE INGENIERÍA Y EDUCACIÓN
EN DOS SISTEMAS DE GESTIÓN DEL APRENDIZAJE

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

JOAQUÍN IGNACIO MUÑOZ DÍAZ

PROFESOR GUÍA:
SERGIO CELIS GUZMÁN

MIEMBROS DE LA COMISIÓN:
OMAR CERDA INOSTROZA
RICHARD WEBER HAAS

Este trabajo ha sido parcialmente financiado por el proyecto FONDECYT #1161413

SANTIAGO DE CHILE
2017

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL
POR: JOAQUÍN IGNACIO MUÑOZ DÍAZ
FECHA: 2017
PROF. GUÍA: SR. SERGIO CELIS GUZMÁN

ESTUDIO DE LA RELACIÓN ENTRE LA ACTIVIDAD ONLINE Y EL
RENDIMIENTO ACADÉMICO DE ESTUDIANTES DE INGENIERÍA Y EDUCACIÓN
EN DOS SISTEMAS DE GESTIÓN DEL APRENDIZAJE

Los avances tecnológicos de las últimas décadas están transformando la forma en la que los estudiantes aprenden. En educación superior, las instituciones académicas han adoptado los Sistemas de Gestión del Aprendizaje, que permiten a alumnos y profesores extender las fronteras de las salas de clase al mundo digital. El auge de estas plataformas ha suscitado el interés de la comunidad investigativa por utilizar la gran cantidad de datos generados por estas tecnologías para determinar y mejorar el efecto de su uso en la forma en la que los estudiantes aprenden y en el impacto en sus resultados académicos.

Este trabajo de memoria se inscribe en el marco del proyecto de investigación FONDECYT #1161413, y busca contribuir a la literatura sobre la relación entre procesos de aprendizaje y desempeño académico, a través del análisis del uso que estudiantes de primer año de la Universidad de Chile y la Pontificia Universidad Católica de Chile dan a los sistemas de gestión del aprendizaje que cada una de ellas posee. Para ello, se analizan las bases de datos de actividad online con herramientas estadísticas y se caracteriza cada plataforma en base a la distribución y frecuencia de uso de sus distintas funcionalidades, empleando variados métodos de visualización que combinan esta información con el desempeño académico. Se elabora además un modelo explicativo para evaluar el poder predictivo de variables construidas a partir de los registros de actividades de los estudiantes.

En términos de la distribución de uso, aun cuando ambos sistemas de gestión del aprendizaje cuentan con funcionalidades similares, existe una cultura de uso de cada plataforma propia a cada institución, la cual es transversal a estudiantes con distintas características o desempeño: mientras en la Universidad Católica la mayor parte de las actividades ocurre en módulos asociados a la gestión del contenido, en la Universidad de Chile estas ocurren en módulos que permiten la interacción entre distintos usuarios. Estos resultados sugieren que los factores que afectan el uso que cada estudiante da a la plataforma se relacionan con su interfaz y funcionalidades; a definiciones propias del equipo docente de cada curso; y al rol y validez que la comunidad educativa le entregue a la plataforma. Relativo a la frecuencia de uso, se determinó que estudiantes que utilizan más las herramientas online logran un mejor desempeño académico al final del semestre. Por último, las regresiones realizadas muestran cómo variables derivadas de la actividad online de estudiantes pueden explicar hasta un 8% de la varianza del desempeño académico para una de las instituciones en estudio. Se destaca la oportunidad de observar la actividad online de estudiantes tanto para medir hábitos de estudio, comprender procesos de aprendizaje y para elaborar modelos de predicción de rendimiento.

Agradecimientos

En primer lugar quiero agradecer a Sergio, mi profesor guía, por la dedicación, el buen humor, la buena voluntad, la ayuda, el tiempo y la confianza de permitirme participar en sus proyectos. Al equipo de LA del que tuve la suerte de ser parte; Augusto, Carlos, Carolina, Dany, Max y Sergio, porque este trabajo refleja el esfuerzo y la motivación de cada uno de ustedes por mejorar la educación.

También agradecer a todos quienes me han orientado en este trabajo de memoria, tanto profesores como compañeros, pues muchas de sus sugerencias me han permitido dar forma a este trabajo.

A mis profesores y mentores, desde el jardín hasta la Universidad, pasando por las prácticas laborales, quienes me enseñaron y guiaron en el camino. Esta memoria es el fruto de su apoyo y dedicación.

A mi familia. En especial a mis padres y hermana, por amarme, aguantarme, y apoyarme todos estos años de las más diversas formas. Su entrega me ha dado la oportunidad de alcanzar los objetivos que me he ido propuesto, muchas gracias.

A mis amigos del colegio y Beauchef, de la 8 y ZD. Pues años de clases, tareas, trabajos y pruebas han sido amenos y llevaderos porque ustedes han estado allí para entregar su alegría. A quienes me acompañaron en la aventura en Francia, por pasar a ser mi segunda familia y entregarme tantos buenos recuerdos y su ayuda en los momentos difíciles. A mis amigos de Centrale, le 2B y CPI, y a quienes conocí durante la pasantía en USA, por compartir todas aquellas enriquecedoras experiencias que dejan recuerdos en el corazón, gracias.

A Kate, quien me ha acompañado estos últimos años de mi vida, por su amor y apoyo incondicional en las buenas y en las malas. Estar contigo es una experiencia que me entrega felicidad día a día. Quiero seguir creciendo y disfrutando junto a ti.

A todos, muchas gracias.

Tabla de Contenido

1. Introducción	1
2. Descripción del Proyecto	3
2.1. Ingeniería UChile y Educación UC	3
2.2. Sistemas de Gestión del Aprendizaje	4
2.3. U-Cursos y Sakai	5
2.4. Proyecto FONDECYT #1161413	7
2.5. Objetivos	8
2.5.1. Objetivo General	8
2.5.2. Objetivos Específicos	8
2.6. Alcances	9
3. Marco Conceptual	10
3.1. Student Learning Research	10
3.2. Educational Data Mining y Learning Analytics	11
3.3. Los LMS en Educación Superior	14
3.4. Elementos Estadísticos	15
3.4.1. Análisis de la Varianza	15
3.4.2. Test de Levene	15
3.4.3. Post hoc tests de Tukey HSD y Games-Howell	16
3.5. Elementos Econométricos	16
3.5.1. Notación	16
3.5.2. Estimador de Mínimos Cuadrados Ordinarios	16
3.5.3. Estimador de Eicker-White y de Clústers	17
3.5.4. Factor de Inflación de la Varianza	18
3.5.5. Evaluación de un modelo: Bondad de Ajuste	18
3.5.6. Test de Razón de Verosimilitud y Test de Wald	18
3.6. Conclusiones Marco Conceptual	19
4. Metodología	20
4.1. Origen de los datos	20
4.2. Datos sociodemográficos	21
4.3. Datos de la actividad online	23
4.3.1. Variables	23
4.3.2. Agrupación de los datos en bloques de tiempo	24
4.3.3. Generación de variables de comportamiento semestral	25

4.3.4.	Horarios de conexión	26
4.4.	Estructura del análisis	27
4.5.	Modelo Explicativo	27
4.5.1.	Estudios anteriores	27
4.5.2.	Formulación	28
5.	Resultados	30
5.1.	Población muestral	30
5.2.	Actividad en los LMS	32
5.2.1.	Actividad diaria y semanal	32
5.2.2.	Actividad anual	35
5.3.	Frecuencia y Distribución de uso de los LMS	37
5.3.1.	Frecuencia de uso	37
5.3.2.	Frecuencia de uso diferenciada por plataforma	39
5.3.3.	Distribución de uso temporal	42
5.3.4.	Distribución de uso según desempeño	44
5.3.5.	Uso por sección	45
5.3.6.	Conclusiones análisis de frecuencia y distribución de uso	46
5.4.	Sociograma	46
5.5.	Regresión Lineal Múltiple	48
5.5.1.	Correlación entre variables	48
5.5.2.	Forma Estructural	49
5.5.3.	Resultados	50
5.5.4.	Regresión Lineal Múltiple con información del desempeño histórico	51
5.5.5.	Limitaciones	54
5.5.6.	Conclusiones Modelo Explicativo	55
6.	Conclusión	57
6.1.	Conclusiones generales	57
6.2.	Toma de decisiones basada en datos	58
6.3.	Investigaciones futuras	59
	Glosario	61
	Bibliografía	62
	Anexos	64

Índice de Tablas

4.1. Descripción de variables sociodemográficas.	22
4.2. Descripción de variables LMS.	24
4.3. Detalle de la cuenta de actividades.	25
5.1. Tasas de reprobación por año y semestre de 0, 1 y 2 o más ramos.	32
5.2. Test de Levene.	39
5.3. Resultados de Welch-ANOVA y Games-Howell.	39
5.4. Test de Levene para cada plataforma.	41
5.5. Resultados de Welch-ANOVA y Games-Howell por plataforma.	42
5.6. Regresión lineal múltiple $zscore_{carrera}$	51
5.7. Regresión lineal múltiple $zscore_{carrera2}$	53
5.8. Coeficientes estandarizados para las regresiones (2) y (4).	54

Índice de Ilustraciones

2.1. Uso de los LMS por industria.	5
2.2. Módulos activos de un curso de la plataforma U-Cursos.	6
2.3. Extracto de la interfaz de Sakai en la UC.	6
3.1. Referencias a EDM y LA en Google Scholar.	12
3.2. Ciclo Educational Data Mining.	12
4.1. Ejemplo de histograma de eventos.	26
5.1. Composición de la muestra según sexo.	31
5.2. Composición de la muestra según colegio de procedencia.	31
5.3. Composición de la muestra según tipo de ingreso a la universidad.	31
5.4. Distribución del uso semanal de los LMS	33
5.5. Horario de las actividades en los LMS.	34
5.6. Histograma de la actividad online.	36
5.7. Diagramas de caja del uso de LMS.	38
5.8. Diagramas de caja del uso de LMS por plataforma.	40
5.9. Evolución en el tiempo del uso de cada categoría de los LMS.	43
5.10. Distribución de uso de la plataforma, según categoría.	44
5.11. Diagrama del uso por sección	45
5.12. Sociograma	47
5.13. Matriz de correlación de variables	48
5.14. Correlación entre zscores.	52

Capítulo 1

Introducción

Los paradigmas existentes sobre los procesos de aprendizaje y enseñanza se encuentran en constante evolución. Desde el punto de vista del conductismo, pasando por el paradigma cognitivo y el constructivista, las diversas concepciones de cómo los estudiantes aprenden dan lugar al diseño e implementación de distintos programas académicos que las comunidades educativas entregan a sus alumnos. Hoy en día, existe cierto consenso en fomentar en educación superior las metodologías de enseñanza que conciben el aprendizaje como un acto activo, en el cual la institución educativa y el cuerpo docente le entregan al estudiante las herramientas para que éste pueda asumir un rol protagónico en el desarrollo de sus competencias.

En este sentido, la irrupción de la tecnología ha tenido un gran impacto en cómo se desarrolla el proceso educativo tanto en el aula, como fuera de ella. En educación superior, las instituciones han adoptado el uso de los Sistemas de Gestión del Aprendizaje (LMS por sus siglas en inglés), que corresponden a plataformas online donde los estudiantes pueden, entre otros, acceder a material académico, interactuar con sus pares y profesores de forma asíncrona, visualizar sus notas, o bien participar en procesos evaluativos. Como consecuencia de la masiva implementación de las herramientas tecnológicas y de la gran cantidad de datos generados en el contexto educativo —en particular por las plataformas LMS— se ha abierto una línea de investigación, Learning Analytics (LA), la cual busca utilizar esta información para mejorar los procesos de enseñanza y aprendizaje, apoyando a la vez a los estudiantes, a docentes y a instituciones educativas. En este sentido, LA comprende metodologías, técnicas, modelos y algoritmos para entender los paradigmas de un sistema educativo del siglo XXI, donde el aprendizaje y la enseñanza no se limitan a la sala de clase ni a la interacción estudiante-profesor.

Entender la relación entre el uso de los LMS y los procesos de aprendizaje de estudiantes ha resultado clave para diseñar e implementar intervenciones que mejoren los sistemas educacionales. Durante la última década, investigadores en distintas universidades han comenzado a realizar proyectos de LA con distintos enfoques, entre los que destacan por ejemplo SNAPP (Social Networks Adapting Pedagogical Practice) desarrollado por Bakharia y Dawson (2011), el cual es un plug-in para navegador web que permite visualizar en forma de grafos la actividad de estudiantes en el foro de un LMS, detectando estudiantes con baja

participación; el proyecto Sherpa (Bramucci y Gaston, 2012) que corresponde a un sistema de recomendación académico que envía sugerencias personalizadas a los estudiantes; el sistema de alerta temprana de estudiantes en riesgo de reprobación de una asignatura desarrollado por Macfadyen y Dawson (2010); y el trabajo de Celis et al. (2015) que permite identificar tempranamente estudiantes que podrían caer en causal de eliminación de la universidad por motivos académicos.

En esta línea, este tema de memoria se enmarca en el proyecto FONDECYT #1161413, el cual busca caracterizar los procesos de aprendizaje de estudiantes de primer año de dos carreras profesionales pertenecientes a universidades distintas, integrando herramientas metodológicas de LA y teorías de aprendizaje tradicionales, con el objetivo de generar un modelo integral que permita apoyar y mejorar el aprendizaje de los estudiantes. En particular, se desea en este trabajo dar luces sobre la relación entre el uso de dos plataformas LMS y el desempeño académico de estudiantes. Se busca determinar relaciones entre cómo los estudiantes utilizan los LMS y sus resultados académicos, con el fin de estudiar la pertinencia de utilizar la actividad online como variable explicativa de un modelo predictivo de desempeño. Se espera, de esta manera, ayudar a la caracterización de parte de los procesos de aprendizaje de estos estudiantes.

Para ello, se realiza un análisis de las estadísticas descriptivas de datos sociodemográficos del universo considerado, así como de los eventos registrados en los LMS. Mediante distintas visualizaciones y un análisis estadístico se estudia la frecuencia con que estudiantes que obtienen distintos niveles de desempeño académico al final del semestre realizan actividades online, así como la forma de la distribución temporal de estas actividades a lo largo del año. Se incluye un sociograma inspirado en el proyecto SNAPP para explorar la visualización a través de grafos de la actividad de los alumnos en los LMS y sus resultados académicos, en miras a una posible implementación en el prototipo del proyecto FONDECYT. Por último, se realiza una regresión con variables derivadas de la actividad en los LMS para predecir desempeño académico.

El documento se estructura de la siguiente forma: En el capítulo 2 Descripción del Proyecto, se presentan los principales antecedentes para el desarrollo de la memoria, y se definen los objetivos y alcances del trabajo; en el capítulo 3 Marco Conceptual, se describen las principales áreas de investigación en educación que fueron revisadas, así como los elementos estadísticos y econométricos utilizados en el trabajo; en el capítulo 4 Metodología, se describe la población muestral, los datos y las variables utilizadas en el trabajo; en el capítulo 5 Resultados, se describen los principales hallazgos encontrados; y por último, en el capítulo 6 Conclusión, se entrega una visión general del trabajo realizado, así como posibles futuras líneas de investigación.

Capítulo 2

Descripción del Proyecto

En este capítulo se introducen los principales antecedentes necesarios para el desarrollo del proyecto de título. Primero, se presentan brevemente el contexto educativo de la población muestral, los sistemas de gestión del aprendizaje, y las plataformas estudiadas: U-Cursos y Sakai. Luego se incluye el proyecto FONDECYT en el que se enmarca el trabajo realizado, para dar paso a los objetivos y alcances del tema de memoria.

2.1. Ingeniería UChile y Educación UC

Se trabaja con una población muestral de estudiantes de primer año cohorte 2013-2014 de tres carreras profesionales pertenecientes a dos instituciones educativas: Pedagogía en Educación Parvularia y Pedagogía en Educación Básica pertenecientes a la Facultad de Educación de la Pontificia Universidad Católica de Chile (UC), y Plan Común de Ingeniería y Ciencias pertenecientes a la Facultad de Ciencias Físicas y Matemáticas (FCFM) de la Universidad de Chile (UChile). Las razones para trabajar con esta población son variadas: (1) Los estudiantes de primer año enfrentan un periodo de transición del colegio a la vida universitaria que es complejo y es cuando existen mayores riesgos de repetición y abandono, por lo que caracterizar sus procesos de aprendizaje puede tener mayor impacto al realizar intervenciones en el futuro; (2) el contar con estudiantes de educación e ingeniería permite diversificar el espectro de perfiles académicos, cercanos tanto a las ciencias sociales como a las ciencias físicas y matemáticas; y (3) ambas universidades utilizan LMS distintos, por lo que será posible analizar características de uso transversales y específicas a las plataformas.

La Universidad de Chile de carácter público y la Universidad Católica de carácter privado, son las dos universidades más antiguas del país, las cuales se ubican en la Región Metropolitana y ofrecen diversos programas conducentes a títulos profesionales. Ambas instituciones educativas reciben a estudiantes de todo el país de forma altamente selectiva, proceso que se desarrolla a través de la Prueba de Selección Universitaria (PSU), la cual los estudiantes pueden rendir anualmente¹.

¹Sin perjuicio de lo anterior, ambas universidades también cuentan con formas de ingreso especial donde

Al ingresar a la FCFM, los estudiantes comienzan un plan común de dos años, consistente principalmente de asignaturas matemáticas y físicas, tras el cual se dividen según el programa específico de su interés, los cuales a nivel de pregrado corresponden a 9 carreras de ingeniería, 3 licenciaturas y la carrera de geología².

Por otro lado, la Facultad de Educación de la UC dicta las carreras de pregrado de pedagogía en educación parvularia, pedagogía general básica con distintas menciones (quienes también siguen un plan común de 4 años) y pedagogía media con distintas especialidades en áreas de las ciencias. Estos distintos programas tienen como objetivo formar profesores con sólidas bases en el ámbito pedagógico y disciplinar preparados para diseñar y conducir el proceso educativo en el aula, facilitando el aprendizaje de niños y niñas del nivel escolar³.

2.2. Sistemas de Gestión del Aprendizaje

Desde principios de los años 2000 ha existido un crecimiento exponencial en el uso de tecnologías de la información en las distintas áreas de la vida diaria. En educación, se han desarrollado diversas herramientas e iniciativas que tienen por objetivo mejorar la forma en que los estudiantes aprenden. En esta línea, especial auge han tenido las plataformas online que permiten a los estudiantes continuar su aprendizaje fuera del aula. Dentro de los distintos tipos de plataformas existentes se encuentran aquellas destinadas a proveer cursos online masivos abiertos (MOOC por sus siglas en inglés) a un gran número de participantes, como Coursera, EdX o Udacity; aquellas plataformas o sistemas que se enfocan en la gestión del contenido del aprendizaje (CMS); y aquellas enfocadas en administrar las actividades de formación no presencial que imparte una institución, los Sistemas de Gestión del Aprendizaje.

El año 2015 se realizó una investigación de la industria de los LMS⁴ que da cuenta de su relevancia: Los LMS son utilizados alrededor del mundo tanto por instituciones académicas (21%), como por empresas privadas de distintos rubros para capacitar a sus empleados (Ver Figura 2.1); el 87% de las instituciones y empresas que utilizan un LMS se inclinan por una versión en la nube; y se estima que el valor de la industria ascenderá de 2.55 a 7.83 billones de dólares en el periodo 2013-2018, equivalente a una tasa de crecimiento anual compuesto del 25.2%.

En el caso particular de la educación superior, los LMS constituyen una plataforma que busca permitir una mejor comunicación entre los estudiantes, el equipo docente y la institución. Corresponden a un canal mediante el cual los actores pueden intercambiar información, tanto con objetivos académicos como sociales, de forma asíncrona y en cualquier instante. Dentro de las actividades que un alumno de algún curso puede realizar en un LMS, si este cuenta con la funcionalidad, se incluyen:

- Compartir su propio material con el resto del curso.

otros criterios son evaluados.

²Fuente: ingeniería.uchile.cl; consultado el 17/12/2016.

³Fuente: educacion.uc.cl; consultado el 17/12/2016.

⁴Fuente: LMS Industry User Research Report. Consultado el 22/04/2017 y disponible en <http://www.capterra.com/learning-management-system-software/user-research>

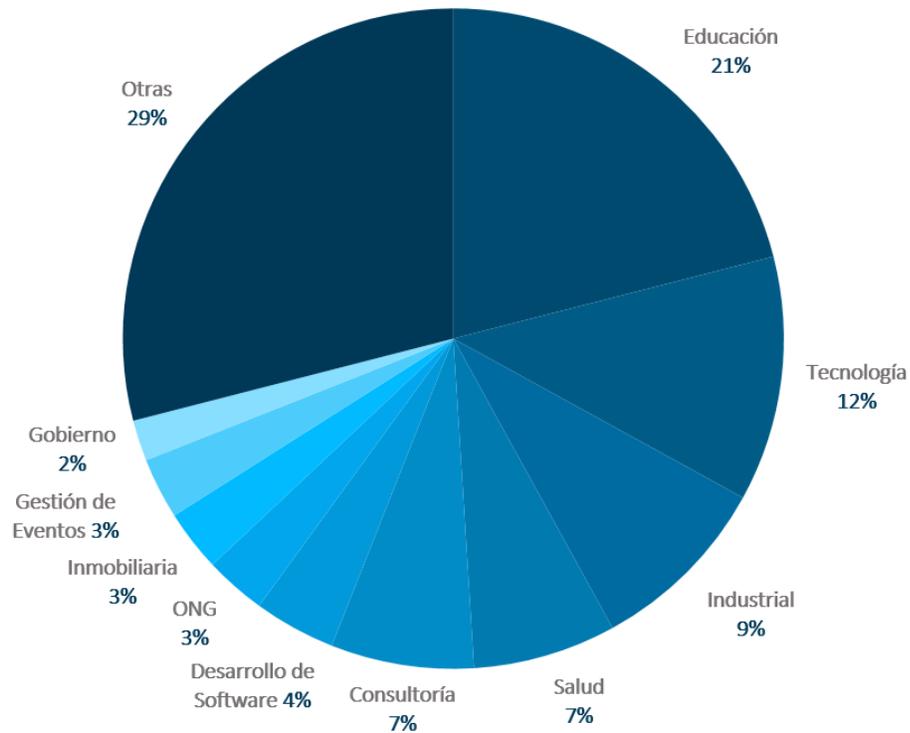


Fig. 2.1: Uso de los LMS por industria. Fuente: traducido de Capterra.com

- Contactar directamente al equipo docente o a compañeros.
- Descargar material subido por el cuerpo docente o por otros compañeros.
- Entregar tareas.
- Evaluar a los docentes anónimamente.
- Interactuar públicamente con sus compañeros y profesores en el foro del curso.
- Participar en votaciones si el equipo docente así lo solicita.
- Visualizar su asistencia presencial a clases, así como los resultados de sus evaluaciones y las estadísticas del desempeño del curso.

2.3. U-Cursos y Sakai

Al año 2017, la Universidad de Chile y la Universidad Católica utilizan principalmente dos plataformas de LMS: U-Cursos y Sakai, respectivamente, las que se pueden acceder a través de un computador, o utilizando las versiones para móviles de los LMS.

Por un lado, U-Cursos corresponde a un servicio desarrollado por el Área de Infotecnologías (ADI)⁵ de la FCFM que le entrega a cada curso un sitio en Internet con numerosas

⁵El equipo desarrollador ahora es conocido como U-Campus.

herramientas o módulos para apoyar el aprendizaje⁶ (Ver Figura 2.2 con un ejemplo de módulos de U-Cursos), haciendo de nexo entre los estudiantes y el equipo docente, y contando con amplio uso dentro de la comunidad académica. Una vez creado el sitio web para cada curso, los profesores pasan a ser los administradores, contando con la posibilidad de activar los módulos que deseen y añadir a su equipo docente, así como fomentar y supervisar el uso del sitio.

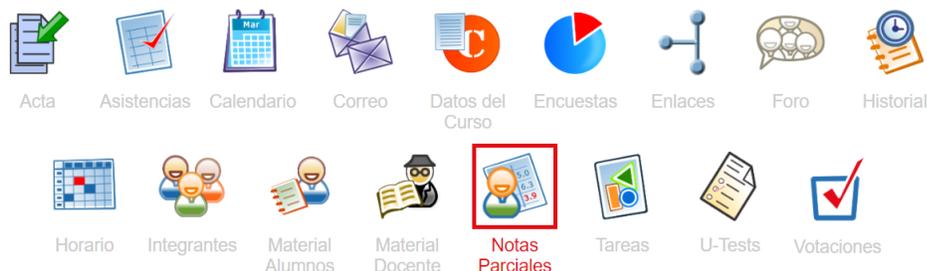


Fig. 2.2: Módulos activos de un curso de la plataforma U-Cursos.

Por otro lado, la PUC utiliza Sakai, un LMS de código abierto programado en Java, el cual ha sido desarrollado gracias a la colaboración de la comunidad educativa de distintos lugares del mundo. En el proyecto Sakai, actualmente en su versión 11, han colaborado más de 350 instituciones⁷, se encuentra disponible en más de 20 idiomas y ofrece una amplia gama de soluciones para facilitar el aprendizaje online de los estudiantes. Una captura de pantalla de la interfaz de Sakai se presenta en la figura 2.3.



Fig. 2.3: Extracto de la interfaz de Sakai en la UC.

⁶Fuente: u-cursos.cl; consultado el 17/12/2016.

⁷Fuente: sakaiproject.org; consultado el 22/04/2017.

2.4. Proyecto FONDECYT #1161413

Esta memoria se enmarca dentro del proyecto de investigación FONDECYT #1161413⁸ “*Integrando Learning Analytics y la experiencia de aprendizaje de estudiantes universitarios de primer año para mejorar sus procesos y resultados de aprendizaje.*” del Centro UC de Estudios de Políticas y Prácticas en Educación (CEPPE), del Centro de Investigación Avanzada en Educación de la Universidad de Chile (CIAE) y de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile.

Learning Analytics⁹ es una disciplina que tiene por objetivo apoyar el aprendizaje de los estudiantes utilizando los datos generados a partir de las plataformas digitales de las instituciones de educación superior durante la última década. En este sentido, LA ha avanzado en la línea del análisis e interpretación de estos datos con el fin de predecir el desempeño académico e identificar alumnos en riesgo de reprobación, buscando poder generar intervenciones adecuadas en los estudiantes.

Por otro lado, para comprender la experiencia de aprendizaje de los estudiantes es necesario mirar la disciplina de investigación en educación superior tradicional, Student Learning Research (SLR)¹⁰, la cual ha descubierto importantes hallazgos. Esta área de la investigación educativa se centra en el desarrollo y la validación de teorías de aprendizaje y enseñanza usando métodos propios de las ciencias sociales, donde cada una de las hipótesis investigadas se analiza desde una perspectiva teórica y práctica, que sientan la base para la elaboración de metodologías para la enseñanza y la planificación curricular.

Si bien ambas disciplinas han producido importantes resultados, también han recibido críticas por distintas razones. En primer lugar, LA no ha podido establecer una relación clara con las teorías tradicionales de aprendizaje, pues los datos que se utilizan en modelos y algoritmos pueden carecer de interpretación para entender la forma en la que los estudiantes aprenden (Dringus, 2012). En segundo lugar, se ha criticado a SLR por fundamentar sus hallazgos principalmente en percepciones auto reportadas por los actores involucrados (Miller, 2012). En este sentido, se plantea que combinar percepciones con datos provenientes de interacción con sistemas digitales puede ayudar a superar estas debilidades, logrando una visión integrada del proceso de aprendizaje de los estudiantes.

De esta forma, la investigación propuesta del proyecto tiene por objetivo general “*Caracterizar los procesos de aprendizaje y predecir los resultados académicos de estudiantes de primer año de dos carreras profesionales (ingeniería y pedagogía) para generar un modelo que integre teorías, constructos y herramientas metodológicas provenientes de Learning Analytics y Student Learning Research, con el objetivo de apoyar y mejorar sus aprendizajes.*” y, en un plazo de 4 años, se articula en 3 etapas:

1. **Etapa Retrospectiva:** enfocada en trabajar con datos de estudiantes de primer año cohortes 2013-2014, con el objetivo de caracterizar, predecir resultados y generar hipótesis relativas a sus procesos de aprendizaje, así como de elaborar, adaptar y pilotear

⁸La fuente de esta sección es el documento de propuesta de investigación de dicho proyecto.

⁹Ver en detalle en la sección 3.2 Learning Analytics y Educational Data Mining.

¹⁰Ver en detalle en la sección 3.1 Student Learning Research.

instrumentos para indagar sobre sus experiencias de aprendizaje.

2. **Etapa Prospectiva:** tiene por fin analizar las experiencias de los estudiantes para describir el contexto de enseñanza-aprendizaje y los enfoques de aprendizaje. Se busca además aplicar el análisis de la etapa anterior con datos levantados para estudiantes de primer año cohortes 2017, 2018 y 2019 con el fin de caracterizar y predecir resultados. Por último, se desea generar un modelo de análisis que combine elementos de LA y SLR con el objetivo de describir las experiencias de aprendizaje de los estudiantes de forma holística y generar información relevante para la toma de decisiones relativa a sus procesos de aprendizaje.
3. **Etapa de Prototipo:** por último, el proyecto busca diseñar y evaluar un prototipo de herramienta web que integre y analice diversas fuentes de información para entregar retroalimentación que apoye la experiencia de aprendizaje de los estudiantes, en base al modelo antes desarrollado.

2.5. Objetivos

Parte del trabajo a realizar en el proyecto FONDECYT corresponde a caracterizar los procesos de aprendizaje de estudiantes en la universidad, lo cual será realizado en la etapa retrospectiva, por un lado, mediante entrevistas a alumnos y la aplicación de una nueva versión del cuestionario LEARN (Parpala y Lindblom-Ylänne, 2012), y por otro lado mediante este trabajo. Teniendo en cuenta la descripción del proyecto realizada anteriormente, los objetivos del trabajo de memoria se detallan a continuación.

2.5.1. Objetivo General

Realizar un análisis comparativo del uso de dos plataformas de gestión del aprendizaje, con el fin de determinar la relación entre la actividad online y el rendimiento académico de estudiantes de contextos educacionales distintos.

2.5.2. Objetivos Específicos

1. Analizar las bases de datos de la actividad online de estudiantes en U-Cursos (UChile) y Sakai (UC), e identificar factores que expliquen similitudes o diferencias en el uso de estas plataformas entre distintos grupos de estudiantes y en distintos contextos académicos.
2. Caracterizar el desempeño académico de estudiantes universitarios de primer año con información sociodemográfica, historial académico e información sobre su comportamiento en una plataforma LMS.

3. Relacionar la actividad online de los estudiantes y su desempeño académico con las teorías de aprendizaje tradicionales.
4. Entregar herramientas claras de visualización de la actividad en un LMS, que puedan ser utilizadas en el desarrollo del prototipo.

2.6. Alcances

Como se ha especificado anteriormente, la memoria tiene como objetivo estudiar la actividad online en LMS de estudiantes en cursos de primer año. En este sentido, los alcances y limitaciones del trabajo son las siguientes:

1. El trabajo se limitará a la interacción online de los estudiantes con los LMS, y se descarta estudiar sus procesos de aprendizajes físicos, tanto en el aula como fuera de ella.
2. Se trabajará con datos de los años 2013-2014 de las carreras de Ingeniería UChile y Educación UC. La información generada por los LMS es extensa y detallada, por lo que se descarta trabajar con el fin de generar nuevos datos.
3. El análisis se realizará sólo sobre las versiones de los LMS para computadores. Existe además información generada por versiones para móviles de los LMS, la cual puede presentar similitudes y diferencias con la versión de escritorio. Sin embargo, debido a la complejidad de las dinámicas que pueden existir en los patrones de uso de la versión móvil, se descarta analizarla en esta instancia.

Capítulo 3

Marco Conceptual

Este capítulo presenta, por un lado, los principales elementos teóricos revisados relativos a tres disciplinas distintas de investigación en educación que buscan apoyar y mejorar el aprendizaje de los estudiantes; una con un enfoque más tradicional, Student Learning Research; y dos que utilizan la gran disponibilidad de datos existente actualmente, Educational Data Mining y Learning Analytics. Además se presentan estudios anteriores que investigan, utilizando información auto reportada, los factores determinantes del uso de LMS por parte de miembros de distintas comunidades académicas.

Por otro lado, se señalan también a continuación algunos elementos estadísticos y econométricos que se utilizan en capítulos posteriores. No se busca realizar un detalle exhaustivo, sino señalar las características principales necesarias para comprender el trabajo realizado.

3.1. Student Learning Research

Student Learning Research es una línea de investigación que utiliza en su análisis tanto métodos cuantitativos como cualitativos. Ha logrado importantes resultados sobre el aprendizaje y la docencia en educación superior, los cuales han servido como fundamento de programas de desarrollo académico para profesores y para ayudar a los estudiantes a comprender mejor sus procesos de aprendizaje (Trigwell et al., 2012).

Una de las teorías más importantes que ha desarrollado esta disciplina es la de Marton y Säljö (1976), y luego Ramsden (1979), la cual establece que los estudiantes abordan el aprendizaje principalmente de tres formas distintas: superficial, profunda y estratégica, las cuales se vinculan con su percepción sobre el contexto de aprendizaje en el que se encuentran y los conducen a distintos resultados académicos. El aprendizaje **superficial** corresponde a un enfoque en memorizar y asociar contenidos sin comprenderlos en detalle, con el fin de responder de buena manera a evaluaciones, generalmente considerando el trabajo como una obligación. Abordar el aprendizaje de forma **profunda** hace referencia a entender el significado de los contenidos que se estudian, a relacionarlos con conocimientos previos y a asociar la teoría con la práctica. De esta forma, se perciben los contenidos de forma íntegra y coherente.

Por último, el aprendizaje del tipo **estratégico** corresponde al cual los estudiantes definen estrategias y organizan su esfuerzo y tiempo de forma tal de obtener mejores resultados.

Tradicionalmente, las distintas percepciones de los procesos de aprendizaje se miden a través de encuestas, en las cuales se interroga a estudiantes sobre su contexto académico y experiencias. De esta forma, los instrumentos buscan características como por ejemplo el enfoque de aprendizaje, el apoyo de pares, de los profesores, la retroalimentación recibida, la carga de trabajo, las creencias de autoeficacia (self-efficacy beliefs), la calidad de las instalaciones y recursos a su disposición, el apoyo familiar, el uso de los LMS, entre otros. Sin embargo, la aplicación de este tipo de instrumentos es costosa y usualmente distanciada en el tiempo, por lo que es difícil utilizar datos medidos de esta manera para dar un seguimiento continuo a los estudiantes.

En el ámbito de la enseñanza, una de las teorías más conocidas de SLR plantea que los profesores pueden hacer docencia centrándose en los contenidos, donde el profesor asume un rol protagónico con la función de transmitir sus conocimientos; o bien centrándose en el aprendizaje, donde el foco está en el desarrollo del pensamiento crítico y reflexivo del estudiante, más que en los contenidos que adquiere. Se ha determinado que la enseñanza basada en el aprendizaje permite a los estudiantes adoptar un enfoque de tipo profundo, y a obtener mejores desempeños académicos (Prosser y Trigwell, 1999). Por esta razón, el movimiento conocido como Scholarship of Teaching and Learning promueve el trabajo y la mejora continua de la docencia enfocada en los estudiantes y sus procesos de aprendizaje.

La importancia de entender los procesos de aprendizaje y enseñanza, es que estos entregan las bases sobre las cuales se diseñan e implementan los programas académicos. Es así como se han desarrollado metodologías basadas en el paradigma conductista, el cognitivo, el constructivista y el basado en competencias. Tal como fue señalado anteriormente, a pesar de los logros de SLR al explicar cómo aprenden los estudiantes, ha recibido críticas sobre el gran peso que le otorga a lo psicológico en las teorías de aprendizaje (Malcolm y Zukas, 2001), así como a fundamentar sus hallazgos en información auto declarada y subjetiva (Miller, 2012). Esto último puede dar lugar a sesgo por deseabilidad social, donde las personas buscan responder lo que creen que el experimentador quiere que respondan. Hoy en día, gracias a la gran cantidad de información generada tras el auge de los MOOC y LMS, existe la oportunidad de revisar y validar las teorías de aprendizaje que analiza SLR desde un enfoque distinto, utilizando la información que estas plataformas entregan.

3.2. Educational Data Mining y Learning Analytics

Tanto Educational Data Mining (EDM) como Learning Analytics son disciplinas emergentes que han tenido su origen en la masificación del uso de la tecnología en educación durante la última década (Ver Fig. 3.1). Por una parte, EDM se enfoca en el desarrollo y la utilización de metodologías propias de la minería de datos para explorar grandes cantidades de información con características similares —proveniente de un contexto educacional— para comprender el desarrollo y el aprendizaje de los estudiantes. Las herramientas, métodos, y algoritmos que se utilizan son variados y han sido desarrollados para satisfacer necesidades

específicas del mundo educativo, utilizando información generada por sistemas educacionales y reportando los descubrimientos a estudiantes, profesores y a la institución educativa (Ver el ciclo de data mining en un sistema educacional en la Fig. 3.2). Entre ellas destacan (Romero et al., 2008):

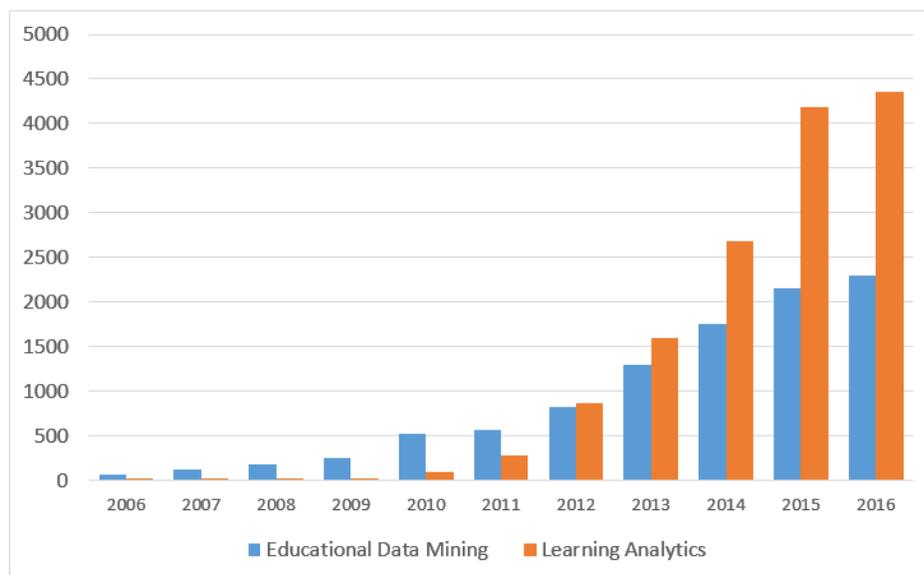


Fig. 3.1: Referencias a Educational Data Mining y Learning Analytics por año en Google Scholar. Fuente: Elaboración Propia¹.

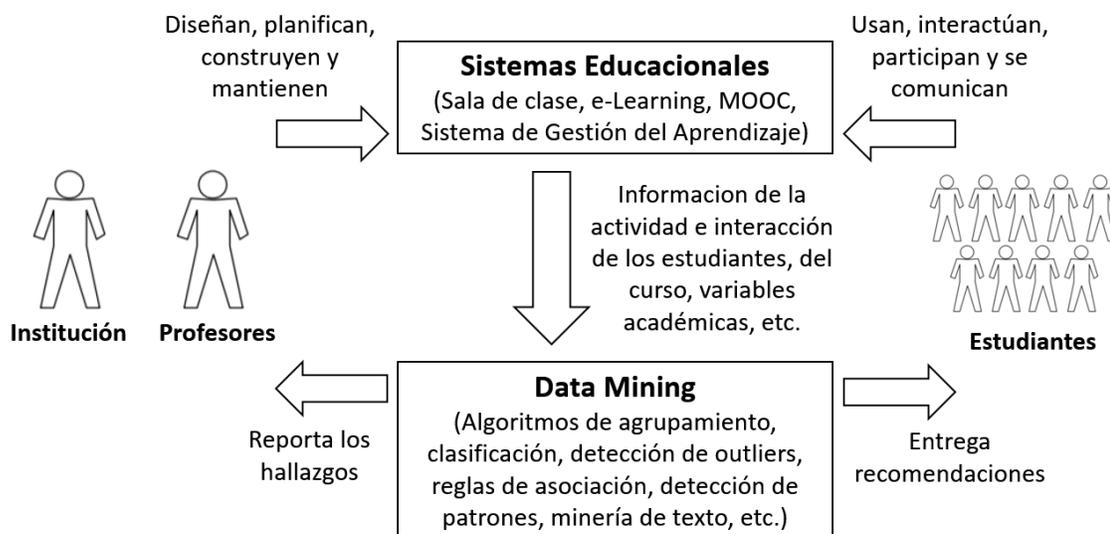


Fig. 3.2: Ciclo de Data Mining en un sistema educacional. Fuente: Romero y Ventura (2007)

- Estadística:** El análisis estadístico de las principales variables recolectadas son frecuentemente el punto de partida en EDM. En el caso de los LMS se pueden estudiar variables como el número de conexiones, número de actividades en cada módulo, duración de la visita, distribución de uso, etc. También se pueden utilizar técnicas como análisis de la varianza, de correlación y de regresión.

¹Elaborada en base a figura 3 de Calvet Liñán y Juan Pérez (2015).

- **Clasificación:** Herramientas como la regresión logística, árboles de decisión, redes neuronales o máquinas de soporte vectorial han sido utilizadas para predecir el desempeño académico de estudiantes y su aprobación o reprobación de un curso, permitiendo a docentes una identificación temprana de alumnos en situación de riesgo.
- **Clustering:** Es el proceso de agrupar objetos similares en una misma clase. En educación, se ha utilizado para identificar alumnos según su trayectoria académica (Castillo, 2016) y también puede usarse para agrupar alumnos con patrones de comportamiento, habilidades, conocimientos o intereses similares con el fin de diseñar intervenciones específicas.
- **Visualización:** Permite el análisis de una gran cantidad de información representando los datos de forma gráfica. La visualización en tiempo real entrega a los profesores y a la institución una importante herramienta para detectar problemas en un sistema educacional y tomar medidas para solucionarlos. Las formas de visualizar la información son diversas, y entre ellas destacan los gráficos de dispersión, de barras, circulares, histogramas, diagramas de caja, entre otros.

A modo general, el proceso que sigue una aplicación de Data Mining y que será utilizado en este trabajo es el de Knowledge Discovery in Databases (KDD, por sus siglas en inglés), el cual consiste en (1) la selección de los datos a utilizar, (2) su preprocesamiento (limpieza, normalización, imputación, etc.), (3) la transformación de los datos en un formato adecuado para extraer patrones, (4) la minería de datos propiamente tal, y (5) la interpretación y evaluación de los patrones descubiertos.

Por otro lado, Learning Analytics es una comunidad de investigación que tiene su origen en el aumento de interés por el mundo del Big Data y el área de Business Intelligence (Ferguson, 2012), la cual utiliza la gran cantidad de información existente en la toma de decisiones para la empresa. Learning Analytics busca servirse de la gran recolección de datos en educación superior de la última década, para mejorar los procesos educativos y diseñar intervenciones para facilitar el aprendizaje de los alumnos. Es así como intervenciones de Learning Analytics en instituciones educativas pueden, por ejemplo:

- Identificar a los estudiantes que requieren atención y apoyo adicional (Macfadyen y Dawson, 2010).
- Reconocer los cursos que tienen mayor relevancia en el éxito de los estudiantes en la carrera (Treaster, 2017).
- Permitir la visualización de forma agregada de los eventos que realizan los estudiantes en una plataforma online (Leony et al., 2012).
- Identificar a los estudiantes que podrían caer en causal de eliminación de la universidad por motivos académicos (Celis et al., 2015).
- Retroalimentar a profesores sobre la situación de un curso en base a datos sobre la interacción online de los estudiantes (Mazza y Dimitrova, 2004).

3.3. Los LMS en Educación Superior

Dada las diferencias presentes en las distintas comunidades académicas y en los diversos sistemas de gestión del aprendizaje, resulta natural suponer la existencia de múltiples factores que determinan el grado de aceptación y uso que los miembros de una institución dan a su LMS. Al respecto, algunos autores han llevado a cabo estudios utilizando información auto reportada. Si bien este trabajo propone considerar datos de los logs de su actividad online, sus investigaciones resultan claves para entender el rol de las plataformas para estudiantes, profesores y las instituciones.

Fathema et al. (2015) investiga los factores que afectan el comportamiento de 560 académicos de dos universidades en un LMS utilizando una extensión del modelo de aceptación de tecnologías (TAM). Los resultados validan el modelo propuesto, que en este contexto se resume de la siguiente manera: En primer lugar los académicos evalúan la dificultad de trabajar con el LMS, luego, evalúan la utilidad de hacerlo. Si resulta ser fácil de usar y útil, entonces desarrollan una actitud positiva frente a él. Esta actitud los lleva a tener una intención positiva de usarlo. Finalmente, esta intención influye en el real uso del LMS. Las variables determinantes de este proceso son tanto las creencias de autoeficacia sobre la capacidad del uso de la plataforma, como la calidad de la misma y sus recursos relacionados (ayuda técnica, infraestructura, hardware, software, etc.).

Tandi Lwoga (2014) por su parte, examina los factores que determinan la intención de uso de un LMS por parte de 272 estudiantes de la Muhimbili University of Health and Allied Science realizando una encuesta y utilizando ecuaciones estructurales. Los resultados indican que factores relacionados con la calidad (del profesor y del LMS) son claves en la percepción de la utilidad de la plataforma y en la satisfacción del usuario. En particular, elementos relativos a la calidad de la plataforma (como tiempo de respuesta, interfaz, diseño, funcionalidades, etc.) influyen fuertemente en la intención de uso de los estudiantes.

Por otro lado, Naveh et al. (2012) encuesta a 8.425 estudiantes y entrevista a 40 para identificar los determinantes de su nivel de satisfacción al usar la plataforma. Sus resultados apuntan a 5 elementos: accesibilidad del contenido (posibilidad de acceder a la totalidad de la información en cualquier momento y lugar), validez del contenido (que la información tanto académica como administrativa este actualizada), facilidad de navegación (relativo a la organización de los distintos aspectos de un curso dentro del LMS), facilidad de acceso (simpleza en la autenticación, velocidad de carga de LMS, interfaz, etc.), y nivel de actividad del equipo docente (percepción del LMS como medio de comunicación con el equipo docente de cada curso).

El objetivo de los trabajos revisados es identificar las razones del comportamiento de usuarios en un LMS. Dado que los factores existen a nivel de cada usuario (como las creencias de autoeficacia), del curso en cuestión (material disponible, actividad del equipo docente, etc.), o de la plataforma (calidad, facilidad de acceso, etc.), se pueden concebir distintos tipos de estrategias, en los distintos niveles mencionados, para aumentar la intención y satisfacción de uso de los LMS. Este trabajo recoge los hallazgos de los autores mencionados para analizar desde un punto de vista cuantitativo el comportamiento en las plataformas.

3.4. Elementos Estadísticos

Para caracterizar e identificar distintos grupos de estudiantes según los valores de distintas variables es necesario utilizar diferentes herramientas estadísticas. A continuación, se presentan de forma breve los elementos más importantes utilizados en el análisis de las bases de datos y descripción de la muestra: el test de ANOVA, de Levene, de Tukey Honest Significant Difference (HSD) y de Games-Howell.

3.4.1. Análisis de la Varianza

El análisis de varianza (ANOVA, por sus siglas en inglés) es un conjunto de modelos estadísticos que, entre otros, provee un test que permite determinar si las medias de distintos grupos son diferentes o no, a una cierta significancia estadística. A continuación, se presentan las principales hipótesis en las que este se sustenta, así como posibles soluciones cuando estas no se cumplen.

1. La distribución de los residuos dentro de cada grupo distribuye de forma normal. ANOVA resiste bien a no normalidad. Sin embargo, la platicurtosis² puede tener un efecto negativo en los resultados si la muestra es pequeña. Las opciones para solucionar este problema incluyen transformar la data, o bien usar el test no paramétrico de Krushal-Wallis H.
2. La varianza de los grupos es la misma (homocedasticidad). Si no se cumple este supuesto puede usarse Welch-ANOVA, la cual es robusta a heteroscedasticidad.
3. Las variables son independientes. Esta hipótesis corresponde a una de las más fuertes en las que se basa el análisis, por lo que no debe usarse ANOVA si existen fuertes razones para creer que los datos no son independientes. La opción es realizar el análisis en una submuestra donde las observaciones efectivamente sean independientes.

3.4.2. Test de Levene

El test de Levene es una herramienta utilizada para determinar si las varianzas de dos o más grupos son iguales o no, a un cierto nivel de significación. Algunos análisis (como ANOVA) asumen que las varianzas de las poblaciones en estudio son iguales (supuesto de homocedasticidad). El test de Levene es útil entonces para evaluar esta hipótesis y, en caso de no cumplirse, utilizar otro tipo de test libres de este supuesto. Si el p-valor del test de Levene es menor que un nivel de significación (en este estudio 0,05), se rechaza la hipótesis nula de homocedasticidad y se concluye que existe una diferencia significativa entre las varianzas de los grupos en la población.

²Una distribución con un grado de curtosis menor a 3 se dice platicúrtica y tiene una forma ‘aplanada’.

3.4.3. Post hoc tests de Tukey HSD y Games-Howell

ANOVA y Welch-ANOVA detectan si existen diferencias significativas entre las medias de distintos grupos, pero no indican cuales de ellos. Para esto se puede proceder a realizar un test post hoc para identificar los pares de grupos que tengan medias diferentes, donde puede utilizarse el test de Tukey HSD en caso de que se cumpla la hipótesis de homocedasticidad (ANOVA), o bien Games-Howell en caso de heteroscedasticidad (Welch-ANOVA). La hipótesis nula de este último test es H_0 : Las medias de ambos grupos son iguales.

3.5. Elementos Econométricos

Para evaluar el poder explicativo de variables relacionadas con la actividad online de estudiantes en el desempeño académico es posible utilizar un modelo tal como una regresión lineal y evaluar la significancia de los parámetros asociados a estas variables explicativas. En esta sección se describen algunos estimadores de una regresión lineal que se pueden usar en distintas situaciones (MCO, Eicker-White, y de clústers); se presenta además la VIF, un valor utilizado para identificar regresores problemáticos; y se introduce una medida de evaluación de la calidad del modelo, la bondad de ajuste, así como dos tests que permiten comparar la diferencia de esta medida entre distintos modelos.

3.5.1. Notación

La regresión lineal es un modelo útil para explicar la relación entre una variable dependiente Y y una o más variables independientes X_j , también conocidas como variables explicativas, predictores, o regresores. Si p es el número de regresores del modelo y N el número de observaciones, entonces la muestra se denota $\{Y_i, (X_{i1}, X_{i2}, \dots, X_{ip})\}_{i=1}^N$ y el modelo se puede escribir del siguiente modo:

$$Y_i = \beta_0 \cdot 1 + \beta_1 \cdot X_{i1} + \beta_2 \cdot X_{i2} + \dots + \beta_p \cdot X_{ip} + \varepsilon_i \quad (3.1)$$

Donde β_0 es conocido como intercepto, β_j con $j \in \{1, \dots, p\}$ los coeficientes de cada regresor y ε_i representa el término aleatorio del error. En notación matricial esto es:

$$y = X\beta + \varepsilon \quad (3.2)$$

3.5.2. Estimador de Mínimos Cuadrados Ordinarios

El teorema de Gauss-Markov muestra que —bajo ciertos supuestos en una regresión lineal— el mejor estimador lineal insesgado de los parámetros β es el de Mínimos Cuadrados Ordinarios (MCO) $\hat{\beta}_{MCO} = (X'X)^{-1}X'y$, que tiene una varianza³ dada por

³La varianza del estimador permite determinar intervalos de confianza y evaluar si los parámetros estimados son significativamente distintos de 0 o no.

$$\hat{V}_{MCO}(\hat{\beta}) = \frac{1}{n-p} (\hat{\varepsilon}'\hat{\varepsilon})(X'X)^{-1} \quad (3.3)$$

Donde $\hat{\varepsilon} = y - X\hat{\beta}$. Estos supuestos son:

1. El modelo está correctamente especificado, es lineal en los parámetros y no existen variables omitidas.
2. No hay multicolinealidad perfecta; $\text{rang}(X) = p$
3. Los errores son independientes e idénticamente distribuidos; $\varepsilon_i \sim \text{i.i.d.}$
4. La esperanza del término del error es cero; $\mathbb{E}[\varepsilon_i] = 0, \forall i$
5. Los errores son homocedásticos; $V(\varepsilon_i) = \sigma^2, \forall i$
6. Los errores no están correlacionados con las variables explicativas; $X_{ip} \perp \varepsilon_i, \forall i, p$
7. Los errores distribuyen de forma normal. Técnicamente esta no es una hipótesis de MCO, pero necesaria para realizar inferencia sobre los parámetros estimados (tests de hipótesis).

3.5.3. Estimador de Eicker-White y de Clústers

En la práctica, las hipótesis del teorema de Gauss Markov no siempre se cumplen, por lo que se utiliza al estimador de MCO como punto de partida para trabajar con otros estimadores en caso de a violaciones de los supuestos. En este trabajo se utiliza el estimador de la varianza de Eicker-White (White, 1980), más resistente a no normalidad de los errores y heteroscedasticidad que el estimador de MCO, dado por:

$$\hat{V}_{EW}(\hat{\beta}) = \frac{n}{n-p} (X'X)^{-1} \left(\sum_{i=1}^N (\hat{\varepsilon}_i x_i)' (\hat{\varepsilon}_i x_i) \right) (X'X)^{-1} \quad (3.4)$$

Donde $\hat{\varepsilon}_i$ es el residuo de la i -ésima observación y x_i es la i -ésima fila de X . También se utiliza un estimador que generaliza al estimador de Eicker-White para el caso en que existan n_c grupos de observaciones (clústers) no independientes (Rogers, 1993):

$$\hat{V}_{cluster}(\hat{\beta}) = \frac{n-1}{n-p} \frac{n_c}{n_c-1} (X'X)^{-1} \left(\sum_{j=1}^{n_c} u_j' \cdot u_j \right) (X'X)^{-1} \quad (3.5)$$

Donde $u_j = \sum_{i \in \text{cluster}_j} \hat{\varepsilon}_i \cdot x_i$

3.5.4. Factor de Inflación de la Varianza

El factor de inflación de la varianza (VIF por sus siglas en inglés) es una medida de la multicolinealidad de las variables dependientes en una regresión lineal. Proporciona un índice que mide hasta qué punto la varianza de un coeficiente de regresión estimado se incrementa a causa de la colinealidad con otras variables. Una regla general es que para un parámetro estimado $\hat{\beta}_i$, una $VIF(\hat{\beta}_i) > 5$ debe ser analizada y una $VIF(\hat{\beta}_i) > 10$ indica una alta colinealidad con otros parámetros, lo que aumenta la varianza de los estimadores de la regresión. La medida más simple a tomar consiste en excluir a las variables colineales de la regresión, dado que su poder explicativo es redundante con otras variables del modelo.

3.5.5. Evaluación de un modelo: Bondad de Ajuste

La medida tradicional del ajuste de un modelo corresponde al R^2 , llamado también bondad de ajuste o coeficiente de determinación. Para las regresiones lineales, el R^2 representa la proporción de la varianza total explicada por el modelo y es ampliamente usada para medir el nivel predictivo de la regresión. Sin embargo, este no corresponde y no debe usarse como una medida de la significancia de los predictores. Esto se ejemplifica en el trabajo de Abelson (1985) donde se estudia mediante una regresión lineal una liga de béisbol con el fin de determinar la relevancia del promedio de bateo histórico para predecir si el siguiente bateo de un jugador será acertado o no. A pesar de que el promedio de bateo histórico es considerado un buen proxy de la habilidad del jugador, y por tanto una característica de éxito, dicha regresión obtuvo un R^2 de tan solo 0,003.

Debido a que el R^2 es creciente en la cantidad de parámetros, en la práctica suele ser conveniente trabajar con una medida conocida como R^2 ajustado, la cual corrige la bondad de ajuste por la cantidad de variables explicativas del modelo.

3.5.6. Test de Razón de Verosimilitud y Test de Wald

El test de razón de verosimilitud se utiliza para comparar la bondad de ajuste entre dos modelos, uno de los cuales se encuentra anidado en el otro, y permite testear cual ajusta mejor a los datos, considerando la cantidad de parámetros adicionales en uno de ellos. Equivalentemente, el test de Wald también puede utilizarse para verificar esta hipótesis⁴. Cuando existen modelos anidados y se calcula una matriz de varianza covarianza con el estimador de clústers, es recomendable utilizar el test de Wald y no el test de razón de verosimilitud, en cuanto este último asume independencia entre las variables.

⁴Utilizar el test de razón de verosimilitud y verificar si el modelo con más parámetros ajusta mejor a los datos es asintóticamente equivalente a usar el test de Wald y verificar si los parámetros extra de uno de los modelos son iguales a 0 (Engle, 1984).

3.6. Conclusiones Marco Conceptual

Con el fin de determinar la relación entre la actividad online y el rendimiento académico en contextos académicos distintos, se han presentado en este capítulo tres distintas disciplinas desde las cuales es posible abordar los procesos de aprendizaje de estudiantes. Se espera realizar un análisis que integre conceptos y metodologías propias a cada una de ellas, con el fin de cumplir los objetivos propuestos. En particular, se piensa hacer uso de herramientas estadísticas, de visualización y de regresión para estudiar las formas en que los estudiantes abordan sus aprendizajes. Con las herramientas mencionadas se espera formalizar cuantitativamente los hallazgos encontrados.

Capítulo 4

Metodología

El equipo que trabaja en el proyecto elaboró un repositorio de datos con información sociodemográfica¹ y de actividad en LMS de 1.940 estudiantes de primer año cohortes 2013-2014: 1.465 de ellos pertenecientes al Plan Común de Ingeniería y Ciencias de la FCFM en la UChile, 138 alumnos de Pedagogía en Educación Parvularia, y 337 de Pedagogía General Básica, ambas carreras de la facultad de Educación UC.

En este capítulo se describe el origen de los datos utilizados y el proceso de limpieza llevado a cabo (preprocesamiento) tanto para los datos sociodemográficos y de actividad online, teniendo en cuenta el contexto en el que se producen y los lineamientos de la literatura revisada. Luego, se describe el trabajo a desarrollar en materia de visualización y el modelo explicativo.

4.1. Origen de los datos

En una fase inicial del proyecto, el equipo recolectó la información necesaria para construir el repositorio de distintas fuentes, entre las que se encuentran:

- La base de datos sociodemograficos de la UChile producida en el marco del proyecto Basal 2012 UCH1898, del que uno de los integrantes del equipo formó parte.
- El ADI, perteneciente a la FCFM, quien dio acceso a las interacciones de los estudiantes de interés con la plataforma de LMS U-Cursos de la UChile.
- La base de datos sociodemográfica, de rendimiento y de la actividades en la plataforma web Sakai de la UC consolidada en el marco del proyecto Fondecyt Regular #1131109, llevado a cabo por uno de los miembros del equipo.

El procesamiento de los datos sociodemográficos y de la actividad online se describen a continuación.

¹Donde se incluye el desempeño académico anterior a entrar a la Universidad y al final de cada semestre.

4.2. Datos sociodemográficos

La base de datos producida por el equipo cuenta con información sociodemográfica y de desempeño de cada estudiante de la población, donde cada observación corresponde a la información de un alumno en un semestre. Debido a que esta base de datos contiene información de dos universidades distintas, el equipo creó ciertas variables a fin de homogeneizar la información, lo cual se detalla a continuación:

- Creación de la variable psu_{esp} definida como el máximo entre psu_{hist} y psu_{cien} . Esta variable representa el máximo entre las dos pruebas PSU específicas, historia y ciencias, rendidas por el estudiante y es útil para realizar comparaciones entre quienes rindieron distintas pruebas.
- Se crea la variable categórica ‘ $\text{ingreso}_{\text{categoría}}$ ’ con valores correspondientes a 3 categorías para el tipo de ingreso a la universidad, creando una codificación única: **Regular**, correspondiente al ingreso por una vía normal a la universidad (prueba de selección); **Especial**, para estudiantes que ingresan de una forma particular a la universidad, como deportistas, estudiantes extranjeros, cambio de carrera o universidad, o buenos estudiantes de colegios vulnerables; y **Frontera**, correspondiente a buenos alumnos que a pesar de no haber sido admitidos en situaciones regulares, son seleccionados bajo criterios especiales. El detalle se presenta en el Anexo A.
- Para poder comparar las notas de alumnos que cursan distintas carreras en distintas universidades, se normalizó la nota tomando como universo cada carrera, definiendo la variable ‘ $\text{zscore}_{\text{carrera}}$ ’ como la posición relativa del alumno respecto a sus compañeros de la carrera-semestre.
- Se crea la variable categórica ‘desempeño’ con el fin de discretizar la variable $\text{zscore}_{\text{carrera}}$ y separar a la población de acuerdo a su desempeño académico. Se define como ‘Bajo’ si $\text{zscore}_{\text{carrera}} < -1$, ‘Promedio’ si $-1 \leq \text{zscore}_{\text{carrera}} \leq 1$, y ‘Alto’ si $1 < \text{zscore}_{\text{carrera}}$.

Tras esto, los datos sociodemográficos corresponden a una tabla que contiene 3.783 observaciones y 27 variables². El detalle de las variables se presenta en la tabla 4.1.

²Donde para la mayoría de los 1940 de los alumnos existe una observación en el semestre 1, y otra en el semestre 2.

Variable	Tipo	NaN	Descripción
id	int	0	Identificador único del alumno.
semester	int	0	Semestre en curso (1 o 2)
sexo	str	0	Masculino (M) o Femenino (F)
año _{ingreso}	int	0	Año de ingreso a la Universidad (2013 o 2014)
dependencia _{colegio}	str	73	Tipo de colegio (Publico, Privado o Subvencionado)
nombre _{colegio}	str	74	Nombre completo del colegio
año _{egreso_colegio}	int	22	Año de egreso del colegio (1981-2013)
nem	int	17	Puntaje de notas de enseñanza media
psu _{mat}	int	44	Puntaje PSU matemática
psu _{len}	int	44	Puntaje PSU lenguaje
psu _{hist}	int	2.868	Puntaje PSU historia
psu _{cien}	int	62	Puntaje PSU ciencias
psu _{esp}	int	62	Máximo entre psu _{hist} y psu _{cien}
puntaje _{ponderado}	int	67	Puntaje PSU ponderado
ingreso	str	0	Tipo de ingreso a la universidad
ingreso _{categoria}	str	25	Categoría del tipo de ingreso a la universidad
cursos _{inscritos}	int	0	Cantidad de cursos inscritos durante el semestre
cursos _{aprobados}	int	0	Cantidad de cursos aprobados durante el semestre
cursos _{reprobados}	int	0	Cantidad de cursos reprobados durante el semestre
créditos _{inscritos}	int	0	Cantidad de créditos inscritos durante el semestre
créditos _{aprobados}	int	0	Cantidad de créditos aprobados durante el semestre
créditos _{reprobados}	int	0	Cantidad de créditos reprobados durante el semestre
carrera	str	0	Nombre de la carrera
nota _{sem}	int	0	Nota final del semestre
zscore _{carrera}	int	0	zscore de nota _{sem} para cada carrera
plataforma	str	0	Plataforma LMS (UCURSOS si estudiante pertenece a UChile y SAKAI si pertenece a PUC)
desempeño	str	0	‘Bajo’ si zscore _{carrera} < -1, ‘Promedio’ si $-1 \leq \text{zscore}_{\text{carrera}} \leq 1$, y ‘Alto’ si $1 < \text{zscore}_{\text{carrera}}$

Tabla 4.1: Descripción de variables sociodemográficas.

4.3. Datos de la actividad online

4.3.1. Variables

El equipo consolidó una base de datos con información de cada clic que realiza un estudiante en su plataforma LMS de ciertos cursos de cada carrera (Ver lista de cursos en Anexo B). Cada observación de esta base de datos contiene información sobre la identificación del alumno, el curso donde fue realizado el clic, la sección a la que pertenece el alumno, el timestamp (la fecha y hora del clic), el semestre académico, la plataforma donde fue realizado el clic (U-Cursos para estudiantes de la UChile y Sakai para la UC) y el evento³.

Un evento corresponde a cualquier actividad que realiza un estudiante en su plataforma, la que puede ser por ejemplo: descargar material, escribir un mensaje en un foro o enviar un correo, entregar una tarea, etc. Cada plataforma registra los eventos de manera diferente: por ejemplo, mientras que en Sakai escribir un mensaje en el foro se codifica como ‘forums.response’, en U-Cursos es ‘Escribir_Mensaje_Foro’. Además, debido a que las funcionalidades de ambas plataformas no son las mismas, hay eventos en un LMS que no tienen símil en el otro. Por estas razones, es necesario primero homogeneizar esta información para poder realizar un análisis de la base de datos. El equipo clasificó esta variable en 5 categorías, en base a juicio experto⁴:

- ‘Academic Content’ (ACA): Eventos que consisten en descargar o acceder a información que tiene relación con las unidades temáticas del curso.
- ‘Administrative Content’ (ADM): Eventos que consisten en descargar o acceder a información que tiene relación con temas administrativos de un curso. Por ejemplo: ver datos del curso, ver notas, ver horario, etc.
- ‘Read Comment’ (RC): Leer comentario en un foro, ver los mensajes de correo, resultados de encuestas, etc.
- ‘Test’ (TEST): Relacionado con descargar o enviar tareas, responder tests online, etc.
- ‘Write Comment’ (WC): Crear o responder un tema en el foro del curso, responder encuestas, subir material para que otros estudiantes puedan descargarlo, etc.

Además, los eventos también fueron clasificados en activos o pasivos, dependiendo del rol que ejerce el estudiante al realizar la actividad. El detalle de ambas clasificaciones se encuentra en el Anexo C⁵.

Por último, se excluyen los eventos registrados antes del inicio de las clases en ambas

³Se descubrió que los logs originales de Sakai tienen un timestamp adelantado de entre 3 y 4 horas debido a que la información se guarda con el timestamp del servidor y no de la hora local, las cuales ambas cambian de acuerdo a los horarios de verano e invierno de los respectivos países. Dada la dificultad de determinar el desfase horario exacto, se decidió corregirlo parcialmente restando 3 horas a todas las observaciones de esta plataforma.

⁴Fuente: Documento del proyecto ‘Documentación LA’.

⁵Debido a la baja cantidad de observaciones de algunos tipos de eventos y la dificultad para catalogarlos, estos fueron descartados de la base de datos.

universidades. Después de este trabajo de preprocesamiento, los datos de la actividad online corresponden a una tabla de $\sim 0,4$ GB que consta de 4.780.064 filas y 9 variables, las cuales se detallan en la tabla 4.2.

Variable	Tipo	NaN	Descripción
id	int	0	Identificador único del alumno
course	str	0	Curso en el que se realiza el evento
section	int	0	Sección del curso
timestamp	datetime	0	Fecha y hora en el que se realiza el evento
semester	int	0	Semestre académico en el que se realiza el evento (1 o 2)
plataforma	str	0	Plataforma LMS (UCURSOS si estudiante pertenece a UChile y SAKAI si pertenece a PUC)
event	str	0	Evento de la observación
category	str	0	Tipo de categoría del evento
action	str	0	Acción realizada (Activa o Pasiva)

Tabla 4.2: Descripción de variables LMS.

4.3.2. Agrupación de los datos en bloques de tiempo

Un elemento importante a considerar al realizar el análisis de las observaciones registradas por parte de un estudiante, es el hecho que puede interactuar en un LMS sin prestar atención a la actividad que realiza, lo que se conoce como ‘Off-task Behaviour’. A modo de ejemplo, Beck (2004) estudia la rapidez de respuesta de estudiantes ante preguntas de selección múltiple elementales, en un computador. Sus resultados muestran que los estudiantes requieren en promedio al menos 4 segundos para procesar la pregunta y responder correctamente. En esta situación, las actividades off-task hacen referencia a estudiantes que responden en un tiempo menor, lo que lleva a respuestas erróneas.

En el caso de los LMS, el off-task behaviour corresponde principalmente a clics realizados en un intervalo de tiempo muy pequeño. Dado el nivel de complejidad de las distintas acciones que se pueden realizar en las plataformas, se decidió solo contabilizar observaciones (clics en el LMS) para un mismo estudiante, dentro de un mismo curso y para la misma categoría de evento, que se encuentran distanciadas temporalmente una de otra por más de $\tau = 10$ segundos⁶. Si una observación se encuentra en un intervalo de tiempo inferior a este de la observación anterior, entonces es catalogada como off-task, representando que la acción realizada no constituye una interacción efectiva con el LMS. Las actividades off-task se analizan como una más de las 5 categorías de eventos antes definidas. De esta forma, de las 4.796.614 observaciones correspondientes a la actividad online, 1.247.186 son clasificadas como Off-task. El detalle por categoría y por plataforma se presenta en la tabla 4.3 a continuación.

⁶Esto es aún un tema de discusión dentro del equipo. Sin embargo, se decidió usar este valor en este trabajo pues es una buena estimación del tiempo necesario para procesar información simple. Además, utilizar 10 segundos entrega una cantidad de observaciones catalogadas como offtask del 26,3%, cantidad razonable de observaciones a separar de la base de datos. τ puede modificarse para realizar pruebas de robustez.

	U-Cursos	Sakai
ACA	676.444	316.744
ADM	424.530	43.709
RC	2.003.956	18.104
TEST	8.852	4.125
WC	51.379	1.585
Off-task	1.194.731	52.455
Total	4.359.892	436.722

Tabla 4.3: Detalle de la cuenta de actividades.

4.3.3. Generación de variables de comportamiento semestral

Como forma de analizar los miles de clics que se realizan en las plataformas, se decidió generar un conjunto de variables que se asocian a cada estudiante-semestre, lo cual corresponde al mismo identificador que define a la tabla de variables sociodemográficas, por lo que es posible combinar estas nuevas variables con las anteriores para realizar inferencias⁷. Esto se explica a continuación.

Para cada semestre cursado por un alumno se definen w_{min} y w_{max} como la primera y última semana en la que el alumno registra actividad online⁸. Luego se define el intervalo de actividad $I_a = [w_{min}, w_{max}]$ que corresponde al rango de semanas en las que el alumno registra alguna actividad. Para cada semana de I_a se cuenta la cantidad de actividades registradas para cada una de las 6 categorías (ACA, ADM, RC, WC, TEST y Offtask) y de forma agregada a nivel de plataforma (PLAT). Luego de esto, se calcula (X representa cada una de las 7 opciones anteriores):

- X_{sum} : La suma total de cantidad de actividades registradas en I_a .
- X_{mean} : El promedio (mean) semanal de cantidad de actividades registradas en I_a .
- X_{std} : La desviación estándar (std) de la cantidad de actividades registradas en I_a , medidas semanalmente.
- X_{skw} : La asimetría (skewness, skw) como el tercer momento estándar de la distribución de las actividades registradas en I_a semanalmente.
- X_{loc} : La variable ‘localización’ (loc) en $[-1, 1]$ que indica donde se encuentra la mediana de las actividades, medida semanalmente con respecto a I_a . Donde $loc = -1$ indica que la mediana se encuentra en la semana w_{min} , y $loc = 1$ en w_{max} .

A modo de ejemplo, en la figura 4.1 se presenta un histograma de los eventos agrupados

⁷Debido a que en general para cada alumno se tienen observaciones en dos semestres, la base de datos no esta compuesta de observaciones independientes. Si se trabaja cada semestre por separado se asume que las observaciones son independientes, es decir, que los datos sociodemográficos, de desempeño histórico y de la actividad de un alumno no dependen de otros

⁸El trabajar con semanas se justifica en el hecho que la carga académica teórica de los cursos se expresa en horas semanales, las cuales pueden ser distribuidas por los estudiantes en este intervalo de tiempo.

de forma semanal que pertenecen a la categoría ‘Read Comment’ realizados por el estudiante con $id = 8$ en el primer semestre académico del año 2013. Se observa alta variabilidad en el número de conexiones por semana, con diversos altos y bajos a lo largo del semestre, reflejando semanas donde el estudiante tuvo mayor y menor interacción con la plataforma. La posición del centro de masa en el histograma (marcado con una cruz) y un valor de $loc = -0,18$ muestra como el estudiante realiza más actividades del tipo RC hacia la primera mitad del semestre que hacia la segunda.

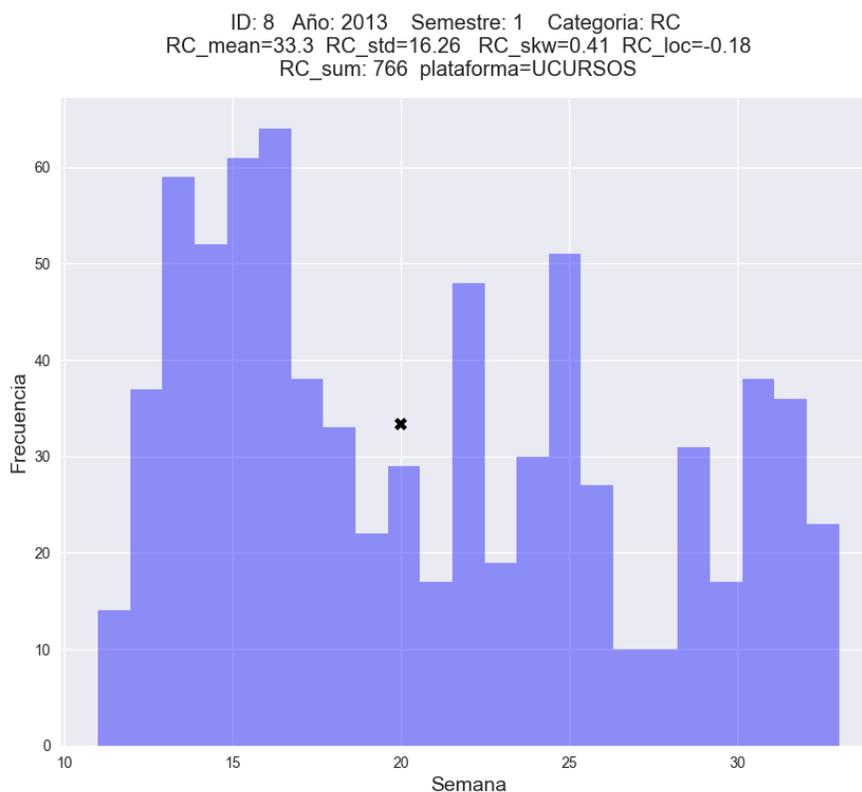


Fig. 4.1: Histograma de los eventos ‘Read Comment’ efectuados por un estudiante durante el primer semestre del año 2013. Se muestra el centro de masa de la distribución con una cruz, donde la abscisa corresponde a la semana que divide la cantidad de actividades en dos, y la ordenada al promedio semanal de actividades.

4.3.4. Horarios de conexión

Para facilitar el analizar la hora de conexión y de actividad de los estudiantes en las plataformas, de la misma forma que en Balmaceda Hidalgo (2012) se definen los horarios de conexión durante el día: ‘Temprano’ (5:00 a 7:59), ‘Mañana’ (8:00 a 10:59), ‘Mediodía’ (11:00 a 14:59), ‘Tarde’ (15:00 a 19:59), ‘Noche’ (20:00 a 23:59) y ‘Madrugada’ (00:00 a 4:59).

4.4. Estructura del análisis

Considerando la gran cantidad de datos de los que se dispone, se plantea realizar un análisis del comportamiento online de los estudiantes utilizando Python 3.6 a la ayuda de las librerías `numpy`, `pandas`, `scikit-learn`, `networkx`, `matplotlib` y `seaborn`, así como MS Excel, Cytoscape y Stata 13. El trabajo a realizar se estructura de la siguiente manera:

- Sección 5.1. Detallar las principales estadísticas descriptivas de los datos sociodemográficos.
- Sección 5.2. Visualizar cuándo los estudiantes interactúan con los LMS: observar los diferentes horarios de la jornada, en distintos días de la semana, y a lo largo del año.
- Sección 5.3. Evaluar la frecuencia con que estudiantes de distinto desempeño académico realizan actividades online. Estudiar la forma y la distribución temporal de estas actividades a lo largo del año, según las categorías señaladas anteriormente.
- Sección 5.4. Ejemplificar, a través del desarrollo de un sociograma inspirado en el proyecto SNAPP, la utilidad de visualizar a través de grafos la actividad y notas de los alumnos en los LMS para realizar inferencias, en miras a una posible implementación en el prototipo del proyecto FONDECYT.
- Sección 5.5. Realizar un modelo explicativo, lo cual se detalla a continuación.

4.5. Modelo Explicativo

Parte del proyecto FONDECYT corresponde a realizar predicciones sobre rendimiento académico incorporando tanto la información con la que se trabaja en este tema de memoria, como datos que se recolectarán a través de la aplicación de cuestionarios a los alumnos. En este contexto, se plantea aquí realizar un primer modelo explicativo —una regresión lineal múltiple— con el fin de evaluar la pertinencia de incluir las variables relativas a la actividad online generadas (de comportamiento semestral) en la predicción y explicación del desempeño académico (variable dependiente) y, de esta forma, evaluar la calidad de estos regresores para implementar un futuro modelo predictivo.

A continuación se presentan algunos trabajos revisados que buscan predecir rendimiento, incluyendo en su metodología diversas variables explicativas. Estos han servido como base para determinar cómo y cuáles variables incluir en los modelos aquí realizados. Luego se plantea la formulación y las limitaciones del modelo aquí desarrollado.

4.5.1. Estudios anteriores

Diversos autores han analizado el rendimiento académico de estudiantes universitarios utilizando diversas metodologías. Ibarra y Michalus (2010) estudia a 589 estudiantes cohortes

1999 a 2003 de la Facultad de Ingeniería de la Universidad Nacional de Misiones. Definiendo el rendimiento académico como el promedio de materias aprobadas anualmente y mediante una regresión logística, logra determinar la incidencia que sobre él tienen diferentes factores de índole personal, socioeconómica, y académica. Los resultados obtenidos permiten concluir que las variables significativas del rendimiento académico son: el promedio de calificaciones anteriores, el tipo de institución donde se cursaron estudios precedentes y el número de asignaturas aprobadas en el primer año de carrera, siendo este último factor el más relevante, destacando la importancia de esta primera etapa de la carrera en los posteriores resultados académicos del estudiante.

Scott et al. (2004) en tanto, examina cómo factores psicosociales (fumar, calidad de vida, apoyo social, etc.) explican hasta el 56 % de la varianza en una regresión lineal múltiple con variable dependiente el CPGA⁹, donde el desempeño anterior (notas de High School y SAT) sólo explica un 25 % de la varianza (Wolfe y Johnson, 1995). Estos factores pueden ser entonces utilizados para identificar alumnos en riesgo de tener un mal desempeño académico en primer año, y para fomentar intervenciones que apunten a corregir estos comportamientos.

Macfadyen y Dawson (2010) por su parte, implementan una regresión lineal múltiple para predecir la nota final de 118 estudiantes en un curso, utilizando como variables independientes el número de mensajes escritos en el foro, emails enviados y tareas completadas en un LMS. Todas ellas resultan significativas (p -valor < 0.05) y la regresión obtiene un $R^2 = 0,33$. Los autores señalan la importancia de verificar la colinealidad y variabilidad de los regresores a utilizar debido a las características de la información con la que se trabaja. Por ejemplo, la cantidad de evaluaciones a realizar puede ser la misma para todos los estudiantes; la cantidad de mensajes escritos en un foro puede estar correlacionada con la cantidad de mensajes leídos; la cantidad de conexiones puede estarlo con la cantidad de actividades en la plataforma, etc.

4.5.2. Formulación

Para la elaboración del modelo explicativo, se recogen diversos factores de la literatura revisada: la importancia del rendimiento académico histórico en el rendimiento futuro, la existencia de factores no intrínsecamente académicos que se relacionan con el desempeño de los estudiantes, y la posibilidad de explicar parte del rendimiento académico por variables originadas en los LMS. Estos factores son indicadores de la posible factibilidad y alcances del trabajo a realizar.

En función de la información disponible, de los resultados de estudios anteriores respecto a la predicción del desempeño académico de estudiantes, y de las diferencias existentes entre ambos LMS, se plantea realizar regresiones lineales diferenciadas por plataforma y comparar la relevancia de las variables explicativas en cada una de las universidades. Para cada plataforma se realizará en un comienzo una regresión sólo con las variables de control (socio-demográficas) y luego las variables de actividad online que se busca estudiar. Así, se analizará el aporte de incluir estos nuevos regresores en la explicación de la variable dependiente. En segunda instancia se repetirá el ejercicio de la regresión, utilizando la submuestra correspon-

⁹Cumulative Grade Point Average, un promedio de calificaciones utilizado en EE.UU.

diente a observaciones de un segundo semestre, e incorporando como regresor el desempeño en un primer semestre. De esta forma, se espera por un lado aumentar la varianza explicada, y por el otro contrastar el comportamiento de los coeficientes de las variables independientes cuando los estudiantes se encuentran ingresando a la universidad y un semestre después. Los resultados se presentan en la sección 5.5 Regresión Lineal Múltiple.

Capítulo 5

Resultados

En este capítulo se articula el trabajo realizado con el fin de dar respuesta a los objetivos propuestos. En primer lugar se entregan las principales estadísticas descriptivas de los datos analizados; luego se identifican comportamientos característicos de distintos grupos de estudiantes en términos de frecuencia y distribución de los LMS donde se espera, a través de la visualización de los datos, dar luces sobre los procesos de aprendizaje en las poblaciones estudiadas en relación a su actividad online. Además se presenta una implementación del proyecto SNAPP mencionado anteriormente y, por último, se detallan los resultados obtenidos de la regresión realizada.

5.1. Población muestral

Como fue mencionado anteriormente, la población muestral corresponde a 1.940 estudiantes de primer año cohortes 2013-2014: 1.465 de ellos de la UChile y 475 de la UC. Es posible evidenciar diferencias entre estudiantes de ambas universidades desde el momento de su ingreso: un 89,2% de los estudiantes de la UChile entran a la universidad inmediatamente tras haber egresado del colegio, un 7,9% espera 1 año para ingresar, y un 2,9% 2 años o más. En la UC en cambio, existe una población más adulta: sólo un 58,1% ingresa inmediatamente después de egresar del colegio, un 18,7% espera 1 año, y un 23,2% 2 años o más.

Además, se destaca la fuerte heterogeneidad en el sexo de los estudiantes, en la UChile los hombres son el grupo mayoritario y en la UC las mujeres, como se logra apreciar en la figura 5.1. A su vez, en el detalle de la distribución de alumnos según tipo de colegio de procedencia presentado en la figura 5.2 se observa que en la UC existe un mayor porcentaje de estudiantes de colegios privados que en la UChile, donde el tipo de colegio de procedencia tiene una distribución más homogénea. En la figura 5.3 se presenta la composición de la población según categoría de ingreso a la universidad, la cual resulta ser similar en ambas universidades. Por último, vale mencionar las tasas de reprobación de la población estudiada en ambas universidades, lo cual se presenta en la tabla 5.1. Las tasas de reprobación de 1 y de 2 o más asignaturas es más alta en la UChile.

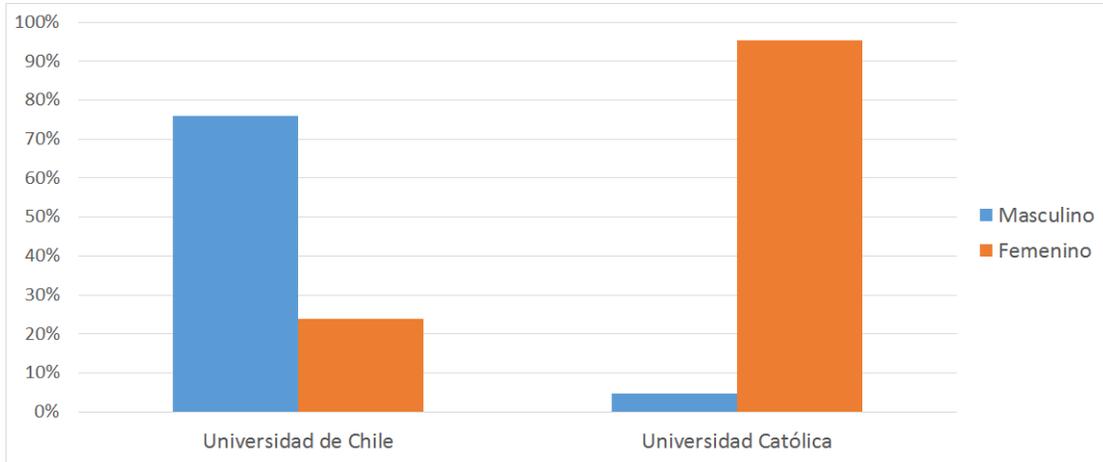


Fig. 5.1: Composición de la muestra según sexo.

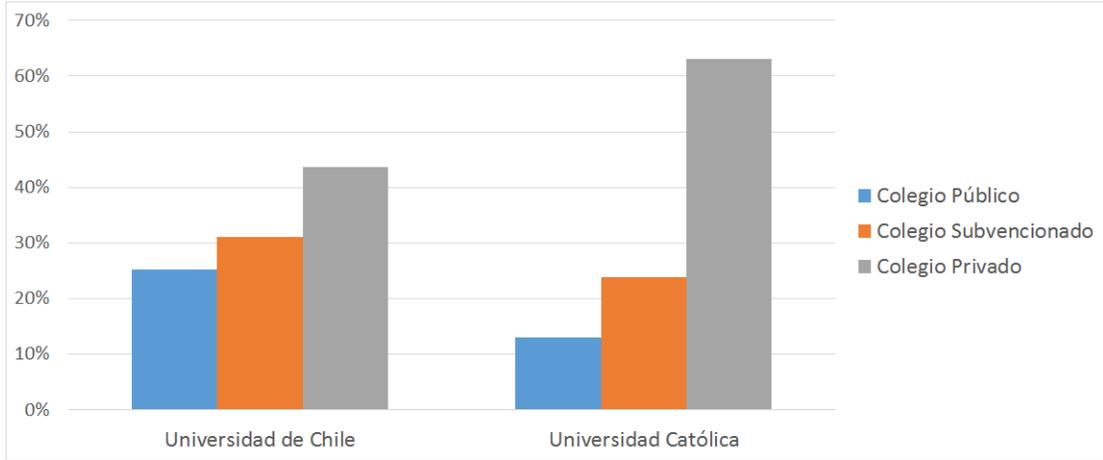


Fig. 5.2: Composición de la muestra según colegio de procedencia.

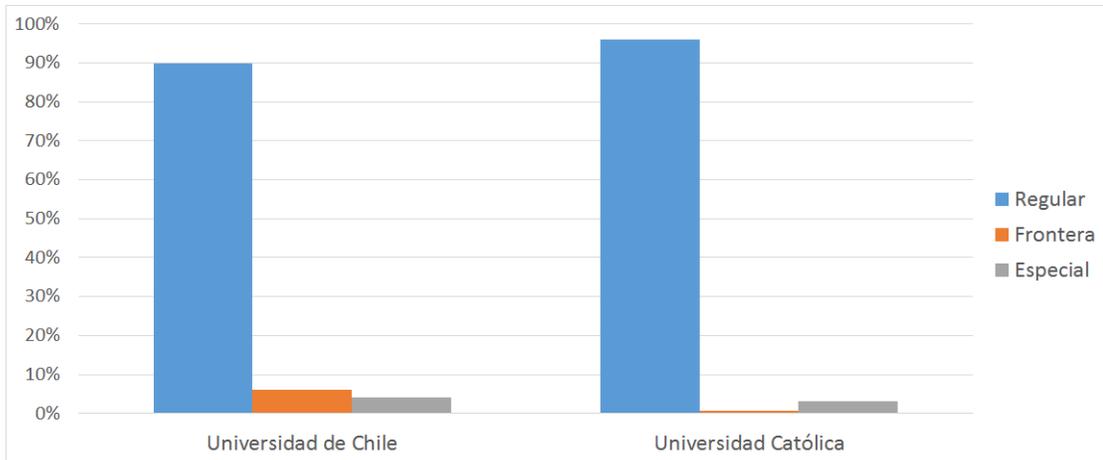


Fig. 5.3: Composición de la muestra según tipo de ingreso a la universidad.

Universidad de Chile					
Ramos Reprobados	Año 2013		Año 2014		Promedio
	Sem. 1	Sem. 2	Sem. 1	Sem. 2	
0	64 %	67 %	68 %	67 %	66 %
1	13 %	21 %	12 %	22 %	17 %
2+	23 %	12 %	20 %	11 %	17 %

Universidad Católica					
Ramos Reprobados	Año 2013		Año 2014		Promedio
	Sem. 1	Sem. 2	Sem. 1	Sem. 2	
0	82 %	78 %	76 %	83 %	80 %
1	13 %	16 %	19 %	10 %	14 %
2+	5 %	5 %	5 %	7 %	6 %

Tabla 5.1: Tasas de reprobación por año y semestre de 0, 1 y 2 o más ramos.

5.2. Actividad en los LMS

5.2.1. Actividad diaria y semanal

Con el fin de analizar hábitos de estudio, es importante estudiar el día y la hora en la cual los alumnos se conectan a las plataformas. Esto fue realizado con los datos disponibles y se muestra en la figura 5.4, donde se presenta el promedio de las distribuciones semanales para estudiantes con desempeño alto, promedio y bajo¹, contando las conexiones en intervalos de una hora. Al respecto, existen bastantes elementos a destacar:

- En Sakai, la distribución se observa menos suave que en U-Cursos. Esto puede deberse a que existen menos observaciones registradas en la UC que en la UChile.
- En ambas universidades se aprecia un uso bastante similar de las plataformas, con una reducción en el número de eventos registrados hacia los fines de semana. Resulta importante destacar que esta disminución de uso comienza los viernes en la UC (flecha 1) y los sábados en la UChile, probablemente debido a que en ingeniería los controles tienen lugar los sábados en la mañana (marcado con una línea punteada roja).
- Durante la tarde y noche de los días de semana en ambas plataformas, pero principalmente en U-Cursos, se observa un aumento de conexiones.
- Durante la madrugada del día sábado en U-Cursos (flecha 2) se registran actividades de estudiantes que preparan el control que rinden a las 9 de la mañana. Estas actividades son realizadas por estudiantes que al final del semestre obtienen un desempeño bajo.
- Los días sábados en ambas plataformas (flechas 3 y 4), un grupo de estudiantes realiza menos conexiones que el resto. Este principalmente corresponde al grupo que, finali-

¹Esto es realizado de forma tal que el aporte de cada uno de los estudiantes es el mismo, y quienes se conectan más no tienen mayor peso en la forma de la distribución.

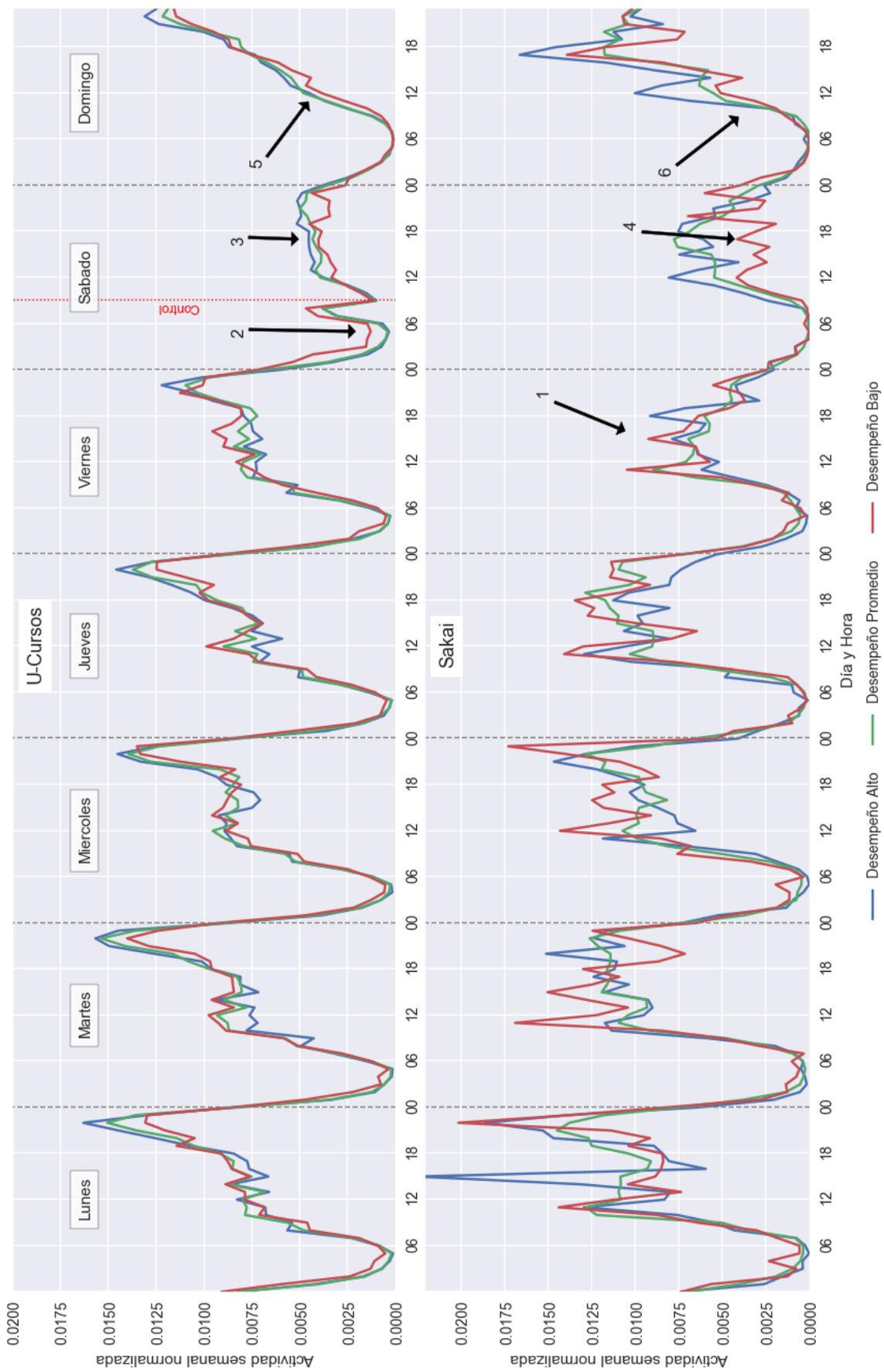


Fig. 5.4: Distribución del uso semanal de los LMS

zando el semestre, obtiene un desempeño bajo.

- Aun siendo un fin de semana, durante la tarde y noche de los días domingos (flechas 5 y 6) los estudiantes registran gran cantidad de conexiones, probablemente para preparar las actividades de la semana.

En la misma línea, se caracteriza a los estudiantes según el momento del día en el que utilizan los LMS. En la figura 5.5 se observa que en ambas plataformas, la mayoría de las conexiones ocurren durante la tarde y noche, y entre un 8 y un 10 % durante la madrugada. Esto es coincidente con los resultados de la encuesta realizada en Balmaceda Hidalgo (2012) a 600 alumnos de primer año de la FCFM donde un 11 % reporta trabajar durante la madrugada.

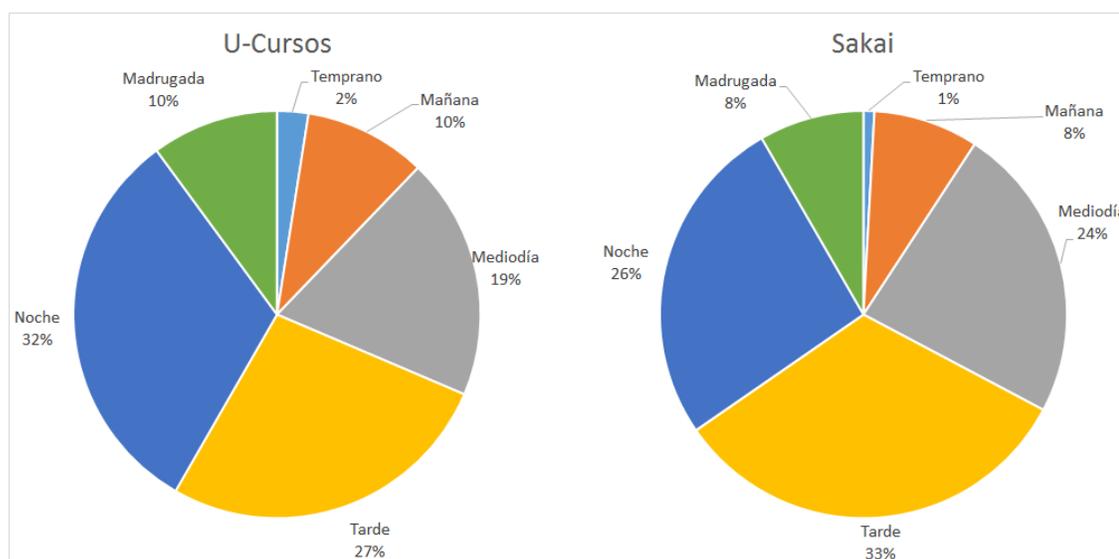


Fig. 5.5: Horario de las actividades en los LMS.

En resumen, los puntos anteriores indican que la organización de la carga académica universitaria por parte de los alumnos los lleva a trabajar durante la noche y los fines de semana, y pone en duda que exista una desconexión de la vida académica en instancias en que los alumnos no se encuentran en la universidad. Este tipo de información podría ser de interés tanto para la toma de decisiones relativa a la carga y planificación académica, como para la realización de talleres de hábitos de estudios.

La importancia de estas cifras y figuras radica en la posibilidad de identificar patrones de estudio en base a registros en los LMS. Se sugiere al equipo, en la etapa de diseño de prototipo, analizar la evolución en tiempo real de estas medidas para monitorear la calidad de vida de los estudiantes. También se recomienda, en la medida de lo posible, conseguir acceso a más datos para poder analizar la evolución de estos hábitos de estudio en años posteriores de la carrera, tras un tiempo de adaptación al ritmo de vida universitario.

5.2.2. Actividad anual

En la figura 5.6 se presenta la actividad online en las plataformas de LMS medida como la cantidad de clics por alumno cada semana del año². Se presentan 2 histogramas (uno para cada año) donde se grafica cada LMS de un color distinto. Se observa una alta variabilidad en el número de registros por semana en ambas plataformas, donde la actividad aumenta hacia el final de cada semestre. Se detecta además como U-Cursos es consistentemente más usado que Sakai. Si bien es difícil determinar las causas de tales patrones, estos podrían estar relacionados, entre otros, a diferencias entre las poblaciones de ambas universidades, a la interfaz de cada LMS, al fomento del uso de la plataforma por parte de la institución, o a la fecha de los períodos evaluativos de cada universidad³.

²Se normaliza la cantidad de clics por número de alumnos para controlar por la diferencia de estudiantes de cada universidad.

³El año 2013, la Universidad de Chile tuvo un paro estudiantil de aproximadamente un mes, causando un desfase en sus actividades que se ve reflejado al comparar las actividades de U-Cursos ambos años.

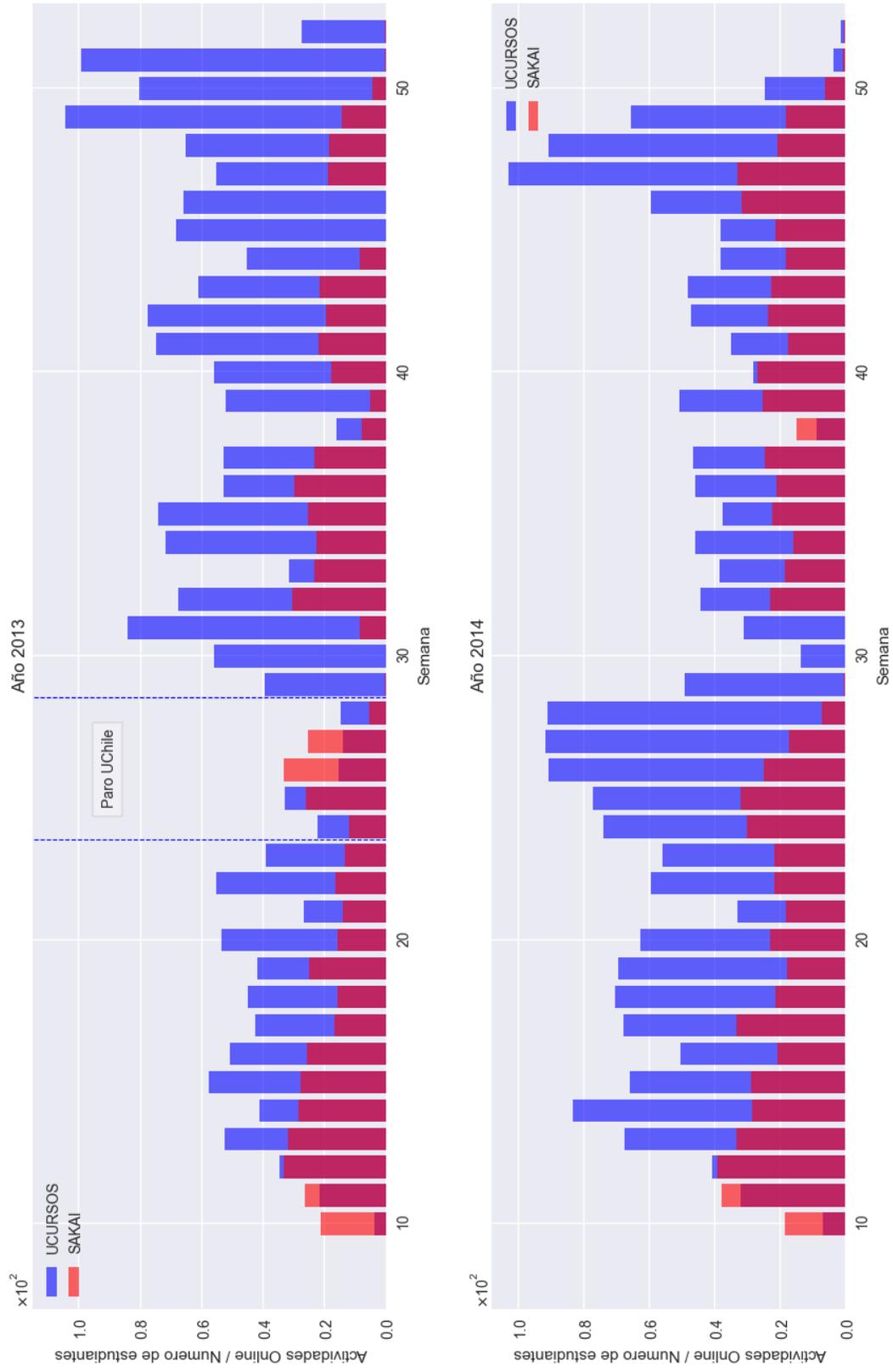


Fig. 5.6: Histograma de la actividad online por año, normalizada por número de estudiantes, según plataforma.

5.3. Frecuencia y Distribución de uso de los LMS

5.3.1. Frecuencia de uso

Para analizar la frecuencia con que los estudiantes utilizan las plataformas se utilizan diagramas de cajas, que permiten visualizar la distribución de los datos⁴. Para cada categoría se realiza un gráfico distinto, donde se divide a los estudiantes según los resultados semestrales que obtienen gracias a la variable ‘Desempeño’. Esto se realiza aquí con datos combinados de ambos LMS, y en la sección siguiente de forma separada.

En la figura 5.7 se observa que, en cada categoría, los estudiantes con mejor desempeño registran más actividades en los LMS que los de desempeño promedio y bajo, incluso en actividades de tipo off-task. Si bien no se puede hablar de causalidad en esta tendencia, si corresponde a una correlación importante. Se destaca que la cantidad de eventos registrados difiere en cada categoría, lo cual es producto de actividades generadas en módulos que cumplen distintas funciones⁵.

Si bien existen diferencias visuales en los diagramas de cajas antes señalados, se desea determinar si las diferencias entre las medias de estas variables son estadísticamente significativas entre estudiantes con distinto desempeño⁶. Sin embargo, para realizar lo anterior no es posible utilizar ANOVA, dado que no se cumple la hipótesis de independencia (como fue explicado anteriormente) y las variables no pasan el test de Levene de homocedasticidad de las varianzas (ver tabla 5.2).

Para solucionar este problema y poder controlar por heterocedasticidad e independizar las muestras, se decide usar dos veces Welch-ANOVA, una vez para estudiantes en primer semestre, y otra para estudiantes en segundo semestre. En caso de rechazar la hipótesis nula de homogeneidad de las medias de los distintos grupos, se realiza el test de Games-Howell para determinar si las diferencias de las medias entre pares de grupos son estadísticamente significativas. Los resultados se muestran en la tabla 5.3. Se observa que efectivamente, se rechaza la hipótesis de homogeneidad de las medias para todas las variables estudiadas (p -valor $< 0,05$) y que para la mayoría de las combinaciones variable-grupo existe una diferencia estadísticamente significativa entre las medias semanales comparadas.

⁴En un diagrama de caja, un punto grande representa la media aritmética de la muestra, la caja a las observaciones del segundo y tercer cuartil, la línea horizontal dentro de la caja la mediana, y las colas 1.5 veces el rango intercuartílico. Los puntos pequeños corresponden a outliers.

⁵Por ejemplo, los test se utilizan en menor medida que ver comentarios.

⁶Los tests realizados en esta sección se consideran significativos con $p < 0,05^*$, $p < 0,01^{**}$ y $p < 0,001^{***}$.

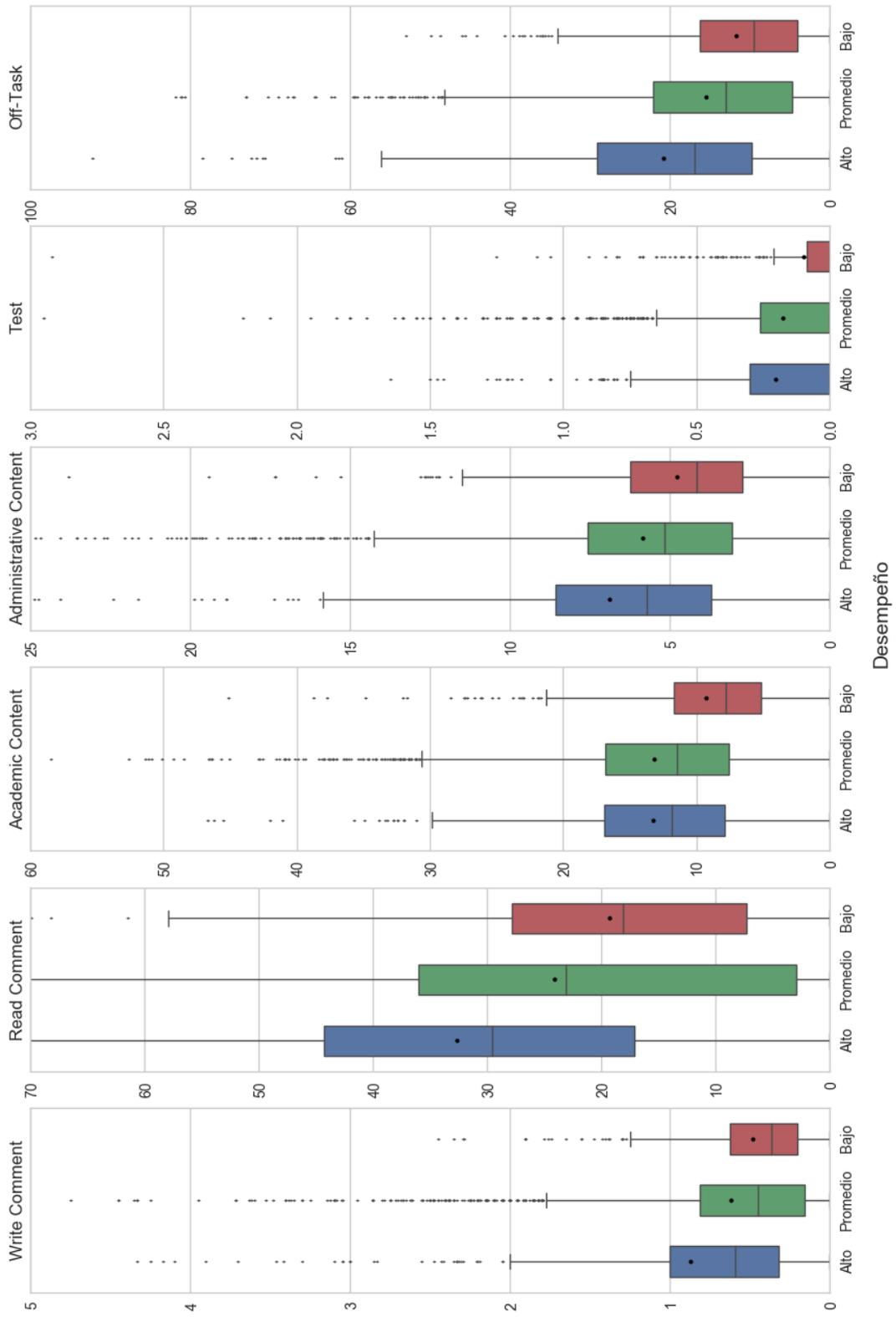


Fig. 5.7: Diagramas de caja de la media del uso de LMS, según desempeño, para las 5 categorías de eventos y actividades off-task (Promedio de las variables X_{mean} , como fue explicado en el capítulo 4 Metodología).

Variable	Semestre 1		Semestre 2	
	Estadístico F	p-valor	Estadístico F	p-valor
WC _{mean}	21,24	7,50E-10***	11,51	1,08E-05***
RC _{mean}	21,40	6,43E-10***	11,08	1,64E-05***
ACA _{mean}	14,00	9,21E-07***	5,72	3,34E-03**
ADM _{mean}	8,20	2,83E-04***	6,65	1,33E-03**
TEST _{mean}	21,78	4,43E-10***	9,55	7,51E-05***
Offtask _{mean}	6,02	2,46E-03**	12,85	2,87E-06***

Tabla 5.2: Test de Levene de estudiantes con desempeño alto, promedio y bajo en semestres 1 y 2, para distintas variables.

Variable	Semestre	Welch-ANOVA		Games Howell		
		Estadístico F	p-valor	Bajo-Promedio	Bajo-Alto	Promedio-Alto
ACA _{mean}	1	31,51	1,37E-13***	sig	sig	sig
ADM _{mean}	1	22,03	7,30E-10***	sig	sig	sig
TEST _{mean}	1	15,91	2,02E-07***	sig	sig	not sig
RC _{mean}	1	34,90	7,72E-15***	sig	sig	sig
WC _{mean}	1	21,87	8,27E-10***	sig	sig	sig
Offtask _{mean}	1	36,49	1,29E-15***	sig	sig	sig
ACA _{mean}	2	50,48	3,54E-20***	sig	sig	not sig
ADM _{mean}	2	23,70	1,90E-10***	sig	sig	not sig
TEST _{mean}	2	10,35	4,08E-05***	sig	sig	not sig
RC _{mean}	2	21,16	2,01E-09***	sig	sig	sig
WC _{mean}	2	8,08	3,74E-04***	sig	sig	sig
Offtask _{mean}	2	17,55	5,40E-08***	sig	sig	sig

Tabla 5.3: Resultados de Welch-ANOVA y test post-hoc de Games-Howell. ‘sig’ representa una diferencia estadísticamente significativa para la variable entre los grupos señalados, al contrario de ‘not sig’.

5.3.2. Frecuencia de uso diferenciada por plataforma

Con el fin de determinar si los hallazgos encontrados son independientes de la plataforma, se realiza el mismo análisis anterior, esta vez haciendo la distinción según LMS. En la figura 5.8 se observa la misma tendencia donde estudiantes con mejores niveles de desempeño realizan más actividades en los LMS. En esta figura se pueden visualizar además las diferencias existentes en la frecuencia de uso de cada categoría por plataforma: en U-Cursos las actividades del tipo ‘Read Comment’ o ‘Write Comment’ se utilizan más que en Sakai, donde se utiliza con mayor frecuencia las actividades del tipo ‘Academic Content’. Esto es analizado en detalle en las secciones a continuación.

Por las mismas razones y de la misma forma que en la sección anterior, se utiliza en primera instancia el test de Levene, y luego Welch-ANOVA y Games-Howell diferenciado por semestre y plataforma para analizar si las medias de las variables son diferentes a una cierta

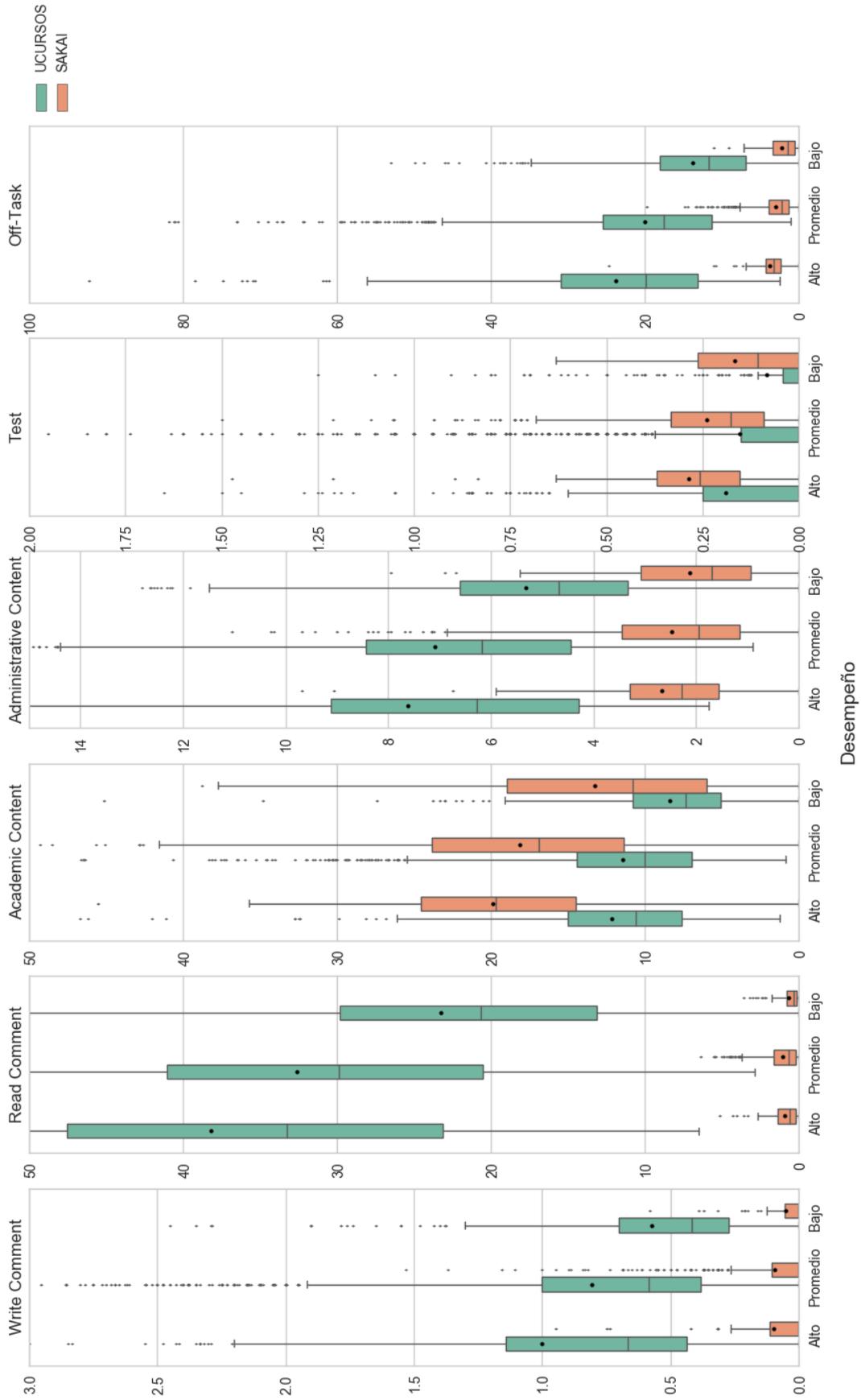


Fig. 5.8: Diagramas de caja de la media del uso de LMS, según desempeño y plataforma, para las 5 categorías de eventos y actividades off-task (Promedio de las variables X_{mean} , como fue explicado en el capítulo 4 Metodología).

significancia estadística (5%). Los resultados del test de Levene se presentan en la tabla 5.4, donde se verifica que en general las variables no cumplen la hipótesis de homocedasticidad (p -valor $< 0,05$) por lo que se utiliza Welch-ANOVA⁷.

Los resultados de Welch-ANOVA y Games-Howell que se presentan en la tabla 5.5 muestran cómo en la UC, el comportamiento de los estudiantes con distinto desempeño es bastante homogéneo, y sólo para algunas variables y algunos grupos es posible determinar diferencias significativas en la frecuencia de uso de la plataforma. En la UChile en cambio, se presentan diferencias entre grupos de todas las variables analizadas, indicando una real frecuencia de uso distinta por parte de estudiantes con distinto desempeño.

Variable	Plataforma	Semestre 1		Semestre 2	
		Estadístico F	p-valor	Estadístico F	p-valor
WC _{mean}	U-Cursos	19,78	3,35E-09***	10,17	4,12E-05***
RC _{mean}	U-Cursos	14,84	4,18E-07***	9,27	1,00E-04***
ACA _{mean}	U-Cursos	1,78	1,69E-01	4,80	8,38E-03**
ADM _{mean}	U-Cursos	6,38	1,74E-03**	6,00	2,54E-03**
TEST _{mean}	U-Cursos	22,62	2,12E-10***	8,82	1,57E-04***
Offtask _{mean}	U-Cursos	3,51	3,03E-02*	11,95	7,13E-06***
WC _{mean}	Sakai	2,59	7,64E-02	3,06	4,79E-02*
RC _{mean}	Sakai	0,89	4,10E-01	4,87	8,04E-03**
ACA _{mean}	Sakai	4,48	1,19E-02*	0,21	8,10E-01
ADM _{mean}	Sakai	2,98	5,16E-02	1,53	2,17E-01
TEST _{mean}	Sakai	0,75	4,71E-01	4,80	8,64E-03**
Offtask _{mean}	Sakai	1,05	3,50E-01	1,17	3,10E-01

Tabla 5.4: Test de Levene para cada plataforma de estudiantes con desempeño alto, promedio y bajo en semestres 1 y 2, para distintas variables.

⁷Algunas variables si pasan el test de Levene, principalmente en Sakai. Sin embargo, se opta por realizar el mismo test de Welch-ANOVA para todas las variables para facilitar la comparación, aun cuando podría tener menos poder estadístico que ANOVA.

Variable	Plataforma	Semestre	Welch-ANOVA		Games Howell		
			F	p-valor	Bajo-Prom.	Bajo-Alto	Prom.-Alto
ACA _{mean}	U-Cursos	1	13,41	2,26E-06***	sig	sig	not sig
ADM _{mean}	U-Cursos	1	34,91	8,77E-15***	sig	sig	not sig
TEST _{mean}	U-Cursos	1	6,77	1,30E-03**	sig	sig	not sig
RC _{mean}	U-Cursos	1	56,94	1,35E-22***	sig	sig	sig
WC _{mean}	U-Cursos	1	28,09	3,53E-12***	sig	sig	not sig
Offtask _{mean}	U-Cursos	1	43,92	2,07E-18***	sig	sig	sig
ACA _{mean}	U-Cursos	2	44,15	1,39E-17***	sig	sig	not sig
ADM _{mean}	U-Cursos	2	27,47	1,03E-11***	sig	sig	not sig
TEST _{mean}	U-Cursos	2	6,64	1,48E-03**	sig	sig	not sig
RC _{mean}	U-Cursos	2	31,64	4,11E-13***	sig	sig	sig
WC _{mean}	U-Cursos	2	8,25	3,33E-04***	sig	sig	sig
Offtask _{mean}	U-Cursos	2	20,65	4,46E-09***	sig	sig	sig
ACA _{mean}	Sakai	1	9,05	5,10E-04***	sig	sig	sig
ADM _{mean}	Sakai	1	0,55	5,81E-01	not sig	not sig	not sig
TEST _{mean}	Sakai	1	2,87	7,03E-02	not sig	not sig	not sig
RC _{mean}	Sakai	1	4,16	2,41E-02*	sig	not sig	not sig
WC _{mean}	Sakai	1	3,45	4,28E-02*	not sig	not sig	not sig
Offtask _{mean}	Sakai	1	0,96	3,91E-01	not sig	not sig	not sig
ACA _{mean}	Sakai	2	11,69	3,72E-05***	sig	sig	not sig
ADM _{mean}	Sakai	2	2,51	8,73E-02	not sig	not sig	not sig
TEST _{mean}	Sakai	2	7,16	1,43E-03**	not sig	sig	sig
RC _{mean}	Sakai	2	6,20	3,03E-03**	sig	not sig	not sig
WC _{mean}	Sakai	2	4,43	1,44E-02*	sig	not sig	not sig
Offtask _{mean}	Sakai	2	9,53	1,97E-04***	sig	sig	not sig

Tabla 5.5: Resultados de Welch-ANOVA y test post-hoc de Games-Howell por plataforma. ‘sig’ representa una diferencia significativa, al contrario de ‘not sig’.

5.3.3. Distribución de uso temporal

Para analizar la forma en que los estudiantes utilizan las plataformas se estudia, para cada LMS en el 2013 y 2014, la evolución anual de la distribución porcentual de los eventos agrupados en las 5 categorías diseñadas. Esto se presenta en la figura 5.9.

Al respecto, se observa que los estudiantes utilizan de distinta manera Sakai y U-Cursos: en el primero se realizan más eventos relacionados con las unidades temáticas de los cursos, mientras que en el segundo se realizan más actividades donde los alumnos escriben y leen mensajes relacionándose unos con otros. Por otro lado, también se advierte que en ambas plataformas, tanto para el 2013 como el 2014, el aporte porcentual de cada categoría al total es relativamente constante a lo largo del año. De la misma forma, la composición porcentual de las categorías en ambas plataformas es similar de un año al otro, aun cuando esto corresponde a poblaciones distintas (cohortes 2013 y 2014).

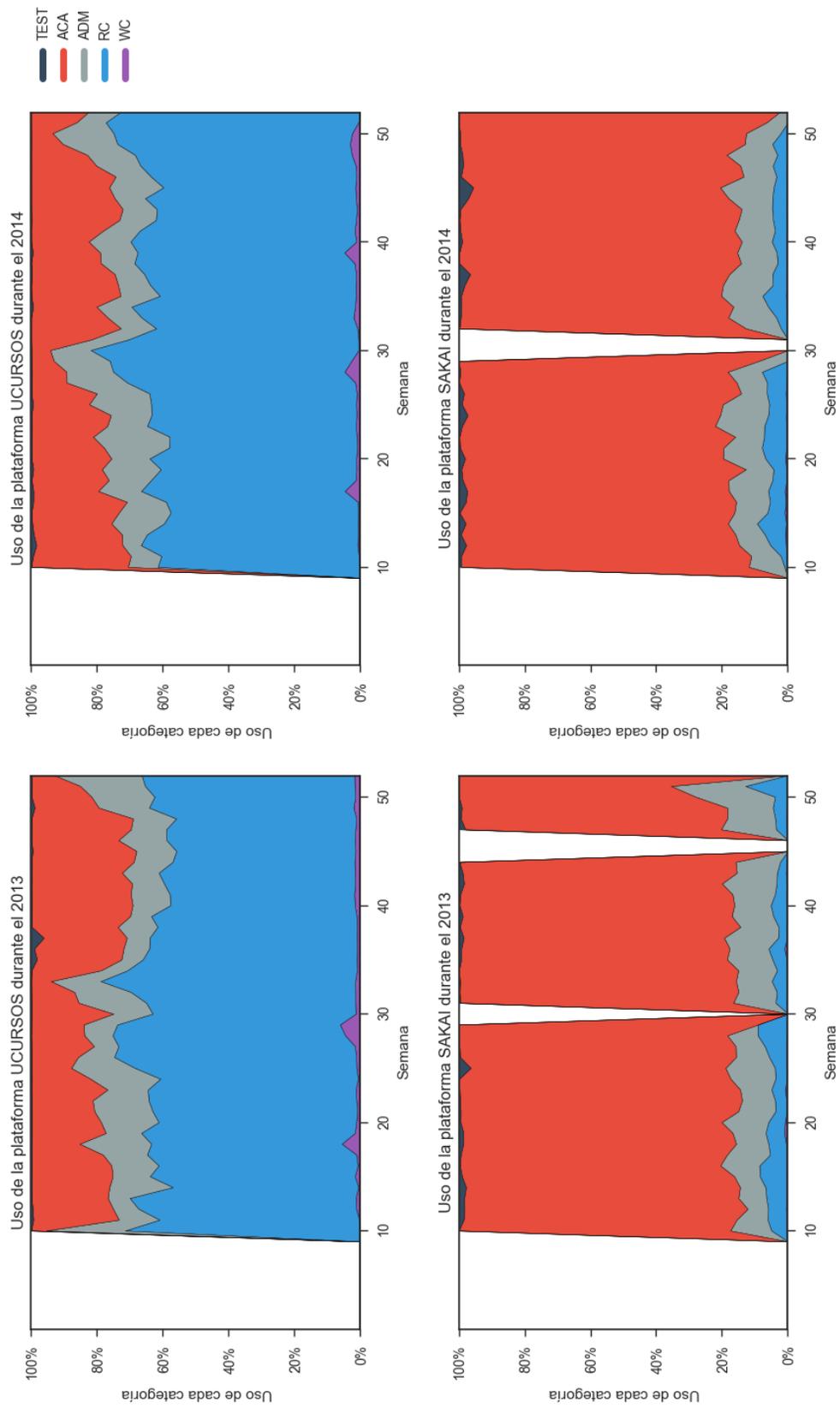


Fig. 5.9: Evolución en el tiempo del uso de cada categoría de los LMS según la cantidad de eventos registrados en un año.

5.3.4. Distribución de uso según desempeño

En la figura 5.10 se presenta la distribución de uso de cada plataforma según el desempeño académico de estudiantes en las distintas categorías de eventos⁸. Se observa que existe una diferencia en el uso que estudiantes de distintas universidades dan al LMS que tienen a disposición, pero no entre los estudiantes que obtienen distintos desempeños. Dicho de otro modo, en cada universidad se le da una forma⁹ de uso distinta a la plataforma que disponen, pero dentro de cada una de las universidades, los estudiantes utilizan de la misma forma su plataforma, independiente de su desempeño al final del semestre.

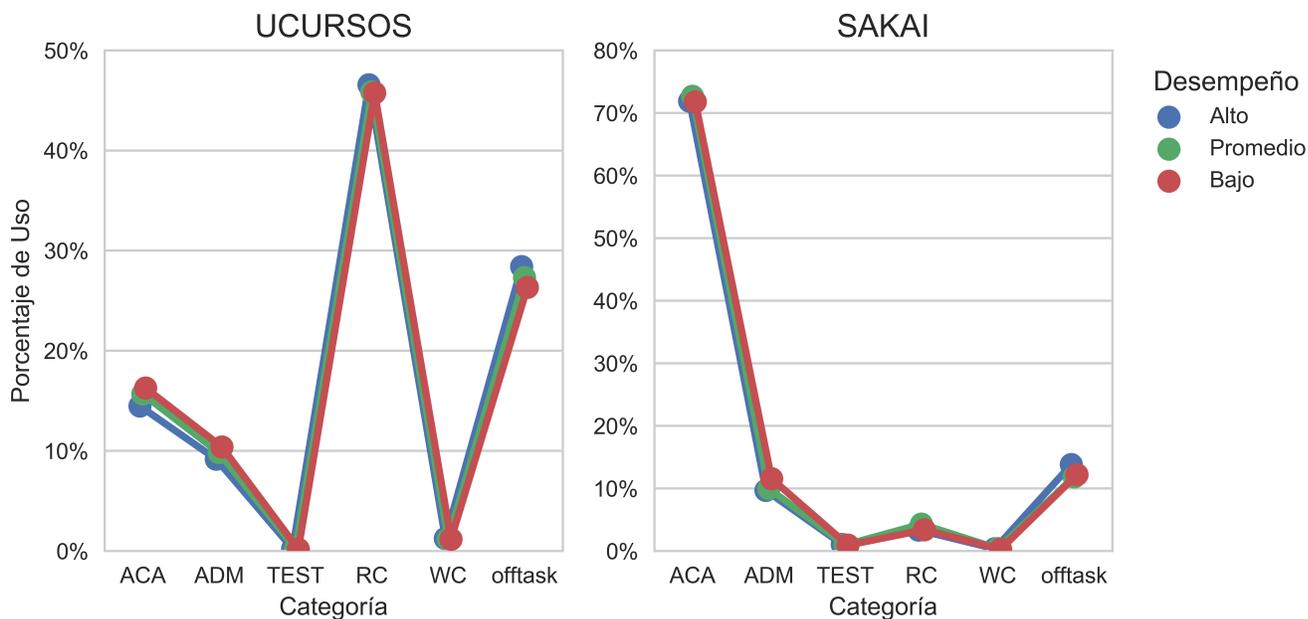


Fig. 5.10: Distribución de uso de la plataforma, según categoría.

Gasevic et al. (2016) plantean que al analizar la actividad en un LMS es necesario considerar las características específicas al contexto educativo donde el aprendizaje tiene lugar. En esta línea, se decide analizar la distribución de uso de las plataformas al interior de cada curso, para lo cual se construyen gráficos de la misma forma que la figura 5.10, sólo con información a nivel de cada asignatura. Estos gráficos se encuentran disponibles en anexos y equivalen a 7 cursos en U-Cursos (Anexo D) y 21 cursos en Sakai (Anexo E). Allí se puede observar cómo se mantiene la diferencia de uso entre U-Cursos y Sakai, y cómo la distribución del uso dentro de la misma plataforma, en términos generales, es similar. Además, dentro de cada curso no existen mayores diferencias entre estudiantes independiente de su desempeño académico al final del semestre.

⁸Esto corresponde a la distribución agregando todas las semanas de los dos años, e incluyendo esta vez en el análisis a las actividades off-task.

⁹Se hace la distinción entre forma (distribución porcentual de las actividades en cada categoría) y frecuencia (cantidad de clics en cada categoría).

5.3.5. Uso por sección

Es factible analizar lo anterior de manera más detallada en la UChile. En este caso particular, los aproximadamente 700 estudiantes que ingresan anualmente son distribuidos en 8 secciones de cerca de 90 estudiantes cada una, de forma tal que los estudiantes de cada sección sean heterogéneos en términos del desempeño académico obtenido para entrar a la carrera de acuerdo a su posición de ingreso, según la fórmula¹⁰

$$Seccion = \begin{cases} posicion_{ingreso} \bmod 8 & , \text{ si } posicion_{ingreso} \bmod 8 \neq 0 \\ 8 & , \text{ si } posicion_{ingreso} \bmod 8 = 0 \end{cases}$$

Al observar la frecuencia y distribución de uso de las 3 categorías más usadas (Academic Content, Administrative Content y Read Comment) en la figura 5.11, se observa que el comportamiento de los mismos 90 alumnos (misma sección) puede variar fuertemente en distintos cursos. Por ejemplo, la sección 5 es una de las que registra menos actividades en FI1001, y más actividades en MA1001. Dado que se han mantenido constantes tanto la plataforma como la población estudiada, y los contenidos son los mismos en cada curso (pues las secciones son coordinadas), la única variable que cambia, y que por tanto afecta la distribución de uso, es alguna dimensión del equipo docente¹¹.

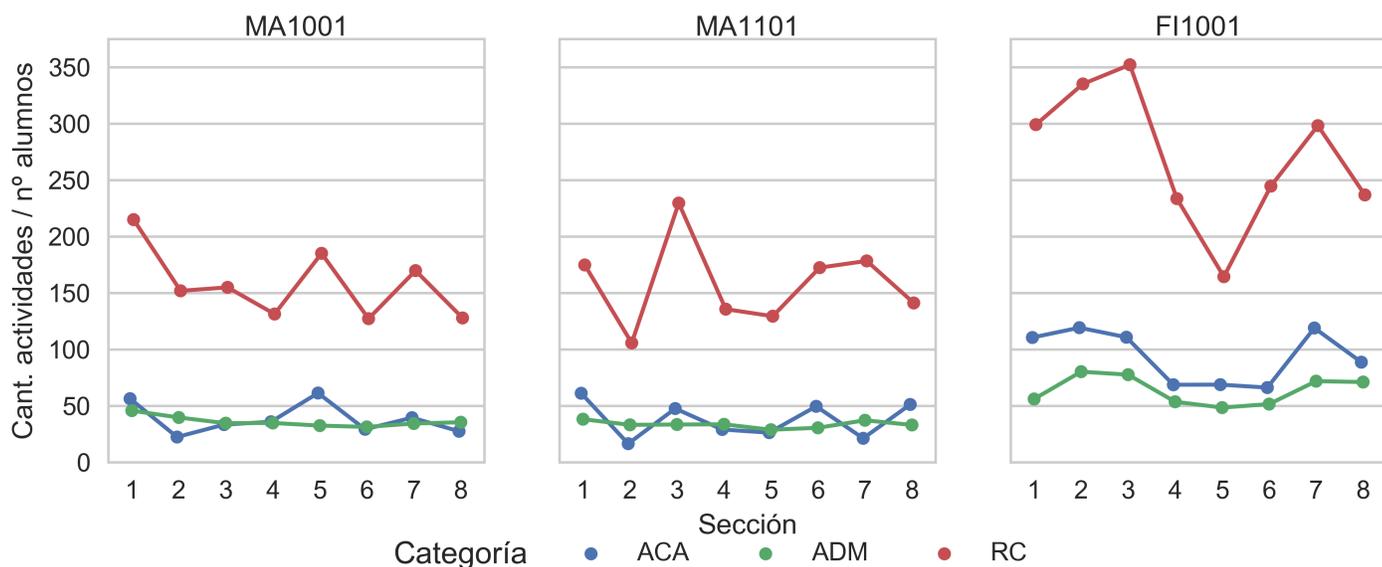


Fig. 5.11: Frecuencia de actividades del primer semestre del 2013 en U-Cursos normalizada por cantidad de alumnos, en las 8 diferentes secciones de 3 cursos de la Universidad de Chile para 3 categorías de actividad. A la izquierda MA1001: Introducción al Cálculo, al centro MA1101: Introducción al Álgebra, y a la derecha FI1001: Introducción a la Física Newtoniana.

¹⁰De esta forma el estudiante que ingresa primero a la universidad es ubicado en la sección 1, el segundo en la sección 2, ..., el octavo en la sección 8, el noveno en la sección 1, el décimo en la sección 2, etc.

¹¹También puede darse el caso que a los estudiantes de la sección 5 les guste la física y no el cálculo. Sin embargo, dado como se conforman las secciones, esto es poco probable.

5.3.6. Conclusiones análisis de frecuencia y distribución de uso

El análisis de la frecuencia y forma de uso realizado sugiere que, en línea con la literatura revisada en la sección 3.3 Los LMS en Educación Superior, existen distintos factores que determinan cómo los estudiantes interactúan con los LMS. Estos pueden ser relativos a los estudiantes y al equipo docente, a la interfaz y funcionalidades del LMS, y al rol y validez que la comunidad educativa le otorga. Es una combinación de estas variables las que motivan la realización de actividades de cierto tipo por sobre otros, moldeando la experiencia de aprendizaje online de los estudiantes.

En ambas universidades se logra identificar que estudiantes que tienen mejores desempeño al final del semestre han utilizado con mayor frecuencia la plataforma. Se observa además que existe una cultura de uso de los LMS compartida dentro de cada institución educativa. Mientras el LMS en la UChile cumple el rol de ser una plataforma que permite la comunicación, en la UC se orienta hacia la gestión del contenido. Por último, se han presentado visualizaciones que permiten identificar hábitos de estudio y distintas situaciones que pueden ocurrir durante el semestre académico. Se espera que este tipo de visualización pueda ser incorporado en la última etapa del proyecto con el fin de la entrega de retroalimentación y recomendaciones a los usuarios.

5.4. Sociograma

Algunos eventos registrados por los LMS corresponden a partes de una estructura que se conforma tras la realización de una secuencia de actividades, y se observa al representar estas actividades como un conjunto. Es el caso, por ejemplo, de las actividades que se registran en los foros de discusión, donde no sólo existen relaciones estudiante-plataforma, sino también estudiante-estudiante o estudiante-profesor. Para entender estas relaciones y sortear las limitantes de observar cada actividad por separado, resulta necesario representarlas de forma tal de poder observar la estructura subyacente a los registros en las plataformas. El proyecto SNAPP permite visualizar, a través de redes, la evolución de las relaciones entre estudiantes y profesores en los distintos foros de discusión de sus cursos. Con ello, a través de herramientas de teoría de grafos es posible identificar, por ejemplo, estudiantes con alto o bajo nivel de participación, la dirección del flujo de la información, o el nivel de interacción entre distintos estudiantes.

Dado lo natural de representar la actividad de los estudiantes en foros como una red (lo que en el caso de relaciones sociales se conoce como sociograma), resulta deseable el poder incorporar en este tipo de visualización la dimensión del desempeño académico de los estudiantes. Para ello en una primera instancia, con base en el trabajo de SNAPP, se diseñó un grafo dirigido (ver ejemplo en la figura 5.12) en el que cada nodo representa un alumno, y una arista de X con dirección a Y indica que el estudiante X escribió un mensaje en un hilo de conversación donde el estudiante Y ya había escrito. Para incorporar el rendimiento, se colorean los nodos según el desempeño (bajo, promedio o alto) que haya obtenido el estudiante

ese semestre¹². Se permite de esta forma observar en un mismo diagrama las dimensiones de actividad online y rendimiento académico, donde las posibilidades de visualización son múltiples. Por ejemplo, puede modificarse el algoritmo (Spring, Circular, Spectral, etc.) que determina la disposición espacial de los nodos para identificar distintas características de la red; visualizar las lecturas de los mensajes en un foro en vez de las respuestas (definiendo la relación de X a Y como “ X leyó un mensaje de Y ”); o bien modificar la forma o tamaño de los nodos para evaluar simultáneamente varias características.

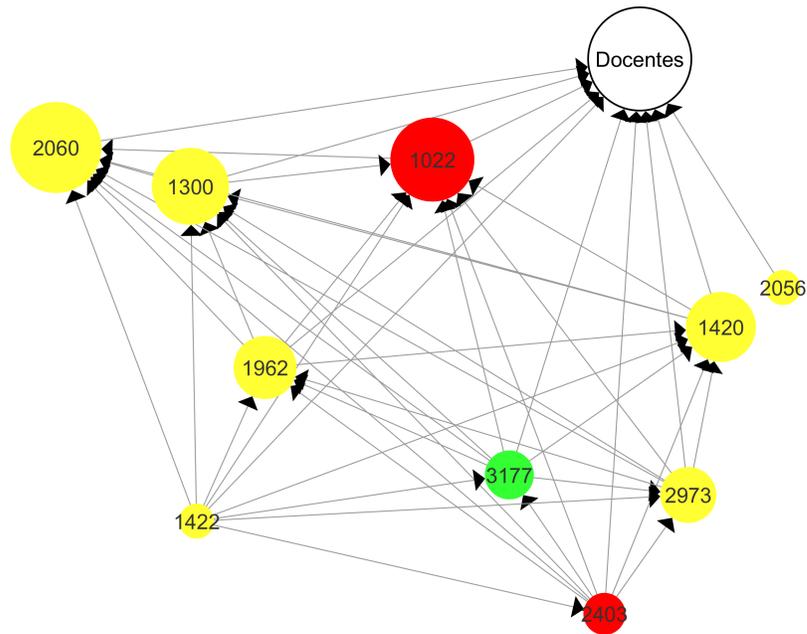


Fig. 5.12: Sociograma (grafo dirigido) de los primeros 15 mensajes del foro de una sección de MA1001 (U-Cursos) año 2014. El nodo blanco corresponde a los integrantes del equipo docente y el tamaño de cada nodo se correlaciona con su grado de entrada (La suma de aristas entrantes). En rojo estudiantes que al final del semestre tienen desempeño bajo, en amarillo un desempeño promedio, y en verde un desempeño alto.

Por temas de alcance no se profundizará en el desarrollo del sociograma que mejor permita representar la información existente en la base de datos de actividad online incorporando desempeño académico. Sin embargo, en línea con determinar variables que predigan desempeño académico y en elaborar un prototipo para ayudar a la toma de decisiones académicas durante el cuarto año del proyecto, se recalca la importancia de incorporar este tipo de técnicas en el análisis. En esta dirección, se plantea en un trabajo futuro el desarrollar distintas representaciones de las aristas del sociograma (como fue explicado anteriormente) e incorporar medidas de centralidad de las redes como predictores del desempeño académico de estudiantes y para identificar estudiantes con baja participación.

¹²Esta información está disponible sólo al final del semestre, por lo que este grafo no puede construirse en tiempo real. Sin embargo, no es difícil concebir otra variable relacionada con el desempeño con la que identificar a cada estudiante, por ejemplo, las notas parciales, o bien la probabilidad de aprobación de un curso como calculada por Celis et al. (2015).

5.5. Regresión Lineal Múltiple

Como fue explicado en la sección 4.5 Modelo Explicativo, a continuación se da cuenta de la regresión lineal elaborada para evaluar la utilidad de incluir regresores relativos a la actividad online en la explicación de la variable dependiente desempeño académico.

5.5.1. Correlación entre variables

Se determina en primera instancia el coeficiente de correlación de Pearson entre los pares de variables cuantitativas explicativas a utilizar, con el fin de identificar posibles problemas de colinealidad que se podrían encontrar. Esto se presenta en la figura 5.13. Debido a la alta correlación entre algunas variables relacionadas a la actividad online (sector inferior derecho de la matriz) y de la psu y nem, se realizará un análisis de la VIF una vez realizada la regresión para evaluar que variables dejar fuera.

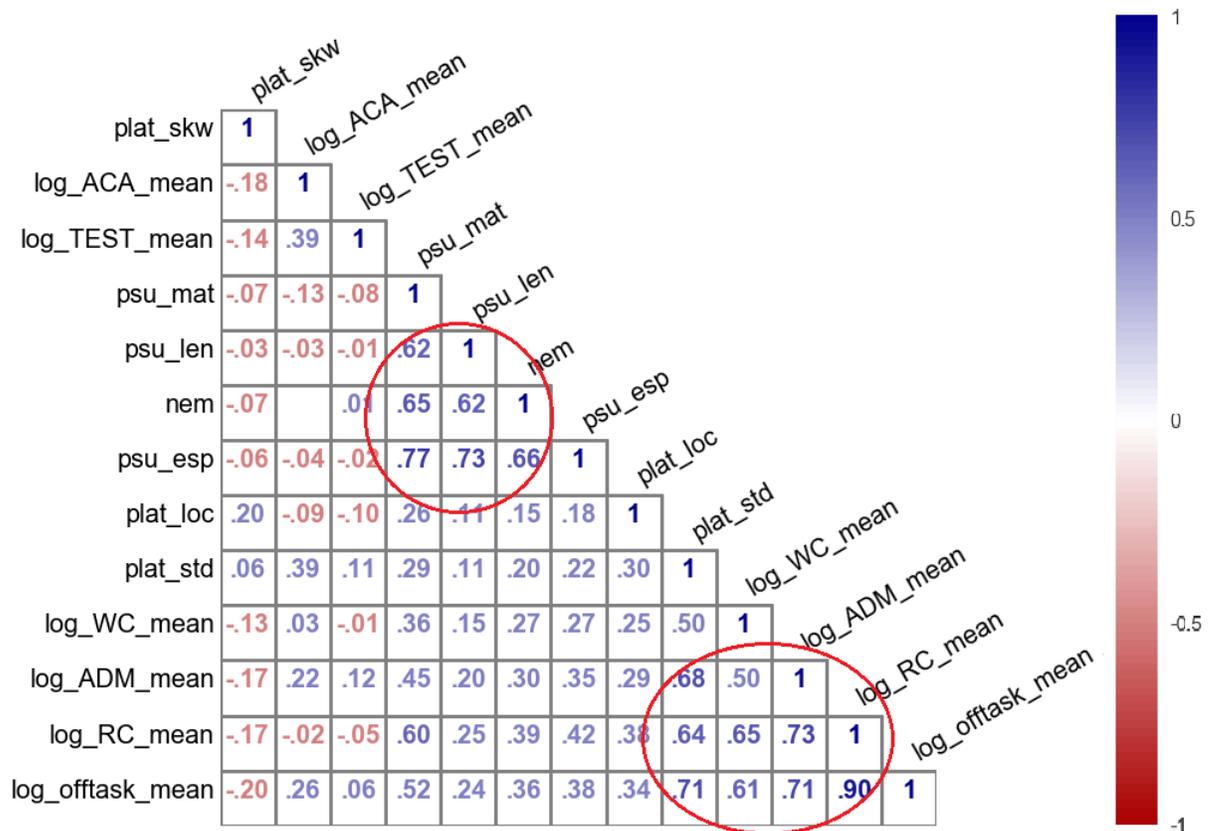


Fig. 5.13: Parte inferior de la matriz de correlación de variables (solo se presenta la mitad debido a la simetría de la matriz).

5.5.2. Forma Estructural

Se realizan cuatro regresiones lineales múltiples, dos para la UChile (1 y 2) y dos para la UC (3 y 4)¹³. En dos de ellas, las regresiones 1 y 3, sólo se incluyen las variables de control, y su especificación queda dada por la ecuación 5.1.

$$\begin{aligned} zscore_{carrera} = & \beta_0 + \beta_1 nem + \beta_2 psu_{mat} + \beta_3 psu_{len} + \beta_4 psu_{esp} + \beta_5 F + \beta_6 ped_{basica} \\ & + \beta_7 privado + \beta_8 subvencionado + \beta_9 ingreso_{especial} + \beta_{10} ingreso_{frontera} + \varepsilon \end{aligned} \quad (5.1)$$

Donde los coeficientes β_i corresponden a los parámetros a estimar. La variable ped_{basica} sólo se utiliza en Sakai y vale 1 si el estudiante es de la carrera pedagogía básica y 0 si no. Las variables $privado$ y $subvencionado$ son variables dummy que indican pertenencia a colegios privados y subvencionados respectivamente, con nivel de referencia a los colegios públicos. Las variables $ingreso_{especial}$ e $ingreso_{frontera}$ indican el tipo de ingreso a la universidad, donde el nivel de referencia es un ingreso de tipo normal.

Por otro lado, en 2 y 4 se incluyen además de las variables de control, aquellas que corresponden a información sobre la actividad online de los estudiantes, como se muestra en la ecuación 5.2. La lógica de analizar las variables de actividad online en regresiones separadas es poder comparar si los resultados de incluir estas variables son estadísticamente mejores que en su ausencia de forma diferenciada por universidad.

$$\begin{aligned} zscore_{carrera} = & \beta_0 + \beta_1 nem + \beta_2 psu_{mat} + \beta_3 psu_{len} + \beta_4 psu_{esp} + \beta_5 F + \beta_6 ped_{basica} \\ & + \beta_7 privado + \beta_8 subvencionado + \beta_9 ingreso_{especial} + \beta_{10} ingreso_{frontera} \\ & + \beta_{11} PLAT_{std} + \beta_{12} PLAT_{skw} + \beta_{13} PLAT_{loc} \\ & + \beta_{14} \ln(ACA_{mean}) + \beta_{15} \ln(ADM_{mean}) + \beta_{16} \ln(WC_{mean}) + \beta_{17} \ln(RC_{mean}) \\ & + \beta_{18} \ln(TEST_{mean}) + \beta_{19} \ln(Offtask_{mean}) + \varepsilon \end{aligned} \quad (5.2)$$

Las variables $PLAT$ asociadas a los coeficientes β_{11} , β_{12} y β_{13} hacen referencia a la distribución del uso que los estudiantes dan a la plataforma, donde $PLAT_{std}$ es una medida de la inconstancia de uso, en tanto $PLAT_{skw}$ y $PLAT_{loc}$ hacen referencia al momento del semestre en que los estudiantes se conectan (de forma temprana o tarde). Por otro lado, el resto de las variables asociadas a los coeficientes β_{14} , β_{15} , β_{16} , β_{17} , β_{18} y β_{19} representan medidas de la frecuencia de uso de las plataformas. La forma funcional $\ln(\cdot)$ tiene por objetivo suavizar las variables entre categorías con frecuencia de uso distintas.

Para el cálculo de la matriz de varianza-covarianza de los errores se utiliza la estimación robusta en clústers de id, debido a que observaciones para el mismo estudiante en distintos semestres no son independientes. Se opta por no incluir efectos fijos por año-semestre en la regresión debido a que la variable objetivo $zscore_{carrera}$ ya corresponde al desempeño de un estudiante en relación a sus compañeros en un contexto académico determinado por el año, semestre y carrera en curso. Por último, se incluyen para las regresiones 2 y 4 un test de Wald

¹³La razón para hacer el estudio diferenciado en la UChile y la UC es poder identificar si existen variables que tengan impacto distinto en el desempeño de estudiantes en cada universidad.

para comprobar que estas variables de actividad online efectivamente aportan información a los modelos explicativos, en comparación a 1 y 3, respectivamente¹⁴.

5.5.3. Resultados

Al realizar una primera regresión se encontraron altos niveles de colinealidad entre algunas variables, medido según la VIF. Estas variables fueron removidas¹⁵, y se reestimaron los coeficientes. Los resultados se muestran en la tabla 5.6, donde a la derecha de cada coeficiente se encuentra entre paréntesis la VIF de cada regresor. Se puede observar como la información de la actividad online ayuda a explicar 8 % más de la varianza en el caso de UCursos, y sólo un 2,5 % más en el caso de Sakai, es decir, un 21,9 % y un 13,8 % del total explicado, respectivamente.

Debido a que $zscore_{carrera}$ es una variable estandarizada y los coeficientes de las distintas variables se encuentran en distintas escalas, no se puede interpretar de manera directa la magnitud del impacto de las variables independientes en esta regresión lineal¹⁶. Sin embargo el signo tiene una interpretación clara: un coeficiente mayor a cero tiene un efecto positivo en la variable dependiente y por tanto en el rendimiento académico, en cambio un coeficiente menor a cero tiene un efecto negativo sobre este. Se observa como:

- Las variables relativas al desempeño histórico tienen efecto positivo en el desempeño al término del semestre (nem en ambas universidades, y psu_{mat} y psu_{esp} en UChile).
- En la UChile, las mujeres obtienen desempeños más bajos que los hombres, *ceteris paribus*.
- De forma contraintuitiva en el contexto chileno, los resultados de alumnos egresados de colegios privados y subvencionados tienen peores desempeños en la UChile.
- Los alumnos que ingresan de forma especial a la UC (cambio de carrera, cambio de universidad) tienen mejores resultados que quienes ingresan por otras vías.
- Con respecto a las variables relacionadas con la distribución de uso, también de forma contraintuitiva los resultados de la variable $PLAT_{std}$ indican que en la UC los estudiantes menos constantes en sus conexiones tienen mejores resultados (al 5 %).
- Por otro lado la variable $PLAT_{loc}$ indica que en la UChile quienes se conectan más hacia la segunda mitad del semestre obtienen mejores resultados. Esto puede deberse al mayor porcentaje de la nota del semestre que se encuentra en juego hacia el final del semestre, particularmente en el período de exámenes.
- Con respecto a la frecuencia de uso, los signos de los coeficientes indican —en concordancia con lo encontrado anteriormente— que en la UChile, los estudiantes que utilizan

¹⁴Debido a que se utiliza la estimación de clústers, no se utiliza el test razón de verosimilitud, como fue explicado en la sección 3.5 Elementos Econométricos.

¹⁵En el caso de U-Cursos corresponde a $\ln(rc_{mean})$ y $\ln(offtask_{mean})$, y en Sakai a psu_{mat} , psu_{len} , psu_{esp} y $\ln(aca_{mean})$.

¹⁶Además, por esta razón, existen variables con 0 en los tres primeros decimales que son significativas al 1 %.

	(1) UChile		(2) UChile ₂		(3) UC		(4) UC ₂	
	zscore _{carrera}		zscore _{carrera}		zscore _{carrera}		zscore _{carrera}	
nem	0,01***	(1,2)	0,01***	(1,2)	0,00***	(3,6)	0,00***	(3,6)
psu _{mat}	0,01***	(1,4)	0,01***	(1,4)				
psu _{len}	-0,00	(1,2)	0,00	(1,2)				
psu _{esp}	0,01***	(1,3)	0,01***	(1,3)				
F	-0,10*	(1,1)	-0,11**	(1,1)	0,11	(1,1)	0,09	(1,1)
ped _{gen_basica}					-0,13	(1,1)	-0,10	(1,4)
privado	-0,22***	(1,6)	-0,18***	(1,6)	0,21	(2,1)	0,23	(2,2)
subvencionado	-0,10*	(1,7)	-0,09	(1,7)	0,13	(2,1)	0,11	(2,1)
ingreso _{especial}	0,20	(1,3)	0,26*	(1,3)	1,09**	(3,4)	1,24***	(3,6)
ingreso _{frontera}	0,01	(1,2)	-0,03	(1,2)	-0,16	(1,0)	-0,06	(1,0)
PLAT _{std}			-0,00	(2,7)			0,02*	(2,8)
PLAT _{skw}			-0,04	(1,5)			-0,13	(1,2)
PLAT _{loc}			0,73***	(1,2)			0,18	(1,1)
ln(ACA _{mean})			0,22***	(2,3)				
ln(ADM _{mean})			0,21***	(2,1)			-0,03	(1,8)
ln(WC _{mean})			0,28***	(1,2)			0,33	(1,3)
ln(TEST _{mean})			-0,13	(1,2)			0,10	(1,3)
ln(RC _{mean})								(1,8)
ln(Offtask _{mean})							0,19*	(2,0)
Constante	-12,69***		-13,97***		-2,84***		-3,04***	
Observaciones	2805		2805		838		838	
R ² Ajustado	0,284		0,364		0,120		0,145	
F _{Wald}			25,95				3,22	
p-valor _{Wald}			0,000***				0,001**	

VIF en paréntesis

* p<0,05, ** p<0,01, *** p<0,001

Tabla 5.6: Regresión lineal múltiple zscore_{carrera}.

más las distintas categorías de U-Cursos tienen mejores resultados al final del semestre (salvo la variable $\ln(test_{mean})$), lo cual no es posible determinar con certeza en la UC.

5.5.4. Regresión Lineal Múltiple con información del desempeño histórico

El objetivo de la regresión anterior no es tener un alto poder predictivo, sino estudiar la significancia de los regresores provenientes de la actividad de estudiantes en LMS en la explicación de la variable dependiente. Sin embargo, el bajo R^2 ajustado podría ser síntoma de alguna variable omitida, causando inestabilidad en la varianza de los estimadores y por lo tanto en la significancia de los predictores. Dado que el desempeño académico histórico según la literatura es uno de los mejores predictores del desempeño en el futuro, surge el interés de

repetir el ejercicio de la regresión lineal agregando esta variable.

En la práctica, lo anterior se verifica particularmente en el caso de la UChile. Definiendo las variables $zscore_{carrera1}$ como el zscore obtenido en el primer semestre y $zscore_{carrera2}$ en el segundo semestre, en la figura 5.14 se puede observar como existe una alta correlación del desempeño en ambos semestres para estudiantes de esta institución.

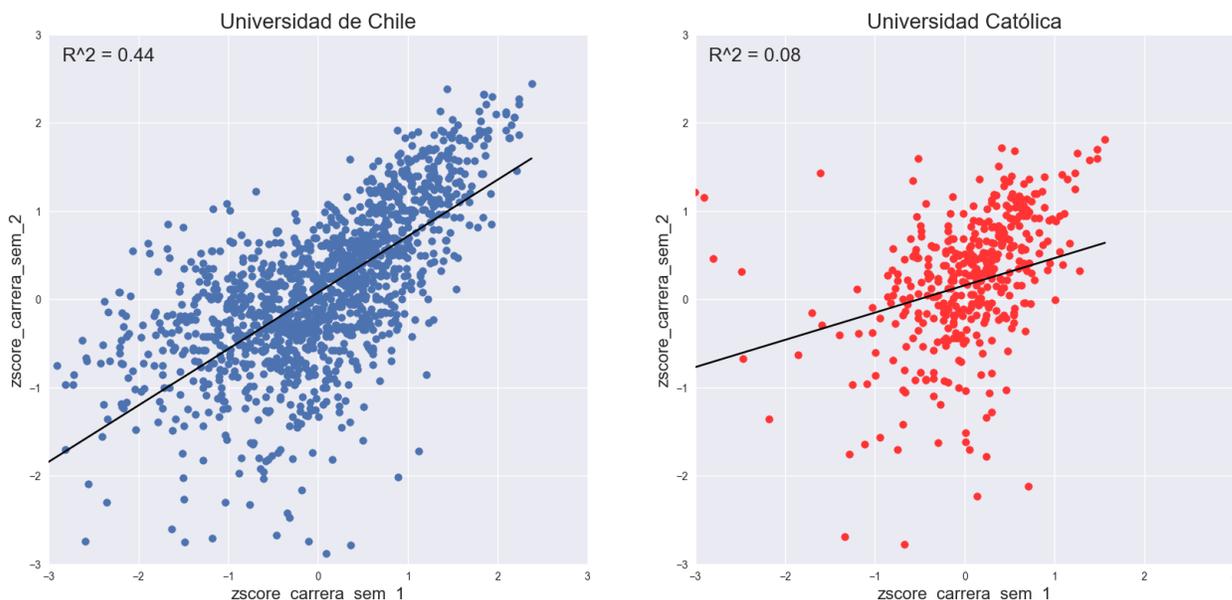


Fig. 5.14: $zscore_{carrera}$ del primer semestre vs $zscore_{carrera}$ del segundo semestre. Izquierda: UChile, Derecha: UC.

De esta forma, se realiza otra regresión lineal múltiple, esta vez sólo para la submuestra correspondiente a estudiantes en el segundo semestre¹⁷, agregando como variable de control el desempeño académico en el primer semestre medido por la variable $zscore_{carrera1}$, lo que equivale a agregar a las ecuaciones 5.1 y 5.2 el término $\beta_{20} zscore_{carrera1}$ y reemplazar la variable dependiente $zscore_{carrera}$ por $zscore_{carrera2}$. Los resultados de esta nueva regresión se presentan en la tabla 5.7. Se observa que los R^2 mejoran sustancialmente en la UChile. En el caso de U-Cursos las variables de la actividad online ayudan a explicar un 5,8 % más de la varianza (11,2 % del total explicado), mientras que en Sakai un 6,6 % (30,1 % del total explicado). Al respecto se señala que:

- Como era previsto, la variable $zscore_{carrera1}$ resulta fuertemente explicativa del desempeño en el segundo semestre (significancia al 1 %). La variable psu_{mat} en U-Cursos deja de ser significativa, probablemente asociado a la inclusión de $zscore_{carrera1}$.
- Los parámetros asociados a la variable F (sexo femenino) ya no son significativos en U-Cursos. No hay evidencia de diferencias en los resultados de estudiantes de distintos sexos en el segundo semestre. De la misma forma, ya no se observan diferencias en los resultados por parte de estudiantes egresados de distintos tipos de colegio. Esto puede deberse a un período de ‘nivelación’ durante el primer semestre de la universidad.

¹⁷Existen estudiantes que sólo cursaron primer semestre, o no finalizaron su segundo semestre. Por razones obvias, se retiran de la muestra. Además, dado que ya no existe el problema de dependencia entre observaciones, se utiliza el estimador robusto de la varianza de Eicker–White.

Variables	(1) UChile		(2) UChile ₂		(3) UC		(4) UC ₂	
	zscore _{carrera2}		zscore _{carrera2}		zscore _{carrera2}		zscore _{carrera2}	
nem	0,00***	(1,3)	0,00***	(1,3)	0,00***	(3,6)	0,00***	(3,7)
psu _{mat}	0,00	(1,6)	0,00	(1,7)				
psu _{len}	-0,00	(1,2)	-0,00	(1,2)				
psu _{esp}	0,00**	(1,5)	0,00**	(1,5)				
F	-0,05	(1,1)	-0,06	(1,2)	0,33	(1,1)	0,31	(1,1)
ped _{gen_basica}					-0,19	(1,1)	-0,11	(2,0)
privado	-0,06	(1,6)	-0,05	(1,7)	0,11	(2,1)	0,15	(2,2)
subvencionado	0,03	(1,7)	0,02	(1,7)	0,01	(2,1)	0,01	(2,1)
ingreso _{especial}	0,43***	(1,3)	0,47***	(1,3)	1,50**	(3,3)	1,98***	(3,4)
ingreso _{frontera}	0,06	(1,2)	0,06	(1,2)	-0,31*	(1,0)	-0,34*	(1,0)
zscore _{carrera1}	0,60***	(1,6)	0,57***	(1,7)	0,20**	(1,2)	0,17*	(1,2)
PLAT _{std}			0,00	(2,9)			0,01	(2,9)
PLAT _{skw}			-0,03	(1,6)			-0,12	(1,2)
PLAT _{loc}			0,78***	(1,3)			0,19	(1,3)
ln(ACA _{mean})			0,20***	(2,6)				
ln(ADM _{mean})			0,11	(2,6)			0,16	(2,0)
ln(WC _{mean})			0,18***	(1,2)			0,53	(1,2)
ln(TEST _{mean})			-0,37***	(1,2)			0,56	(1,3)
ln(RC _{mean})							-0,09	(1,5)
ln(Offtask _{mean})							0,23*	(2,1)
Constante	-2,53***		-4,50***		-2,64***		-3,42***	
Observaciones	1373		1373		404		404	
R ² Ajustado	0,459		0,517		0,148		0,214	
F _{Wald}			16,65				3,93	
p-valor _{Wald}			0,000***				0,000***	

VIF en paréntesis

* p<0,05, ** p<0,01, *** p<0,001

Tabla 5.7: Regresión lineal múltiple zscore_{carrera2}.

- Las conclusiones derivadas de variables de uso de las plataformas a modo general se mantienen, salvo aquella que indicaba que los estudiantes menos constantes tienen mejores resultados, lo cual no se observa en estas regresiones.

Con el fin de comparar las magnitudes de los coeficientes estimados, se presentan en la tabla 5.8 los coeficientes estandarizados¹⁸ de las variables independientes para las regresiones (2) y (4). Estos permiten determinar cuales variables independientes tienen mayor impacto en la variable dependiente. En el caso de la Universidad de Chile, son las variables zscore_{carrera1}

¹⁸Esto corresponde a normalizar las variables independientes (que pasan a tener media 0 y varianza 1) y calcular el estimador $\hat{\beta}_{MCO}$. La interpretación es en términos de desviaciones estándar, por ejemplo, un coeficiente estándar de 0,5 indica que un aumento de una desviación estándar de la variable independiente implica un aumento de 0,5 desviaciones estándar en la variable dependiente.

(0,59), nem (0,149), y $plat_{loc}$ (0,144). En la Universidad Católica estas corresponden a nem (0,476), $ingreso_{especial}$ (0,339), y $zscore_{carrera1}$ (0,164).

	(2)	(4)
	UChile ₂	UC ₂
	$ZSCORE_{carrera2}$	$ZSCORE_{carrera2}$
nem	0,149	0,476
psu_{mat}	0,04	
psu_{len}	-0,006	
psu_{esp}	0,078	
F	-0.026	0,064
ped_{gen_basica}		-0,05
$privado$	-0,028	0,074
$subvencionado$	0,01	0,003
$ingreso_{especial}$	0,095	0,339
$ingreso_{frontera}$	0,014	-0,03
$zscore_{carrera1}$	0,59	0,164
$PLAT_{std}$	0,037	0,092
$PLAT_{skw}$	-0,023	-0,074
$PLAT_{loc}$	0,144	0,052
$\ln(ACA_{mean})$	0,116	
$\ln(ADM_{mean})$	0,057	0,076
$\ln(WC_{mean})$	0,066	0,054
$\ln(TEST_{mean})$	-0,092	0,082
$\ln(RC_{mean})$		-0,036
$\ln(Offtask_{mean})$		0,139

Tabla 5.8: Coeficientes estandarizados para las regresiones (2) y (4).

5.5.5. Limitaciones

Una de las hipótesis del estimador de MCO es que los errores no se encuentran correlacionados con las variables dependientes. Cuando se viola este supuesto surge un problema de endogeneidad, el cual puede sesgar los coeficientes estimados y llevar a falsas conclusiones al realizar tests de hipótesis. En el caso de las regresiones aquí realizadas, la endogeneidad puede deberse a un error de medición, causalidad inversa, por problema de simultaneidad, sesgo de selección y de variables omitidas:

- **Error de medición:** La única variable que podría ser objeto de críticas por estar mal medida (salvo error en las bases de datos) es nem , en tanto el mismo alumno en dos colegios distintos podría tener un nem diferente (dado que es una variable que depende de la nota del estudiante en la enseñanza media y se mide de forma distinta en cada colegio). Sin embargo, esto no corresponde, dado que esta variable puede considerarse como el desempeño del estudiante en el colegio dado su contexto académico.

- **Causalidad inversa y simultaneidad:** Estos problemas podrían encontrarse si la variable $zscore_{carrera}$ causase alguna de las otras variables. Debido que la nota que obtiene el estudiante se da después (al final del semestre) que los demás factores, no puede existir ni causalidad inversa ni simultaneidad.
- **Sesgo de selección:** Existiría si se considerase a solo parte del universo de cada carrera para estimar la variable independiente ($zscore$). Es equivalente a, por ejemplo, intentar estimar las causas de buenos o malos rendimientos tomando como muestra sólo estudiantes de alto rendimiento. Como este no es el caso, se descarta sesgo de selección.
- **Variables omitidas:** Para tener variables omitidas se deben cumplir dos condiciones. 1) La variable omitida debe ser determinante de la variable dependiente (existen razones para creer que si se incluye en la regresión su coeficiente no sería nulo) y 2) la variable omitida se encuentra correlacionada con alguna variable independiente de la regresión. Existe la posibilidad que en las regresiones realizadas existan variables omitidas¹⁹; sin embargo, no es posible verificarlo sin efectivamente tener la variable o un instrumento adecuado de ella, lo cual no es el caso. Las opciones para solucionar este problema son (1) utilizar variables instrumentales o un experimento natural y (2) obtener variables que midan otras dimensiones de los procesos de aprendizaje de los estudiantes. Es esta última la dirección a la que apunta el proyecto FONDECYT mediante la elaboración del instrumento LEARN+.

5.5.6. Conclusiones Modelo Explicativo

Por un lado, los resultados de la presente sección sugieren la mayor facilidad con la que se puede predecir desempeño académico utilizando las variables señaladas para estudiantes de la FCFM que de educación UC. Al respecto, se remarca la importancia de analizar los detalles de cada contexto educativo al momento de elaborar modelos predictivos, pues este resulta relevante en la selección de atributos. De forma particular en los LMS, es necesario que el rol de la plataforma en la comunidad educativa sea importante para que incluir variables que representen su uso sea un aporte a la mejora de los resultados del modelo.

Por otro lado, en el caso particular de la UChile se muestra cómo es posible incluir regresores derivados de la actividad online de estudiantes para mejorar los resultados de predicción de desempeño. Estos regresores resultan estadísticamente significativos y mejoran el R^2 ajustado en un rango del 11,2 al 21,98 % comparado a cuando no son incluidos. Al respecto, se identifica que (1) las personas que se conectan más hacia el final del semestre tienden a tener mejores resultados académicos, lo cual puede tener origen en el mayor porcentaje de la nota final en juego durante los últimos controles y el examen, (2) realizar actividades del tipo contenido académico y escribir comentario tienen un impacto positivo en el desempeño, reflejando cercanía a los procesos de aprendizaje y (3) realizar más actividades del tipo test tiene un impacto negativo en el desempeño, resultado contraintuitivo que podría estar influido por

¹⁹Por ejemplo, el enfoque del alumno al aprendizaje (superficial, profundo o estratégico) podría ser determinante tanto de su desempeño histórico (*nem*, *psu*), de la forma en la que utilice las plataformas (suponiendo por ejemplo, que un estudiante de tipo superficial no interactúa en el foro), y de la nota que obtiene al final del semestre.

la baja cuenta total de este tipo de actividades, que no se utiliza en todos los cursos. En la UC, en cambio, la baja varianza explicada no permite concluir que incluir estas variables en un modelo sea un aporte importante.

Se logra validar para la muestra analizada uno de los antecedentes importantes encontrados en la literatura: el desempeño histórico está fuertemente correlacionado con el desempeño futuro. Además, se verifica una de las ideas centrales que fundan este estudio: Incorporar información derivada de la actividad online proveniente de los registros de las plataformas en la predicción de desempeño, las cuales representan medidas objetivas del comportamiento de los estudiantes a diferencia de la información auto reportada, permite una mejora de los resultados obtenidos.

Capítulo 6

Conclusión

6.1. Conclusiones generales

La investigación realizada en esta memoria ha tenido un enfoque en analizar, desde un punto de vista cuantitativo con herramientas derivadas de EDM y LA, cómo los estudiantes utilizan los sistemas de gestión del aprendizaje que tienen a disposición, y en comprender las relaciones que existen entre este uso y el desempeño académico en dos carreras y universidades diferentes.

Al respecto, se logran identificar diferencias y semejanzas en el uso de las plataformas U-Cursos y Sakai. Por un lado, el rol primordial que los estudiantes de la FCFM le dan a U-Cursos es el de una plataforma que permite la comunicación y el intercambio de información. Por otro lado, en educación UC, Sakai tiene un enfoque en la gestión del contenido y en las actividades de tipo académicas. También, se observa en ambas instituciones que estudiantes que utilizan con más frecuencia las herramientas que tienen a su disposición obtienen mejores resultados académicos al final del semestre. Se observan patrones que, en línea con la literatura revisada, sugieren una cultura de uso distinta en cada institución, donde estudiantes interactúan con las plataformas de acuerdo a una combinación de factores relativos tanto a características propias del alumnado y del equipo docente; a las instituciones que fomentan y moldean el rol del LMS; y a la interfaz y funcionalidades de cada una de ellas. En otro ámbito, se destaca la importancia de visualizar las bases de datos constituidas por gran cantidad de observaciones y variables para descubrir información subyacente. En este trabajo se desarrollaron distintos tipos de gráficos con la información disponible, los cuales permitieron analizar conductas de distintos grupos de estudiantes.

La regresión llevada a cabo permite confirmar la línea de trabajo planteada por el proyecto FONDECYT. Es posible incluir variables de este tipo para obtener mejores resultados en la predicción de desempeño académico, en función del rol y del grado de apropiación que el LMS tenga dentro de la comunidad educativa. En la Universidad de Chile incorporar información de los LMS ayuda a explicar entre un 5,8 y un 8 % más de la varianza, lo que equivale a una mejora entre el 11,2 y 21,9 % del R^2 ajustado cuando no son incluidos, respectivamente. En el caso de Sakai también se observa una mejora en la varianza explicada por el modelo; sin

embargo, el bajo R^2 ajustado obtenido hace dudar de la calidad del mismo.

De este modo se logra determinar para el universo estudiado que, tal como se encuentra en la literatura revisada, el desempeño histórico está fuertemente correlacionado con el desempeño futuro. Además, se valida una de las hipótesis en las que se basa esta investigación: Utilizar datos de la actividad de los estudiantes en los LMS para predecir desempeño permite mejorar los resultados que se obtienen. Los beneficios de usar este tipo de variables son diversos, entre los que destacan: (1) data obtenida a partir de los logs de LMS no es sutil al sesgo por deseabilidad social, a diferencia de información auto reportada por los estudiantes en encuestas, (2) es información que puede obtenerse en tiempo real, permitiendo actualizar los modelos de predicción durante el transcurso del semestre y (3) permite incorporar nuevas dimensiones de análisis con información que es difícil de obtener por otros medios, como por ejemplo las medidas de centralidad de un sociograma representativas de la actividad en un foro.

Por último, también queda en evidencia la dificultad de predecir rendimiento académico. Como es usual en modelos que involucran el comportamiento de personas, existe una parte importante de la varianza no explicada por características sociodemográficas de los estudiantes, ni por su desempeño histórico, ni por la forma de usar las herramientas que tienen a su disposición. Lo anterior sugiere que existe un amplio espacio de superación académica en los estudiantes que no es condicional a estas variables, y procesos de aprendizaje que no se ven reflejados en este tipo de datos. Se espera poder explicar de mejor manera estos procesos en base a información recolectada por el cuestionario LEARN+.

6.2. Toma de decisiones basada en datos

Tras los resultados obtenidos en el desarrollo del trabajo nace la oportunidad de hacer uso de esta información. Si bien la última etapa del proyecto Fondecyt corresponde a la elaboración de una plataforma que aúne resultados de distintas fuentes que permitan: concebir el aprendizaje de forma holística, evaluar la situación académica de los estudiantes, y tomar decisiones respecto a ella, aquí se discuten tres posibles medidas para apoyar a los estudiantes en función de los resultados parciales obtenidos.

En primer lugar, con los resultados de la regresión se puede determinar cuales elementos tienen impacto positivo en el rendimiento, dando lugar a tomar medidas para potenciarlos. En el caso de la Universidad de Chile, estas corresponden a actividades donde el estudiante se relaciona con sus compañeros y profesores mediante el foro, y accediendo al material académico del curso. Debido a lo anterior se sugiere incluir en la planificación docente el entregar a parte del equipo docente la responsabilidad explícita de fomentar los intercambios estudiante-estudiante y estudiante-profesor mediante el LMS, así como de mantener el material del curso accesible y válido¹ durante el semestre. En la Universidad Católica, los resultados no son concluyentes en tanto el uso de la plataforma es bajo y los estudiantes la utilizan mayoritariamente con fines administrativos; se espera poder identificar oportunidades de intervención interrogando directamente a los estudiantes en la siguiente fase del

¹De acuerdo a las definiciones de Naveh et al. (2012).

proyecto Fondecyt.

En segundo lugar, los datos muestran la existencia del hábito de algunos alumnos con desempeño final bajo de trabajar hasta altas horas de la madrugada. Esta información puede ser transmitida al área de bienestar estudiantil para su uso, la cual podría realizar talleres de hábitos de estudio personalizados, entrevistar a los alumnos en cuestión con el fin de determinar y corregir las causas de este comportamiento, o bien recomendar reducir la carga de trabajo.

Por último, es posible informar a los estudiantes que ingresan a la universidad sobre las prácticas que han tomado un rol importante en el desempeño de estudiantes de años superiores, motivando de esta forma su adopción temprana. Sobre este punto, es necesario desarrollar una estrategia comunicativa que permita evitar suponer que realizar este tipo de prácticas implica tener buen rendimiento, o situaciones del tipo profecía autocumplida en la cual informar a los estudiantes tenga un impacto negativo en su desempeño. Del mismo modo, es posible automatizar la generación de las visualizaciones realizadas con el fin de retroalimentar en forma continua a los estudiantes.

6.3. Investigaciones futuras

El análisis de las plataformas deja entrever la posibilidad de medir y monitorear ciertas dimensiones de los procesos de aprendizaje utilizando datos de la actividad online de los estudiantes, entregando información en tiempo real sobre distintas variables que permitan apoyar a la toma de decisiones académicas. Por ejemplo, midiendo para cada estudiante su nivel de organización observando la varianza en la frecuencia de sus conexiones; medir cuanto se relaciona con sus pares a través de la participación en los foros; o determinar sus hábitos de estudio y carga de trabajo observando cuanto se conecta al LMS durante la madrugada y los fines de semana. Para validar este tipo de variables proveniente LMS como medidas de distintos constructos de los estudiantes, es necesario primero verificar que exista una correlación entre ellas y pruebas, experimentos o instrumentos que ya hayan sido previamente validados como proxy del constructo. La aplicación del cuestionario LEARN+ del proyecto FONDECYT para obtener percepciones auto reportadas de los estudiantes podría permitir realizar este tipo de trabajo en un futuro.

Por el lado del modelo a desarrollar, con el fin de obtener un mejor poder predictivo en miras al prototipo que será elaborado por el equipo se sugiere formular el problema como una clasificación de estudiantes (aprueba o no reprueba), y utilizar modelos con enfoque en reducir errores del tipo I y II, donde la interpretación del efecto que cada variable tiene menos relevancia. Así mismo, se recomienda en primera instancia separar grupos de estudiantes de acuerdo a su contexto educativo (institución, semestre, curso o sección), para luego entrenar modelos de forma separada. Esto debido a que variables relevantes en un contexto podrían no serlo en otro, como pudo observarse en el trabajo realizado. Lo anterior podría ser realizado mediante algoritmos como arboles clasificadores, bosques aleatorios o XGBoost, que seleccionan los atributos predictivos en función de medidas de la información entregada por las variables, como la impureza de Gini o la divergencia de Kullback-Leibler.

Con respecto al punto 3 de la sección 2.6 Alcances, resulta claro que el uso de versiones móviles de los LMS da lugar a una multitud de posibles trabajos en Student Learning Research, Educational Data Mining y de Learning Analytics. Queda propuesto integrar al análisis realizado en este tema de memoria ese tipo de datos, así como el entender las similitudes y diferencias del comportamiento de estudiantes con las versiones de escritorio.

Para finalizar, remarcar como el entender los procesos de enseñanza-aprendizaje que ocurren en distintos contextos educativos permite identificar, desarrollar y utilizar las metodologías y herramientas que mejor se adecuen a cada comunidad académica. Dada la gran importancia que han adquirido los sistemas de gestión del aprendizaje en educación superior, existe la oportunidad no solo de contar con esta herramienta, sino de comprender como los estudiantes la utilizan para poder incluirla en la planificación tanto a nivel institucional como docente, haciendo uso idóneo de las funcionalidades que ellas poseen, y potenciando el rol activo de los estudiantes para permitir un enfoque de tipo profundo hacia el aprendizaje.

Glosario

ACA: Academic Content

ADI: Área De Infotecnologías

ADM: Administrative Content

ANOVA: Analysis of Variance

CGPA: Cumulative Grade Point Average

DM: Data Mining

EDM: Educational Data Mining

FCFM: Facultad de Ciencias Físicas y Matemáticas

FONDECYT: Fondo Nacional de Desarrollo Científico y Tecnológico

KDD: Knowledge Discovery in Databases

LA: Learning Analytics

LMS: Learning Management System

MCO: Mínimos Cuadrados Ordinarios

MOOC: Massive Open Online Course

RC: Read Comment

SNAPP: Social Networks Adapting Pedagogical Practice

UC: Pontificia Universidad Católica de Chile

UChile: Universidad de Chile

VIF: Variance Inflation Factor

WC: Write Comment

Bibliografía

- Abelson, R. P. (1985). A Variance Explanation Paradox: When a Little is a Lot. *Psychological Bulletin*, 97(1), 129–133.
- Baker, R. S. (2007). Modeling and understanding students' off-task behavior in intelligent tutoring systems. *Proceedings of the SIGCHI Conference on Human Factors in computing Systems*, 1059.
- Bakharia, A., y Dawson, S. (2011). SNAPP: A Bird's-Eye View of Temporal Participant Interaction AneeshaBakharia. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge - LAK '11*, 168.
- Balmaceda Hidalgo, S. I. (2012). Diseño de una comunidad virtual de aprendizaje para estudiantes de primer año de Ingeniería de la Universidad de Chile. *Departamento de Ingeniería Industrial, Universidad de Chile*.
- Beck, J. (2004). Using response times to model student disengagement. *Proceedings of the ITS2004 Workshop on Social and Emotional Intelligence in Learning Environments*, 13–20.
- Bramucci, R., y Gaston, J. (2012). Sherpa. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK '12*(May), 82.
- Calvet Liñán, L., y Juan Pérez, Á. A. (2015). Educational Data Mining and Learning Analytics: differences, similarities, and time evolution. *RUSC. Universities and Knowledge Society Journal*, 12(3), 98. doi: 10.7238/rusc.v12i3.2515
- Castillo, C. P. A. (2016). Superación Académica En Primer Año De Ingeniería Y Ciencias: Mecanismos De Permanencia Y Mejoramiento Académico. *Departamento de Ingeniería Industrial, Universidad de Chile*.
- Celis, S., Moreno, L., Poblete, P., Villanueva, J., y Weber, R. (2015). Un modelo analítico para la predicción del rendimiento académico de estudiantes de ingeniería. *Revista Ingeniería de Sistemas*, 29(1), 5–24.
- Dringus, L. L. P. (2012). Learning Analytics Considered Harmful. *Journal of Asynchronous Learning Networks*, 16(3), 87–101.
- Engle, R. F. (1984). Chapter 13 Wald, likelihood ratio, and Lagrange multiplier tests in

- econometrics. En *Handbook of econometrics* (Vol. 2, pp. 775–826).
- Fathema, N., Shannon, D., y Ross, M. (2015). *Expanding The Technology Acceptance Model (TAM) to Examine Faculty Use of Learning Management Systems (LMSs) In Higher Education Institutions* (Tesis Doctoral no publicada).
- Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5/6), 304–317.
- Gasevic, D., Dawson, S., Rogers, T., y Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *Internet and Higher Education*, 28, 68–84.
- Ibarra, M. d. C., y Michalus, J. C. (2010). Análisis Del Rendimiento Academico Mediante Un Modelo Logit * Academic Performance Analysis Through a Logit Model. *Revista Ingeniería Industrial*, 47–56.
- Kovanović, V., Gašević, D., Dawson, S., Joksimović, S., Baker, R. S., y Hatala, M. (2015). Penetrating the black box of time-on-task estimation. *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge - LAK '15*(October), 184–193.
- Leony, D., Pardo, A., de la Fuente Valentín, L., Sánchez De Castro, D., y Delgado Kloos, C. (2012). GLASS: a learning analytics visualization tool. *Proceedings of the 2nd international conference on learning analytics and knowledge*(May), 162–163.
- Macfadyen, L. P., y Dawson, S. (2010). Mining LMS data to develop an “early warning system for educators”: A proof of concept. *Computers and Education*, 54(2), 588–599.
- Malcolm, J., y Zukas, M. (2001). Bridging Pedagogic Gaps: conceptual discontinuities in higher education. *Teaching in Higher Education*, 6(1), 33–42.
- Marton, F., y Säljö, R. (1976). On Qualitative Differences in Learning — II Outcome as a Function of the Learner’s Conception of the Task. *British Journal of Educational Psychology*, 46(1947), 115–127.
- Mazza, R., y Dimitrova, V. (2004). Visualising student tracking data to support instructors in web-based distance education. *Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters - WWW Alt. '04*, 154.
- Miller, A. L. (2012). Investigating social desirability bias in student self-report surveys. *Educational Research Quarterly*, 36, 30.
- Naveh, G., Tubin, D., y Pliskin, N. (2012). Student satisfaction with learning management systems: a lens of critical success factors. *Technology, Pedagogy and Education*, 21(3), 337–350. Descargado de <https://www.tandfonline.com/doi/full/10.1080/1475939X.2012.720413> doi: 10.1080/1475939X.2012.720413
- Parpala, A., y Lindblom-Ylänne, S. (2012). Using a research instrument for developing quality at the university. *Quality in Higher Education*, 18(3), 313–328.

- Prosser, M., y Trigwell, K. (1999). *Understanding Learning and Teaching: The Experience in Higher Education*.
- Ramsden, P. (1979). Student learning and perceptions of the academic environment. *Higher Education*, 8(4), 411–427.
- Rogers, W. H. (1993). Regression standard errors in clustered samples. *Stata Technical Bulletin*, 13(May), 19–23.
- Romero, C., y Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 33(1), 135–146.
- Romero, C., Ventura, S., y García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers and Education*, 51(1), 368–384.
- Scott, D., Spielmans, G. G., Julka, D., y DeBerard, M. S. (2004). Predictors of Academic Achievement and Retention Among College Freshmen: a Longitudinal Study. *College Student Journal*, 38(1), 66–80.
- Tamhane, A. (2009). *Statistical Analysis of Designed Experiments: Theory and Applications*. John Wiley & Sons.
- Tandi Lwoga, E. (2014). Critical success factors for adoption of web-based learning management systems in Tanzania. *International Journal of Education and Development using Information and Communication Technology*, 10(1), 4–21.
- Treaster, J. B. (2017, feb). *Will You Graduate? Ask Big Data*. Descargado de <https://www.nytimes.com/2017/02/02/education/edlife/will-you-graduate-ask-big-data.html>
- Trigwell, K., Caballero Rodriguez, K., y Han, F. (2012). Assessing the impact of a university teaching development programme. *Assessment & Evaluation in Higher Education*, 37(4), 499–511.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, 48(4), 817–838.
- Wolfe, R. N., y Johnson, S. D. (1995). Personality as a predictor of college performance. *Educational and psychological measurement*, 55(2), 177–185.

Anexos

Anexo A

Tipo de Ingreso	Categoría	Universidad
PAA o PSU	Regular	U. de Chile
Carreras Paralelas	Regular	PUC
No Especial	Regular	PUC
Deportista	Especial	U. de Chile
Extranjero	Especial	U. de Chile
Sistema de Ingreso Prioritario de Equidad Educativa	Especial	U. de Chile
Cambio Carrera Santiago	Especial	PUC
Cambio de Universidad	Especial	PUC
5% Colegios Municipalizados	Frontera	U. de Chile
Equidad de Genero	Frontera	U. de Chile
Beca Excelencia Académica	Frontera	PUC
Hijo de Profesor PUC	Frontera	PUC
Ingreso Especial	NA	U. de Chile
Cambio Carrera y Universidad	NA	PUC
Cambio de Carrera por Vocación	NA	PUC
Enseñanza Media en el Extranjero	NA	PUC
Egresado o Titulado	NA	PUC
Reingreso Abandono	NA	PUC

Anexo A: Categorías de los tipos de ingreso a la universidad.

Anexo B

Curso	Nombre	Carrera
MAT1920	Números	Pedagogía Básica
MAT1930	Geometría I	Pedagogía Básica
EDU0310	Aprendizaje y desarrollo: teoría y práctica	Pedagogía Básica
LET001B	Fundamentos lingüísticos	Pedagogía Básica
EDU0315	Aprendizaje y desarrollo escolar	Pedagogía Básica
EDU0311	Teoría de la educación	Pedagogía Básica
EDU0318	Curriculum en la educación básica	Pedagogía Básica
LET002B	Oralidad y desarrollo semiótico infantil	Pedagogía Básica
EDU0338	practica i: educación básica	Pedagogía Básica
EDU0329	Desarrollo y enseñanza del lenguaje artístico visual	Pedagogía Básica
EDU0428	Aprendizaje y desarrollo del niño I: teoría y práctica	Párvulo
MAT1900	Números, geometría, datos y azar para la ed. parvularia	Párvulo
EDU0424	Adquisición y desarrollo de la lengua	Párvulo
EDU0427	Ciencias sociales	Párvulo
EDU0441	Introducción al campo profesional de la ed. parvularia	Párvulo
EDU0420	Curriculum y modalidades curriculares en ed. parvularia	Párvulo
EDU0425	Literatura en la infancia	Párvulo
MAT0100	Razonamiento cuantitativo	Párvulo
TSL100P	Familia, comunidad y educación	Párvulo
EDU0421	Fundamentos de la educación parvularia	Párvulo
BIO100P	Ciencias naturales	Párvulo
MA1002	Cálculo dif. e integral	Ing. Plan Común
CM1001	Química	Ing. Plan Común
MA1101	Intro. al álgebra	Ing. Plan Común
MA1001	Intro. al cálculo	Ing. Plan Común
FI1001	Intro. a la física newtoniana	Ing. Plan Común
CC1001	Computación	Ing. Plan Común
MA1102	Álgebra lineal	Ing. Plan Común
FI1002	Sistemas newtonianos	Ing. Plan Común
CC1000	Herramientas computacionales para Ing. y Cs.	Ing. Plan Común

Anexo B: Lista de cursos analizados.

Anexo C

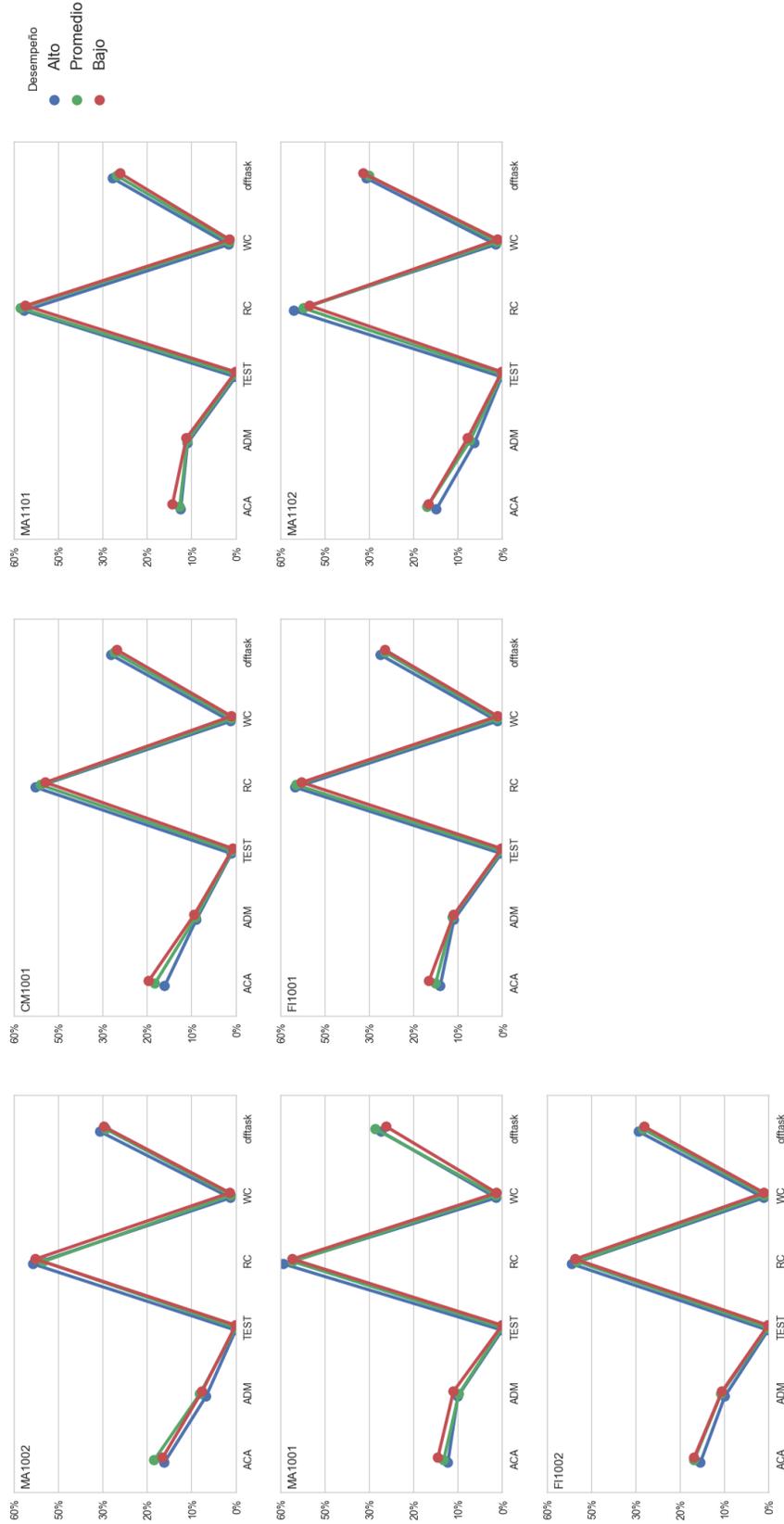
Event	Platform	Category	Action
content.read	Sakai	Academic Content	Pasivo
melete.section.read	Sakai	Academic Content	Pasivo
webcontent.read	Sakai	Academic Content	Pasivo
syllabus.read	Sakai	Administrative Content	Pasivo
forums.read	Sakai	Read Comment	Pasivo
messages.read	Sakai	Read Comment	Pasivo
news.read	Sakai	Read Comment	Pasivo
poll.viewresult	Sakai	Read Comment	Pasivo
wiki.read	Sakai	Read Comment	Pasivo
asn.revise.assignment	Sakai	Test	Activo
sam.assessment.submit	Sakai	Test	Activo
sam.assessment.take	Sakai	Test	Activo
asn.read.submission	Sakai	Test	Pasivo
asn.revise.assignmentcontent	Sakai	Test	Pasivo
chat.new	Sakai	Write Comment	Activo
forums.newtopic	Sakai	Write Comment	Activo
forums.response	Sakai	Write Comment	Activo
mail.create	Sakai	Write Comment	Activo
messages.forward	Sakai	Write Comment	Activo
messages.new	Sakai	Write Comment	Activo
messages.reply	Sakai	Write Comment	Activo
poll.vote	Sakai	Write Comment	Activo
wiki.new	Sakai	Write Comment	Activo
wiki.revise	Sakai	Write Comment	Activo
Bajar_Material_Alumnos	U-Cursos	Academic Content	Pasivo
Bajar_Material_Docente	U-Cursos	Academic Content	Pasivo
Ver_Bibliografia	U-Cursos	Academic Content	Pasivo
Ver_Enlaces	U-Cursos	Academic Content	Pasivo
Ver_Material_Alumnos	U-Cursos	Academic Content	Pasivo
Ver_Material_Docente	U-Cursos	Academic Content	Pasivo
Visitar_Enlace	U-Cursos	Academic Content	Pasivo
Buscar_Buscador	U-Cursos	Administrative Content	Activo
Resumen_Evaluaciones_Notas	U-Cursos	Administrative Content	Pasivo
Ver_Buscador	U-Cursos	Administrative Content	Pasivo
Ver_Calendario	U-Cursos	Administrative Content	Pasivo
Ver_Datos_Curso	U-Cursos	Administrative Content	Pasivo
Ver_Horario	U-Cursos	Administrative Content	Pasivo
Ver_Notas	U-Cursos	Administrative Content	Pasivo
Leer_Mensaje_Foro	U-Cursos	Read Comment	Pasivo
Recibir_Correo	U-Cursos	Read Comment	Pasivo
Ver_Bandeja_Salida_Correo	U-Cursos	Read Comment	Pasivo
Ver_Blog	U-Cursos	Read Comment	Pasivo

Continúa en la siguiente página

Ver_Correo	U-Cursos	Read Comment	Pasivo
Ver_Encuesta	U-Cursos	Read Comment	Pasivo
Ver_Foro	U-Cursos	Read Comment	Pasivo
Ver_Mensajes	U-Cursos	Read Comment	Pasivo
Ver_Novedades	U-Cursos	Read Comment	Pasivo
Ver_Votaciones	U-Cursos	Read Comment	Pasivo
Entregar_Tarea	U-Cursos	Test	Activo
Responder_Pregunta	U-Cursos	Test	Activo
Descargar_Tarea	U-Cursos	Test	Pasivo
Descargar_Tarea_Entregada	U-Cursos	Test	Pasivo
Ver_Detalle_Tarea	U-Cursos	Test	Pasivo
Ver_Historial_Tarea	U-Cursos	Test	Pasivo
Ver_Tarea	U-Cursos	Test	Pasivo
Ver_Test	U-Cursos	Test	Pasivo
Agregar_Material_Alumnos	U-Cursos	Write Comment	Activo
Agregar_Post_Blog	U-Cursos	Write Comment	Activo
Crear_Tema_Foro	U-Cursos	Write Comment	Activo
Editar_Mensaje	U-Cursos	Write Comment	Activo
Enviar_Correo	U-Cursos	Write Comment	Activo
Escribir_Mensaje_Foro	U-Cursos	Write Comment	Activo
Responder_Encuesta	U-Cursos	Write Comment	Activo
Responder_Votacion	U-Cursos	Write Comment	Activo

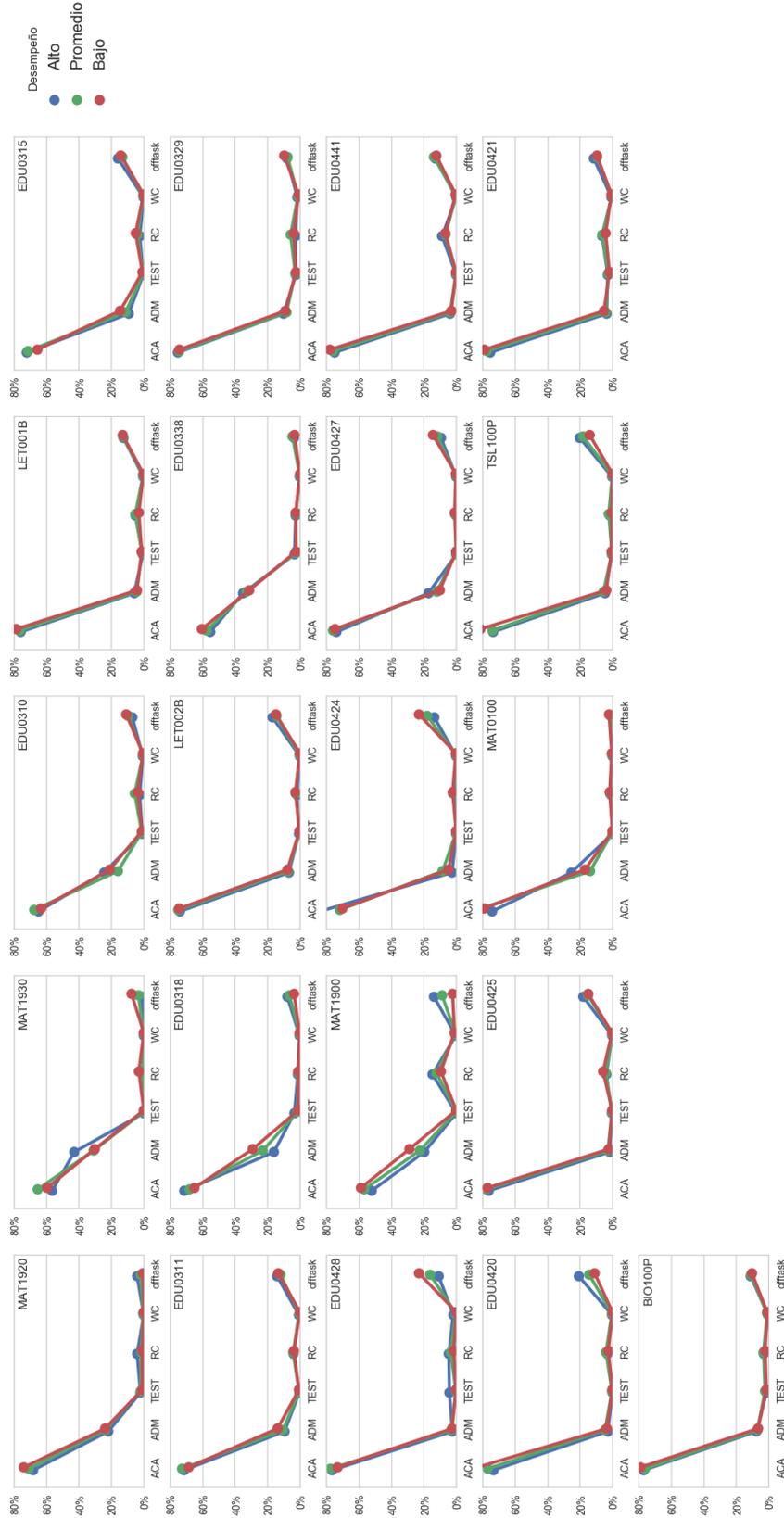
Anexo C: Categorización de los distintos eventos dentro de un LMS.

Anexo D



Anexo D: Distribución de uso de los distintos cursos en la plataforma U-Cursos, según desempeño.

Anexo E



Anexo E: Distribución de uso de los distintos cursos en la plataforma Sakai, según desempeño.