



**Universidad de Sevilla**

**Departamento de Estadística e Investigación Operativa**

**TESIS DOCTORAL**

---

**CONSTRUCCIÓN DE UN MODELO PARA DETERMINAR EL  
RENDIMIENTO ACADÉMICO DE LOS ESTUDIANTES  
BASADO EN LEARNING ANALYTICS (ANÁLISIS DEL  
APRENDIZAJE), MEDIANTE EL USO DE TÉCNICAS  
MULTIVARIANTES**

**Autora:** Daysi Karina García Tinisaray

**Directores:** José Luis Pino Mejías y Juan Manuel Muñoz Pichardo



José Luis Pino Mejías, Profesor Titular del Departamento de Estadística e Investigación Operativa, de la Universidad de Sevilla

Juan Manuel Muñoz Pichardo, Titular del Departamento de Estadística e Investigación Operativa, de la Universidad de Sevilla

Certificamos que la presente memoria titulada:

**CONSTRUCCIÓN DE UN MODELO PARA DETERMINAR EL RENDIMIENTO ACADÉMICO DE  
LOS ESTUDIANTES BASADO EN LEARNING ANALYTICS (ANÁLISIS DEL APRENDIZAJE),  
MEDIANTE EL USO DE TÉCNICAS MULTIVARIANTES**

Ha sido realizada bajo nuestra dirección por

**Daysi Karina García Tinisaray**

Y constituye su tesis para optar al grado de Doctora

Y para que conste, en cumplimiento de la legalidad vigente y a los efectos que haya lugar, firmamos el presente documento

En Sevilla, 28 de septiembre de 2015

Fdo. Dr. D. José Luis Pino Mejías

Fdo. Dr. D. Juan Manuel Muñoz Pichardo



## DEDICATORIA

*A mi vida el Señor, a mi amor mi esposo, a mi  
alegría mi hija y a mi fuerza mi familia.*



## **AGRADECIMIENTO**

Mis más sinceros agradecimientos a los directores de esta tesis, José Luis Pino Mejías y Juan Manuel Muñoz Pichardo, del Departamento de Estadística e Investigación Operativa de la Universidad de Sevilla, quienes con sus conocimientos y paciencia me guiaron en el desarrollo de la tesis.

A la Universidad Técnica Particular de Loja y muy especialmente al Padre Luis Miguel Romero quien cuando fue Rector puso en mí su confianza, para alcanzar aquel sueño de la universidad del futuro donde los docentes somos uno de los principales motores.

A la SENECYT (Secretaría de Educación Superior, Ciencia y Tecnología) por concederme una beca de apoyo económico, permitiéndome de esta forma ser partícipe de esta nueva revolución educativa en mi país Ecuador.

Finalmente a toda aquella lista interminable de familiares, amigos, colegas que directa o indirectamente me han apoyado para cumplir con este reto, personas que aportaron con datos, ideas, un aliento, infinitas gracias de corazón a todos ellos.





# ÍNDICE

---

ÍNDICE .....	IX
ÍNDICE DE TABLAS .....	XIII
ÍNDICE DE ILUSTRACIONES .....	XV
INTRODUCCIÓN .....	1
<b>CAPÍTULO 1</b>	
<b>ANÁLISIS DE DATOS EN EL CAMPO EDUCATIVO .....</b>	<b>7</b>
1.1 Enfoques de análisis de datos .....	8
1.2 Minería de datos para la educación .....	9
1.2.1 Definición y usos de la EDM .....	10
1.2.2 La estadística aplicada en el DM .....	12
1.3 Análisis académico .....	13
1.3.1 Definición y logros del AA .....	14
1.3.2 Proceso del análisis académico .....	15
1.4 Análisis del aprendizaje .....	16
1.4.1 Definición de learning analytics .....	17
1.4.2 Proceso del learning analytics.....	18
1.4.3 Usos potenciales y aplicaciones del LA .....	19
<b>CAPÍTULO 2</b>	
<b>RENDIMIENTO ACADÉMICO .....</b>	<b>35</b>
2.1 Conceptualización.....	35
2.2 Indicadores de medida del rendimiento académico.....	37
2.3 Variables que predicen el rendimiento académico universitario .....	39

2.4 Técnicas estadísticas aplicadas en la medición del rendimiento académico: Evidencia empírica .....	42
---	----

### **CAPÍTULO 3**

<b>MODELOS MULTIVARIANTES COMO METODOLOGÍA PARA LA INVESTIGACIÓN EDUCATIVA.....</b>	<b>49</b>
---	-----------

3.1 Análisis a través de una regresión multinivel.....	49
3.1.1 Definición formal de los modelos multinivel.....	50
3.1.2 Características de los modelos multinivel.....	51
3.1.3 Métodos de estimación de los modelos multinivel.....	52
3.1.5 Proceso de modelización de 2 y 3 niveles.....	56
3.1.6 Interpretación de resultados.....	65
3.1.7 Modelos multinivel en el campo educativo .....	66
3.2 Análisis con un modelo logístico bivalente.....	68
3.2.1 Tabla de asociación .....	68
3.2.2 Odds Ratio.....	69
3.2.3 Modelización de la regresión logística bivalente .....	70
3.2.4 Modelización de la regresión logística bivalente ordinal.....	72
3.2.5 Modelos bivariantes aplicados a la medición del rendimiento académico .....	74

### **CAPÍTULO 4**

<b>METODOLOGÍA DEL MODELO DE ESTUDIO.....</b>	<b>75</b>
---	-----------

4.1 Variables objetivo consideradas.....	76
4.1.1 Justificación de las variables.....	76
4.2 Población objeto de estudio .....	84
4.3 Análisis estadístico descriptivo.....	85
4.4 Fuentes de datos .....	93
4.5 Técnicas multivariantes aplicadas.....	94

## **CAPÍTULO 5**

<b>ANÁLISIS DE RESULTADOS DEL MODELO MULTINIVEL .....</b>	<b>99</b>
5.1 Ajuste del modelo multinivel con 2 niveles .....	100
5.2 Ajuste del modelo multinivel con 3 niveles .....	119

## **CAPÍTULO 6**

<b>ANÁLISIS DE RESULTADOS DEL MODELO LOGÍSTICO BIVARIANTE .....</b>	<b>123</b>
6.1 Objetivos de la regresión logística bivalente .....	125
6.2 Estudio de la relación entre las variables de respuesta.....	127
6.3 Análisis logístico bivalente con dos variables de respuesta binarias .....	130
6.4 Análisis logístico bivalente con dos variables de respuesta ordinales .....	140

## **CAPÍTULO 7**

<b>DISCUSIÓN Y FUTURAS INVESTIGACIONES .....</b>	<b>145</b>
7.1 Del análisis multinivel.....	146
7.2 Del análisis logístico bivalente .....	149
7.3 Futuras investigaciones .....	150
<b>REFERENCIAS BIBLIOGRÁFICAS.....</b>	<b>153</b>
<b>ANEXOS .....</b>	<b>165</b>



# ÍNDICE DE TABLAS

---

Tabla 1. Usos de la minería de datos en el ámbito educativo .....	11
Tabla 2. Procesos estadísticos usados en la minería de datos .....	12
Tabla 3. Aplicaciones para learning analytics .....	31
Tabla 4. Metodología estadística aplicada para la medición del rendimiento académico ..	44
Tabla 5. Ecuaciones básicas para modelos multinivel de 2 niveles.....	54
Tabla 6. Ecuaciones básicas para modelos multinivel de 3 niveles.....	55
Tabla 7. Asociación de 2 variables de respuesta .....	68
Tabla 8. Variables como indicadores del rendimiento académico .....	77
Tabla 9. Calificación promedio expresada en escala.....	78
Tabla 10. Agrupación por créditos universitarios acumulados.....	79
Tabla 11. Variables explicativas.....	80
Tabla 12. Análisis descriptivo de las variables utilizadas en el estudio de la influencia de variables individuales, grupales y contextuales sobre el rendimiento académico.....	87
Tabla 13. Estadística descriptiva de las variables de la segunda muestra .....	89
Tabla 14. Resultados del análisis discriminante.....	92
Tabla 15. Regresión modelo nulo 2 niveles .....	102
Tabla 16. Estimación del intercepto con variables del nivel 2.....	105
Tabla 17. Estimación multinivel con variables del nivel 1 .....	109
Tabla 18. Estimación de la regresión con interacciones y variables de los niveles 1 y 2 .....	113
Tabla 19. Estimación de la variabilidad en los coeficientes de los predictores del nivel 1 .....	117
Tabla 20. Comprobación de ajuste de 2 niveles.....	118
Tabla 21. Estimación ANOVA con efectos aleatorios de 3 niveles.....	120
Tabla 22. Asociación entre Grados y Créditos, rendimiento académico óptimo .....	128
Tabla 23. Asociación entre Grados y Créditos, rendimiento académico positivo.....	128
Tabla 24. Asociación entre Grados y Créditos como variables de respuesta ordinales..	130
Tabla 25. Estimación logística bivalente con dos variables de respuesta binarias .....	133

Tabla 26. Probabilidades conjuntas -rendimiento académico óptimo .....	135
Tabla 27. Valores ajustados - rendimiento académico positivo .....	136
Tabla 28. Estimación logística bivariante con dos variables de respuesta binarias .....	138
Tabla 29. Probabilidades conjuntas -rendimiento académico positivo.....	139
Tabla 30. Valores ajustados -rendimiento .....	140
Tabla 31. Estimación de la regresión logística bivariante ordinal .....	142
Tabla 32. Log-Odds Ratios globales .....	143
Tabla 33. Asociación Grados y Créditos con log-GORs.....	143

## ÍNDICE DE ILUSTRACIONES

---

Ilustración 1. Beneficiarios del análisis de datos académicos.....	8
Ilustración 2. Logros del análisis académico .....	14
Ilustración 3. Proceso del análisis académico .....	15
Ilustración 4. Proceso del learning analytics.....	18
Ilustración 5. Centroides de la variable de participación en actividades en línea.....	92
Ilustración 6. Técnicas multivariantes aplicadas para la predicción del rendimiento académico .....	95
Ilustración 7. Estructura jerárquica del modelo.....	95
Ilustración 8. Varianza explicada por nivel .....	121
Ilustración 9. Enfoques de la regresión bivariante .....	124
Ilustración 10. Valores ajustados –Casos de éxito .....	136
Ilustración 11. Valores ajustados –Casos sobre la media .....	140





# INTRODUCCIÓN

---

Las técnicas multivariantes se aplican en todos los campos científicos, sobre todo en la investigación educativa se han constituido como técnicas de análisis importantes. En este trabajo se hace una revisión de dos metodologías que se adaptan a las características de los datos en la educación superior y se consideran como las más adecuadas. Un análisis multinivel para hacer mediciones desde un ámbito cuantitativo y un análisis logístico bivariante que permite hacer mediciones desde un ámbito cualitativo, siendo esta última metodología la que amplía la escasa información que existe sobre estudios en el campo educativo que asocian dos variables de respuesta correlacionadas.

Los datos utilizados provienen de una de las universidades ecuatorianas con más número de estudiantes a nivel de educación superior a distancia en Latinoamérica, a partir de esta información se desarrollan los dos análisis antes mencionados cuya variable objetivo es el rendimiento académico.

El primer análisis se basa en una regresión Multinivel de 2 niveles (estudiantes y aulas) y de tres niveles (estudiantes, aulas y escuelas), el objetivo de este primer análisis es explicar la influencia de 27 covariables que comprenden la dimensión sociodemográfica, antecedentes académicos, enfoque learning analytics, todas estas a “nivel individual”; y las características de la asignatura, del docente y de la escuela, a “nivel contextual”, sobre la variable objetivo el rendimiento académico (cuyo indicador es la nota promedio final de una asignatura troncal). La población objeto de estudio está constituida por 23583 estudiantes, 468 aulas y 19 escuelas correspondientes al año 2014, el análisis del rendimiento académico se realiza en el transcurso de un periodo de estudio.

El segundo análisis se basa en una regresión logística bivariante binaria y ordinal, el objetivo es analizar el rendimiento académico universitario a través de dos variables de respuesta asociadas, el grado o calificación académica y los créditos universitarios acumulados con cuatro covariables (edad de ingreso, género, región de procedencia y participación en actividades en línea). La población objeto de estudio está constituida por 410 estudiantes matriculados en una carrera de 5 años equivalente a 282 créditos, cuyo

tiempo de estudio comprende el periodo abril 2009- abril 2014, es decir, se realiza un análisis del rendimiento académico al finalizar el periodo de estudio de una carrera universitaria.

### **Alcance de la investigación**

En materia de rendimiento académico en la educación superior, la mayoría de las investigaciones relevantes presentan un marcado interés en la inclusión de factores personales como se demostrará mas adelante en el análisis de la evidencia empírica existente, y son pocos los estudios que hacen un abordaje multivariante que incluya variables del enfoque “learning analytics”.

Las plataformas de enseñanza virtual tales como WEbCT, Moodle, Blackboard, Claroline, Dokeos y recientemente las plataformas MOOC (Massive Open Online Courses) han permitido disponer en las universidades que las utilizan, la opción de monitorizar en tiempo real la actividad de los estudiantes. La integración de esta información con otras variables, se constituye en el origen del concepto del learning analytics.

Al ser el “learning analytics” un campo emergente, en el caso de la educación ecuatoriana aún no se ha desarrollado la cultura de utilizar y analizar los datos generados por el estudiante en los procesos de formación y aprendizaje en una plataforma virtual para determinar su influencia en el rendimiento académico, la deserción y/o la graduación.

Bajo estas premisas la presente investigación se centra en determinar la incidencia del uso de las plataformas virtuales sobre el rendimiento académico a través de un grupo de variables de interacción (participación en foro, chat, videocolaboración, número de mensajes enviados al profesor, número de comentarios en el curso de la asignatura, N° de accesos al LMS ), haciendo uso de las ventajas que presentan los dos análisis multivariantes seleccionados. Ya que a diferencia de la regresión clásica, los modelos multinivel permiten incluir en una misma ecuación variables independientes de diferentes niveles de agregación y en el caso de los modelos bivariantes dos variables dependientes cualitativas asociadas.

Además se pretende realizar un trabajo que responda a las necesidades de mejorar el rendimiento académico de la educación superior, de tal forma que se convierta en un aporte que sirva de referente para las universidades.

## **Objetivos**

El objetivo central de esta tesis es construir un modelo para determinar el rendimiento académico de los estudiantes basado en learning analytics, mediante el uso de técnicas multivariantes.

Para guiar el desarrollo teórico, empírico y aplicado en la presente investigación se plantean los siguientes objetivos específicos:

- ✓ Analizar el rendimiento académico universitario a través de un modelo con dos variables de respuesta asociadas.
- ✓ Analizar el rendimiento académico universitario a nivel individual y contextual
- ✓ Determinar el aporte de las variables individuales, grupales o contextuales sobre la conducta del rendimiento académico universitario.
- ✓ Identificar la naturaleza de la relación entre las variables del enfoque learning analytics y el rendimiento académico de los estudiantes.

## **Preguntas de investigación**

1. ¿Qué proporción de la variación en el rendimiento académico puede atribuirse a las variables que engloba el learning analytics?

2. ¿Cuál es la influencia que existe entre variables sociodemográficas y académicas sobre el rendimiento académico del colectivo de estudiantes universitarios ecuatorianos de modalidad a distancia?
3. ¿Existe una relación entre el rendimiento académico y el contexto de los estudiantes y aulas así como entre éstas dos a través del contexto de las escuelas?

## **Hipótesis**

*Hipótesis relacionadas con la pregunta de investigación 1:*

Las variables que engloba el learning analytics influyen positivamente en el rendimiento académico.

*Hipótesis relacionadas con la pregunta de investigación 2:*

La influencia de las variables sociodemográficas y académicas, sobre el rendimiento académico es significativa.

*Hipótesis relacionadas con la pregunta de investigación 3:*

Existe una relación significativa entre el rendimiento académico y el contexto (estudiantes, aulas y escuelas).

## **Estructura del documento**

La presente investigación se desarrolla en siete capítulos. En el capítulo 1 se revisan los tipos de análisis de datos que se están aplicando actualmente en el ámbito educativo, como son: Minería de datos ( Data Mining ), análisis académico (academic analytics) y análisis del aprendizaje (learning analytics), ampliando la revisión de este último ya que es parte del centro de la investigación.

En el capítulo 2 se hace una revisión teórica sobre el rendimiento académico un aspecto importante a la hora de analizar los resultados sobre el proceso de enseñanza aprendizaje y también se incluye un apartado empírico con la finalidad de conocer que modelos estadísticos se han venido aplicando a la hora de medir el rendimiento académico.

El capítulo 3 recoge una revisión teórica, matemática y empírica de las dos técnicas multivariantes aplicadas en la investigación educativa.

En el capítulo 4 se hace referencia a la metodología de estudio, selección de datos, establecimiento de variables que permiten justificar porque se trabaja con un modelo multinivel y logístico bivariante.

En el capítulo 5 se obtienen los resultados del modelo empírico multinivel al estimar el modelo jerárquico con 2 y 3 niveles: estudiante (nivel inferior), aula (nivel intermedio) escuela (nivel superior). Los modelos se ejecutan usando el software Stata/SE 12.0.

En el capítulo 6 se presentan los resultados del modelo logístico bivariante binario y ordinal. Los modelos se ejecutan usando la herramienta Rstudio.

Finalmente, en el capítulo 7 se discuten los resultados y se plantean futuras investigaciones.



# **CAPÍTULO 1**

## **ANÁLISIS DE DATOS EN EL CAMPO EDUCATIVO**

---

El análisis de datos está en auge en el área de la educación sobre todo porque existen herramientas para procesar un volumen creciente de datos, facilitando de esta forma el uso de la información relacionada con el estudiante, el docente, la entidad educativa, etc. con fines de mejorar el aprendizaje.

La tendencia en las universidades es trabajar con tecnologías que faciliten y mejoren el desarrollo de las actividades académicas, creándose de esta forma un aprendizaje electrónico añadido al tradicional. El uso de estos nuevos medios tiene grandes ventajas. Una de ellas es que permite generar una gran cantidad de datos producidas por el estudiante durante el proceso de aprendizaje. Esta información procesada con las herramientas adecuadas permite predecir, prevenir y/o actuar para mejorar el rendimiento académico de los estudiantes.

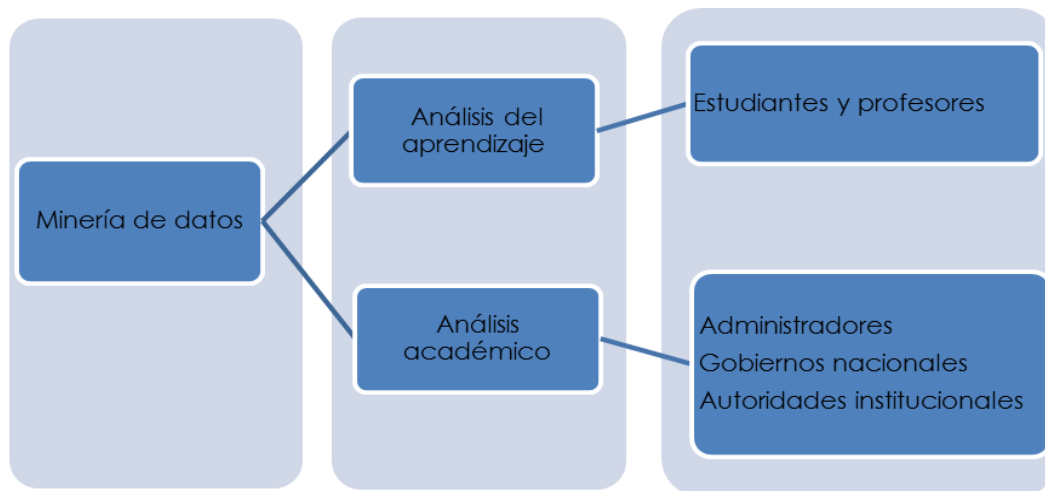
Los principales beneficiarios de este análisis de datos son los estudiantes seguidos de los docentes, administradores, autoridades de la institución académica y gobierno nacional.

## 1.1 Enfoques de análisis de datos

En la actualidad, los enfoques de análisis de datos más usados en el ámbito de la educación superior son la minería de datos educativos (del inglés, Educational Data Mining, EDM), el análisis académico (del inglés, Academic Analytics, AA) y el análisis del aprendizaje (del inglés, Learning Analytics, LA).

El análisis del aprendizaje, análisis académico y minería de datos se centran específicamente en herramientas y métodos para la exploración de datos que provienen de contextos educativos (Dyckhoff y otros 2012). Hoy en día se considera que estas técnicas ayudan a moldear el futuro de la educación superior y a generar nuevos enfoques y estrategias en mejora de la enseñanza y del aprendizaje.

En general, como se observa en la Ilustración 1 tanto el análisis del aprendizaje como el análisis académico utilizan la información que la minería de datos obtiene a través de la búsqueda de patrones o aplicación de algoritmos de procesamiento de datos.



**Ilustración 1. Beneficiarios del análisis de datos académicos**

**Fuente:** Elaboración propia



Para establecer una diferencia entre éstos tres enfoques se puede mencionar los planteamientos que establece Ferguson (2012):

- La minería de datos es un desafío técnico ¿Cómo se puede extraer valor de los grandes conjuntos de datos relacionados con el aprendizaje?
- El análisis del aprendizaje es un desafío educativo ¿Cómo se puede optimizar las oportunidades para el aprendizaje en línea?
- El análisis académico es un desafío económico / político ¿Cómo se puede mejorar sustancialmente las oportunidades de aprendizaje y los resultados educativos a nivel nacional o internacional?

Estos enfoques no solo recogen y exploran grandes cantidades de información, sino que permiten construir y poner a prueba modelos que se centran en el estudiante, ya sea de forma individual o en el contexto de la institución, con la finalidad de predecir o mejorar el rendimiento académico.

## **1.2 Minería de datos para la educación**

La DM surge en los años 90 con una finalidad empresarial que consistía en utilizar toda la información que se podía obtener para conocer el comportamiento de los consumidores o clientes. La minería está basada en técnicas estadísticas que generan algunos aspectos positivos:

- Deduce patrones y tendencias que existen en los datos.
- Encuentra relaciones entre variables.
- Pronostica conductas individuales.
- Genera modelos descriptivos (regresión simple, múltiple, logística, análisis de series de tiempo, etc.) y modelos de predicción (clúster, reducción de dimensiones, reglas de asociación, etc.).

- Descubre patrones ocultos.

### **1.2.1 Definición y usos de la EDM**

La gran cantidad de información generada hoy en día por los estudiantes permite que la EDM obtenga datos relevantes y, a través de métodos estadísticos y otras herramientas, relacione la información para conocer si el proceso de enseñanza aprendizaje ha dado resultados positivos.

Es así que Bienkowski y otros (2012) definen “Educational Data Mining” como la teoría que desarrolla métodos, aplica técnicas estadísticas y de aprendizaje automático para analizar los datos recogidos durante el proceso de la enseñanza y aprendizaje.

Actualmente los usos más generales que se le están dando a la EDM básicamente se enfocan en mejorar la estructura del conocimiento y determinar el apoyo pedagógico. En la Tabla 1 se muestra este enunciado.

Según Luan (2002) la minería de datos utiliza una combinación de bases de conocimientos explícita, conocimientos analíticos complejos y conocimiento de campo para descubrir las tendencias y los patrones ocultos. Estas tendencias y patrones forman la base de los modelos predictivos que permiten a los analistas realizar nuevas observaciones de los datos existentes.

Con respecto a los trabajos que se han desarrollado en el campo educativo, existen diversos estudios que han implementado el uso de la minería de datos como son los que se citan a continuación:

Ayesha y otros (2010) realizan un análisis de conglomerados a través del método de las k-medias (en inglés, k-means clustering) para determinar el comportamiento de aprendizaje del estudiante. Consideran los factores que involucran la evaluación del

estudiante y concluyen que si se trasmite con anticipación los resultados se reducirá la deserción y se mejorará el rendimiento académico de los estudiantes.

**Tabla 1. Usos de la minería de datos en el ámbito educativo**

Actores	Usos
Estudiantes	<ul style="list-style-type: none"> <li>- Personalizar las actividades de aprendizaje de acuerdo al perfil del alumno.</li> <li>- Determinar qué actividades, recursos, tareas, etc. mejoran el rendimiento académico.</li> </ul>
Docentes	<ul style="list-style-type: none"> <li>- Buscar patrones de comportamiento.</li> <li>- Cuantificar la efectividad del proceso de enseñanza-aprendizaje.</li> <li>- Auditar actividades.</li> <li>- Mejorar o corregir la estructura de un curso.</li> <li>- Clasificar o agrupar alumnos en base a sus características.</li> <li>- Personalizar y adaptar el contenido de cursos.</li> </ul>
Institución educativa	<ul style="list-style-type: none"> <li>- Organizar los recursos institucionales.</li> <li>- Optimizar recursos.</li> <li>- Mejorar diseño de horarios.</li> <li>- Adquirir material adecuado.</li> <li>- Mejorar la oferta educativa.</li> <li>- Mejorar de la eficiencia de los portales.</li> </ul>

**Fuente:** Elaboración propia

En el trabajo de Zhan y otros (2010) se analiza el comportamiento académico de los estudiantes y se proporciona una base para las estrategias de intervención y disminución de la deserción estudiantil.

Dickerson & Hazelton (2012) exploran y ponen en práctica las técnicas de minería de datos para crear un nuevo circuito de retroalimentación hacia los profesores para la evaluación de programas y la adaptación sistemática a los cambios que requiere el sector empresarial de sus estudiantes.

Guruler & Istanbulu (2014) evalúan y desarrollan un enfoque basado en datos relacionados con la mejora del rendimiento de los estudiantes universitarios mediante la

aplicación de técnicas incluidas en el ámbito del "Descubrimiento de Conocimiento en Bases de Datos" (conocido como KDD, las iniciales de Knowledge Discovery in Databases), combinado con minería de datos.

Khan & Choi (2014) a través de un árbol de decisión y algoritmos procesan los datos de los estudiantes para calcular las posibilidades de ganar una beca en función de su grado de semestre, la ubicación del alumno en clase, la cantidad máxima y mínima de horas de crédito tomadas y permitidas y las actividades extracurriculares.

### 1.2.2 La estadística aplicada en el DM

En la Tabla 2 se resume las técnicas que aplica "Educational Data Mining" para la extracción de la información y los procesos estadísticos que se usan para facilitar el análisis de los datos, sin embargo, estos términos pueden ir en aumento debido a que el abanico de técnicas y procedimientos englobados en la Minería de Datos está creciendo exponencialmente.

**Tabla 2. Procesos estadísticos usados en la minería de datos**

Técnicas de minería de datos	Procesos estadísticos
Predicción por redes neuronales	Ecuaciones de regresión, chi-cuadrado, ecuaciones estructurales.
Regla inductiva	Análisis de componentes principales, análisis discriminante, análisis factorial.
Mapas de Kohonen	Análisis cluster, función de densidad.
Visualización espacial	Estadística descriptiva: gráficos de dos y tres dimensiones.
Espacio euclídeo	Modelo de ecuaciones estructurales y regresiones lineales y no lineales.
Clasificación	Regresión logística.
Segmentación	Análisis cluster y análisis factorial.
Detección de outliers	Desviación estándar y análisis de errores.

**Fuente:** Elaboración propia, en base a Luan (2002).

### 1.3 Análisis académico

El AA aparece en el ámbito de la enseñanza desde el 2005, pero previamente fue acuñada por la compañía WebCT (ahora Blackboard) para describir las funciones de recolección de datos que el CMS (Content Management System) tiene habilitadas (Baepler & James, 2010).

Generalmente se lo conoce como “Business intelligence” aplicado a la educación superior. El análisis académico permite a los administradores de centros educativos monitorear el progreso de los objetivos institucionales, eficiencia terminal de las carreras, impacto de la difusión, etc.

Hoy en día las instituciones educativas tienen que rendir cuentas no solamente ante entes internos sino también ante organismos estatales que están midiendo y evaluando la eficiencia de la educación, por lo que están aplicando este tipo de análisis que les ha permitido cambiar su forma de tomar decisiones, modificar en algunos casos la planificación, elaboración de informes y la asignación de recursos.

Algunos trabajos interesantes que se han realizado aplicando “academic analytics” son: Morris y otros (2005) a través de técnicas multivariantes y análisis discriminante predictivo determinan la relación entre la deserción y el promedio de calificaciones.

Sharkey (2011) crea un repositorio de datos integrados con todas las partes de la institución para predecir la persistencia de los estudiantes a través de indicadores tales como: horarios, calificaciones, uso del contenido de clases y datos demográficos.

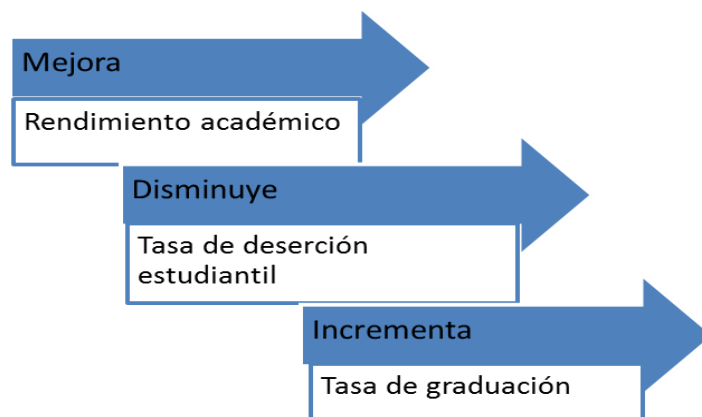
Olmos & Corrin (2012) determinan que el uso de análisis académico es valioso para el desarrollo curricular y que además permite tener fundamentos para realizar críticas en mejora de la excelencia educativa.

Palmer (2013) demuestra en su investigación que los datos del estudiante almacenados en los sistemas institucionales se pueden utilizar para predecir el rendimiento académico. Además identifica variables predictoras significativas que ofrecen la capacidad de desarrollar intervenciones específicas para mejorar el rendimiento académico de los estudiantes y disminuir las tasas de deserción.

Lauría y otros (2013) aplican modelos predictivos probando su factibilidad y demostrando que los resultados obtenidos permiten generar estrategias de detección e intervención sobre los estudiantes en riesgo académico.

### 1.3.1 Definición y logros del AA

El análisis académico fortalece la responsabilidad y el cumplimiento de los objetivos de las instituciones de educación superior, permitiéndoles trabajar eficientemente con los recursos, cuya finalidad es lograr los aspectos presentados en la Ilustración 3.



**Ilustración 2. Logros del análisis académico**

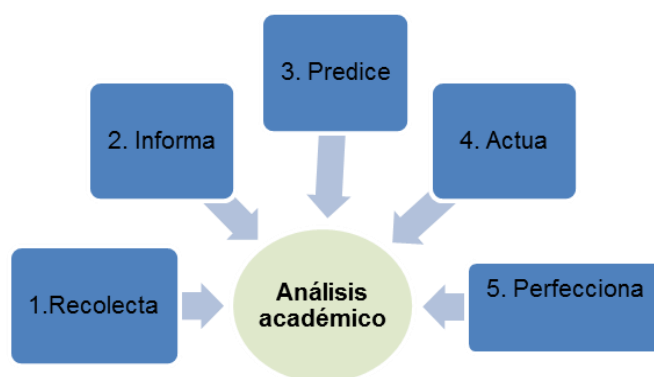
**Fuente:** Elaboración propia

Baepler & James (2010) señalan que el análisis académico combina datos institucionales, análisis estadístico y modelos de predicción con la finalidad de crear conciencia sobre los

estudiantes, instructores o administradores para que puedan cambiar el comportamiento académico.

Goldstein & Katz (2005) usan el término “academic analytic” como el equivalente de “business intelligence”, para describir la intersección entre la tecnología, la información, la cultura de gestión y la aplicación de información para gestionar la empresa académica.

### 1.3.2 Proceso del análisis académico



**Ilustración 3. Proceso del análisis académico**

**Fuente:** Elaboración propia

En la Ilustración 4 se observa que el proceso del análisis académico consta de cinco pasos. Se inicia con la *recolección* de los datos. Estos pueden ser internos o externos a la institución educativa, por ejemplo dentro de los datos típicos que se levantan se encuentran los demográficos, habilidades académicas, rendimiento académico, historial académico, etc.

El segundo paso implica la presentación del *informe*, aquí se examinan los datos y se identifican tendencias, posteriormente se continúa con la *predicción* que mediante modelos estadísticos se determina cual será el rendimiento del estudiante. En base a las predicciones y probabilidades, las instituciones deben *actuar* planteando estrategias de

mejora, finalmente la institución tomando en cuenta los resultados que ha obtenido del análisis académico debe *perfeccionar* las actividades que permitan a la institución ser más eficiente.

## 1.4 ANÁLISIS DEL APRENDIZAJE

El “learning analytics” surge a partir de dos tendencias convergentes: el uso cada vez mayor de los Entornos Virtuales de Aprendizaje en las instituciones educativas y la aplicación de técnicas de minería de datos para los procesos de inteligencia de negocios en sistemas de información de la organización (Agudo y otros 2012).

El informe Horizont (Johnson y otros 2013) menciona que el learning analytics tiene su origen en la minería de datos aplicada al sector comercial en donde se realizaban análisis de las actividades de los consumidores con la finalidad de personalizar la publicidad.

En el trabajo de Ferguson (2012) se delinearán tres grandes factores que han impulsado el desarrollo del análisis del aprendizaje y que benefician a los gobiernos, instituciones educativas, docentes y estudiantes:

- 1) La existencia de grandes conjuntos de datos que se obtienen de los entornos virtuales de aprendizaje (VLE) también conocidos como sistemas de gestión de aprendizaje (LMS).
- 2) Búsqueda de la optimización del aprendizaje “on line”.
- 3) La preocupación de los gobiernos e instituciones internacionales que buscan medir y mejorar el rendimiento estudiantil.



### **1.4.1 Definición de learning analytics**

El análisis del aprendizaje aprovecha la gran cantidad de información que generan los estudiantes durante el proceso educativo con la finalidad de mejorar la calidad del aprendizaje y el rendimiento académico.

Según Brown (2011) este tipo de análisis permite usar los datos asociados con el aprendizaje de los estudiantes y generar informes que sean útiles para los docentes (actividades y progreso de los estudiantes), para los estudiantes (retroalimentación) y para los administradores (incremento de aulas de clase, tasa de graduación, etc.).

Learning analytics es la medición, recopilación, análisis y presentación de datos sobre los alumnos y sus contextos, a efectos de entender y optimizar el aprendizaje y los entornos en los que ocurren los sucesos de aprendizaje (Siemens, 2011).

Por otra parte Johnson y otros (2011) indican que el learning analytics busca aprovechar los avances en la minería de datos, la interpretación y modelación, para mejorar la comprensión de la enseñanza y el aprendizaje para adaptar eficazmente la educación a los estudiantes.

Estas definiciones nos permiten corroborar dos afirmaciones planteadas por Ferguson (2011).

- El learning analytics utiliza datos pre-existentes.
- El learning analytics puede ser usado para manejar grandes conjuntos de datos, que no son posibles de manejar de forma manual.

## 1.4.2 Proceso del learning analytics

La inclusión de tecnología en los procesos de enseñanza aprendizaje tiene como ventaja el contar con el registro de actividades que tanto el docente como el estudiante efectúan en un entorno virtual de aprendizaje mientras realizan su actividad académica.

El hecho de contar con una gran cantidad de datos generados en estas plataformas permite utilizar esta información para mejorar y personalizar el aprendizaje. Estos datos pasan por una serie de procesos, que le permiten al docente y al estudiante determinar si se han alcanzado sus objetivos, parte de este procesamiento de datos se lo realiza con “data mining”.

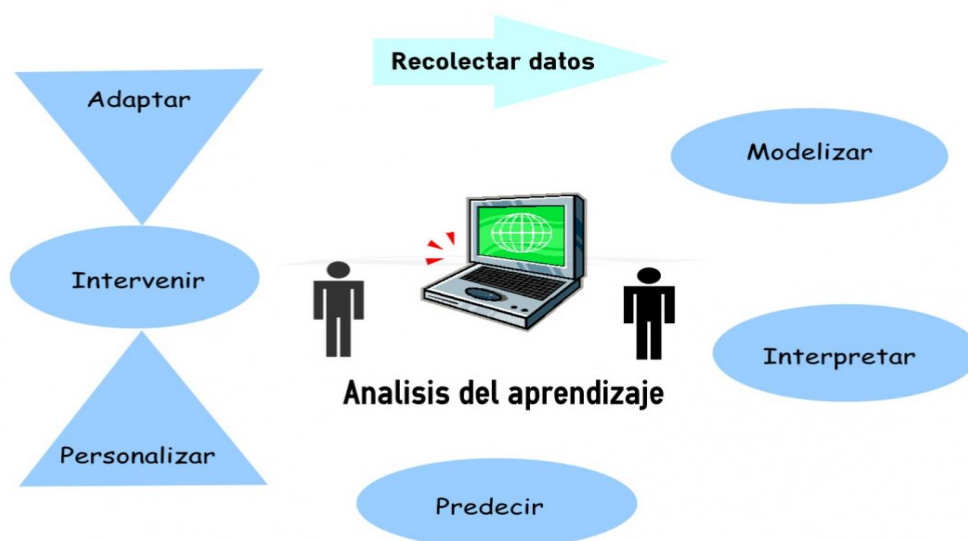


Ilustración 4. Proceso del learning analytics

Fuente: Elaboración propia

El proceso que se observa en la Ilustración 5 inicia con la **recolección de datos** proporcionados por la actividad académica del estudiante en un entorno virtual de aprendizaje. Posteriormente se continúa con el tratamiento de los mismos para obtener las variables deseadas y plantear un **modelo** utilizando algunas técnicas de minería que permitan determinar la relación entre los instrumentos de enseñanza y el aprendizaje.

Luego la institución debe realizar la **interpretación** de los resultados y reflexionar sobre el impacto del método de enseñanza en el comportamiento de aprendizaje y el rendimiento de los estudiantes, permitiendo de esta forma hacer una **predicción** de las tendencias y del comportamiento de los estudiantes, las mismas que al ser negativas provocan una **intervención** pedagógica que consiste en trabajar adecuadamente sobre las actividades específicas que no están funcionando, a través de la **adaptación** del estudiante a la metodología o viceversa. La **personalización** de los entornos de enseñanza-aprendizaje en función de las necesidades de cada estudiante es otro tipo de intervención pedagógica que puede ayudar a mejorar el proceso de enseñanza aprendizaje.

### **1.4.3 Usos potenciales y aplicaciones del LA**

Varias instituciones educativas están utilizando sus bases de datos generadas en las herramientas de educación virtual para realizar el learning analytics enfocado en los siguientes aspectos:

- Evaluación diagnóstica y formativa
- Evaluación del currículo
- Evaluación de programas académicos
- Mejora del rendimiento estudiantil
- Adaptabilidad de recursos de aprendizaje
- Planificación personalizada
- Evaluación de las tutorías
- Desarrollo de nuevos procesos y herramientas destinadas a mejorar el aprendizaje y la enseñanza
- Identificación de docentes con buen desempeño y docentes que requieren capacitación para mejorar los métodos de enseñanza
- Establecimiento de modelos asociados al aprendizaje por competencias
- Identificación de puntos conflictivos de un proceso de enseñanza-aprendizaje

Para cumplir con estos usos potenciales, actualmente se han desarrollado varias aplicaciones como las expuestas en la Tabla 3, cuya finalidad primordial es mejorar el rendimiento académico.

**Tabla 3. Aplicaciones para learning analytics**

Aplicaciones	¿Qué hace?	Objetivo	Enlace
Signals	Proporciona un indicador de riesgo a través de un semáforo para cada estudiante mediante el uso de datos en tiempo real (verde da la señal de que todo está bien, amarillo sugiere algunas áreas de preocupación y rojo indica problemas significativos).	Mejorar el rendimiento académico del estudiante en el aula detectando tempranamente características de advertencia que eviten que el estudiante se ubique en un estado crítico.	<a href="http://www.itap.purdue.edu/studio/signals/">http://www.itap.purdue.edu/studio/signals/</a> Purdue University- Estados Unidos.
SNAPP (Social Networks Adapting Pedagogical Practice)	Realiza representaciones visuales de la interacción que se genera en los mensajes y respuestas en los foros de discusión, los datos los obtiene de los sistemas de aprendizaje, como Blackboard, Moodle y Sakai.	Ser un instrumento de diagnóstico evaluando los patrones de comportamiento e identificando los estudiantes en riesgo de acuerdo a los niveles de participación.	<a href="http://www.snappvis.org/">http://www.snappvis.org/</a> University of Wollongong – Australia.
Socrato	Facilita pruebas de admisión estandarizadas para que el docente estudiante o escuela interesada las desarrolle y posteriormente genera informes de diagnóstico y rendimiento.	Identificar fortalezas y debilidades y crear planes de aprendizaje personalizados para mejorar las puntuaciones.	<a href="http://www.socrato.com/">http://www.socrato.com/</a> SOCRATO- Boston, Massachusetts.

Aplicaciones	¿Qué hace?	Objetivo	Enlace
LOCO-Analyst (Learning Object Context Ontologies)	Realiza el seguimiento del estudiante analizando los datos que genera su interacción en línea con los contenidos de aprendizaje (cuestionarios, lecturas, foros, etc.).	Proporcionar a los docentes una retroalimentación de la participación de los alumnos en el proceso de aprendizaje que se ha generado en la web.	<a href="http://jelenajovanovic.net/LOCO-Analyst/">http://jelenajovanovic.net/LOCO-Analyst/</a> Apoyado por: Canadian International Development Agency (CIDA). Canada's NSERC-funded LORNET research network. WINTEL project funded by the Serbian Ministry of Science.
GISMO(Graphical Interactive Student Monitoring Tool for Moodle)	Genera representaciones gráficas que pueden ser exploradas y manipuladas por los docentes del curso para examinar los aspectos sociales, cognitivos y de comportamiento de los estudiantes.	Proporcionar una visualización útil de las actividades de los estudiantes en los cursos en línea para docentes.	<a href="http://gismo.sourceforge.net/">http://gismo.sourceforge.net/</a> eLearning laboratory (eLab) of the University of Lugano in Switzerland.
CMA (Check My Activity)	Los estudiantes pueden revisar su actividad y la frecuencia de su uso en relación a un resumen anónimo de los compañeros de curso.	Lograr que en base a las comparaciones por pares se pueda influir en los estudiantes en la metacognición.	<a href="http://www.umbc.edu/oit/newmedia/blackboard/stats/">http://www.umbc.edu/oit/newmedia/blackboard/stats/</a> University of Maryland, Baltimore County (UMBC).
Junyo	Desarrolla plataformas educativas para analizar los datos que generan los estudiantes de primaria en tiempo real.	Realizar una retroalimentación en tiempo real y mejorar el proceso de aprendizaje.	<a href="http://junyo.com/">http://junyo.com/</a> Silicon Valley's Menlo Park, California.

Aplicaciones	¿Qué hace?	Objetivo	Enlace
Persistence Plus	A través de una plataforma móvil, realiza intervenciones conductuales permitiendo a los estudiantes organizar su tiempo y responsabilidades.	Incrementar tasas de retención y graduación.	<a href="http://www.persistenceplusnetwork.com/">http://www.persistenceplusnetwork.com/</a>  Persistence Plus™ - Boston, Massachusetts.
eLAT (exploratory Learning Analytics Toolkit)	Permite explorar y correlacionar el uso del contenido, las propiedades y el comportamiento del usuario, así como los resultados de evaluación basada en indicadores gráficos.	Apoyar la reflexión y el mejoramiento de los métodos de enseñanza en línea sobre la base de los intereses personales y observaciones.	<a href="http://www.elearn.rwth-aachen.de/Dyckhoff">http://www.elearn.rwth-aachen.de/Dyckhoff</a>  RWTH Aachen University, Germany
Moodog (Watchdog Moodle)	Muestra como los estudiantes interactúan con los materiales de un curso en línea, proporciona recordatorios automáticos, alentándolos a ver los materiales disponibles aún no revisados.	Permitir que los estudiantes comparen su progreso propio con los demás en la clase.	<a href="http://www.nmsl.cs.ucsb.edu/papers/152.pdf">http://www.nmsl.cs.ucsb.edu/papers/152.pdf</a>  University of California in Santa Barbara.
D2L (Student Success System)	Utiliza conocimientos analíticos para hacer predicciones sobre el rendimiento del estudiante y los niveles de riesgo en sus cursos.	Monitorear el compromiso del estudiante y las expectativas de logro para sus cursos, controlar y diseñar intervenciones específicas para los estudiantes en riesgo mejorar el rendimiento académico	<a href="https://documentation.desire2learn.com/en/student-success-system">https://documentation.desire2learn.com/en/student-success-system</a>  D2L Corporation - Kitchener Ontario, Canadá.

Aplicaciones	¿Qué hace?	Objetivo	Enlace
jPoll	Realizar un sondeo móvil que involucra a los estudiantes a través de iPhone, Android, Blackberry, tabletas, netbooks y ordenadores portátiles. Los docentes crean listas de preguntas para sondear a los estudiantes.	Proporcionar retroalimentación directa e inmediata a los instructores y proporcionar a los estudiantes información instantánea sobre la comprensión de los contenidos para ayudarles a controlar su propio entendimiento.	<a href="http://app.griffith.edu.au/poll/landing.html">http://app.griffith.edu.au/poll/landing.html</a>  Griffith University - Australia
Sherpa (Predictive analytics module)	Proporciona una alerta temprana de la conducta del estudiante que tiene una mayor probabilidad de fracaso y fomenta un comportamiento productivo.	Mejorar el rendimiento académico a través de un motor de recomendación.	<a href="http://www.socccd.edu/sherpa/">http://www.socccd.edu/sherpa/</a>  South Orange County Community College District- Estados Unidos

Fuente: Elaboración propia



# CAPÍTULO 2

## RENDIMIENTO ACADÉMICO

---

El rendimiento académico es el principal indicador de éxito o fracaso del estudiante, por tal motivo ha sido considerado como uno de los aspectos importantes a la hora de analizar resultados sobre el proceso de enseñanza aprendizaje. En el ámbito de la educación superior el rendimiento académico es uno de los indicadores principales de la eficiencia y calidad educativa que representa a la institución.

Causa controversia ya que no existe una teoría definitiva acerca de una metodología para su medición o un indicador para su valoración. Al ser multidimensional, el rendimiento académico depende de múltiples aspectos tales como los objetivos del docente, de la institución, del estudiante, etc.

En este capítulo se hace una revisión de los indicadores, las variables más usadas para la medición del rendimiento académico y las técnicas estadísticas aplicadas en la evidencia empírica revelante.

### 2.1 Conceptualización

El rendimiento académico se determina tomando en cuenta variables cualitativas y/o cuantitativas que permiten determinar si el estudiante, la institución y el docente han tenido éxito o han sido eficientes durante el desarrollo del proceso de enseñanza aprendizaje. Suele tomar algunas denominaciones como logro académico, rendimiento universitario, desempeño académico, éxito escolar, etc.

Edel Navarro (2003) refiere al rendimiento académico como una intrincada red de articulaciones cognitivas generadas por el hombre que sintetiza las variables de cantidad y cualidad como factores de medición y predicción de la experiencia educativa y que contrariamente de reducirlo como un indicador de desempeño escolar, se considera una constelación dinámica de atributos cuyos rasgos característicos distinguen los resultados de cualquier proceso de enseñanza aprendizaje.

Al ser un tema complejo y un aspecto primordial en la búsqueda de la calidad educativa el rendimiento académico es de gran interés a nivel social, institucional y económico.

Fullana (1992) hace énfasis en el enfoque multidimensional del rendimiento académico y destaca que es un resultado del proceso de aprendizaje académico en el que convergen los efectos de numerosas variables sociales, personales, escolares y de sus interrelaciones.

En un estudio reciente Rojas (2013) menciona algunas definiciones de investigadores que han conceptualizado el rendimiento académico como son: Garbanzo (2007) que recoge que rendimiento académico es la suma de diferentes y complejos factores que actúan en la persona que aprende, y ha sido definido con un valor atribuido al logro del estudiante en las tareas académicas; Tournon (1984) que afirma que el rendimiento académico es un resultado del aprendizaje, suscitado por la intervención pedagógica del profesor o profesora y producido en el alumno; y Forteza (1975) quien definió el rendimiento académico como una serie de factores que giran alrededor de los resultados finales del esfuerzo hecho por el o la estudiante.

En definitiva se pueden presentar infinidad de definiciones, ya que al existir varios actores involucrados también existen varias visiones que implican que al rendimiento académico se lo debe medir o definir basados en la conducta, esfuerzo, estrategias de enseñanza-aprendizaje, aprovechamiento escolar, etc.

## 2.2 Indicadores de medida del rendimiento académico

Siendo el rendimiento académico un indicador multidimensional, la forma de valolarlo debe ser objetiva y fiable. Existen algunos indicadores que se han venido usando en alguna investigaciones, tal vez unos con más fuerza que otros, algunos autores han hecho ya una diferencia del rendimiento académico basado en el indicador de medición.

Latiesa (1992) diferencia el rendimiento académico en sentido amplio considera el éxito, retraso y abandono de estudios; y en sentido estricto las notas obtenidas por los estudiantes. Mientras que, De Miguel y otros (2002) distinguen entre rendimiento inmediato considerando las notas y el mediato en el que incluye logros personales y profesionales.

A continuación se hace una revisión de los indicadores más usados en la medición del rendimiento académico.

### a) **Calificaciones**

Las notas representan una medición cuantitativa o cualitativa del rendimiento, fácil de analizar puesto que una nota alta puede indicar un excelente rendimiento académico y una nota baja todo lo contrario.

En la mayoría de las investigaciones se toma en cuenta la calificación final por ser una medida exacta y accesible. Sin embargo, Page (1990) indica que las calificaciones poseen un valor relativo como medida de rendimiento, ya que no existe un criterio estandarizado para todos los centros, para todas las asignaturas, para todos los cursos y para todos los profesores.

Se podría decir que el mismo hecho de que no haya un criterio estandarizado, en ocasiones puede implicar que no haya la suficiente objetividad, por tal motivo, Escudero

(1999) señala que las calificaciones son una medida de los resultados de la enseñanza, pero no estrictamente de su calidad pues están condicionadas no solo por la calidad de los alumnos, sino también por el criterio y el rigor personal del profesor a la hora de diseñar la enseñanza y valorar y calificar el aprendizaje y el rendimiento académico.

#### **b) Pruebas objetivas**

Como medidas del rendimiento, ofrecen mayores ventajas; entre éstas destaca su grado de objetividad, debido a que las respuestas son cortas y precisas, sin la influencia subjetiva del profesor; estas pruebas poseen un alto grado de validez, debido a que cumplen específicamente los propósitos para los que fueron elaboradas (Page, 1990; González, 1988; Marrero & Espino, 1988).

#### **c) Número de asignaturas aprobadas**

Este indicador permite determinar el rendimiento académico a través de una medida de éxito/fracaso, por ejemplo, si un estudiante ha superado la mayor parte de las asignaturas en las que se matriculó un determinado periodo, indicaría que tuvo éxito; si aprobó menos de la mitad se podría decir que tuvo un bajo desempeño académico y peor aún sino superó ninguna de las materias se concluiría indicando que el estudiante fracasó académicamente.

El estudio realizado por Di Gresia & Porto (2005) concluye que el número de asignaturas aprobadas por año es un indicador de rendimiento estudiantil más adecuado que el promedio, ya que sintéticamente la cantidad de materias aprobadas presenta mayor variabilidad que las notas obtenidas en las materias.

#### **d) Créditos acumulados**

Al igual que el anterior indicador representa el éxito o fracaso de un estudiante, pues indica el avance que ha tenido en la carrera, es decir, permite hacer una comparación

entre los créditos acumulados por el alumno durante cierto tiempo de estudio y los créditos que, de acuerdo al plan de estudios, debió acumular en el tiempo programado (Rodríguez & Ruiz, 2011).

## **2.3 Variables que predicen el rendimiento académico universitario**

El rendimiento académico está relacionado con múltiples variables de diferente tipo, se presentan variables controlables como es el caso de las que se engloban en los aspectos institucionales y pedagógicos, también considera las variables no controlables que comprenden el ámbito sociocultural, intelectual, demográfico, etc.

Algunos trabajos recientes han aplicado métodos estadísticos con un enfoque generalmente basado en factores demográficos, económicos y psicológicos, como son los de Tomás y otros (2014); Kotzé & Niemann (2013), Miguel-Vázquez y otros (2013); Oladejo y otros (2010); entre otros.

Varios factores pueden influir en el rendimiento académico de un estudiante y explicar el éxito o fracaso académico. Las características del estudiante, del lugar donde estudia, del profesor que dicta las clases, de la familia que lo rodea, etc., se convierten en variables que pueden determinar el rendimiento académico tomando en cuenta el contexto, como son:

### **Demográficas**

- Género
- Edad
- Estado Civil
- Experiencia laboral
- Porcentaje de discapacidad
- Tipo de colegio donde culminó la educación secundaria

### **Socioculturales**

- Nivel educativo de los padres
- Número de miembros de la familia
- Ocupación de los padres

### **Socioeconómicas**

- Nivel de ingresos
- Número de autos
- Casa propia o arrendada

### **Institucionales**

- Infraestructura
- Políticas
- Horarios de los cursos
- Número de libros en biblioteca
- Organización y planificación del personal
- Tipo y tamaño de la institución
- Procesos de funcionamiento

### **Pedagógicas**

- Número de alumnos por maestro
- Métodos y materiales didácticos utilizados
- Motivación de los estudiantes
- Tiempo dedicado por los profesores a la preparación de sus clases
- Expectativas y actitudes del profesor
- Formación académica y experiencia del profesor

### **Intelectuales**

- Concentración
- Memoria
- Comprensión verbal

- Razonamiento
- Fluidez verbal

### **Afectivas**

- Sociabilidad
- Seguridad
- Integración
- Relación padre-hijo
- Relación Profesor-alumno

### **Académicas**

- Resultados académicos obtenidos en el colegio
- Nota de la pruebas de admisión a la universidad
- Asistencia y participación en clases

### **Psicológicas**

- Esperanza
- Optimismo
- Resiliencia

### **Tecnológicas**

En el ámbito **tecnológico** también se obtienen variables como determinantes del rendimiento académico, es un factor muy importante puesto que la mayoría de las instituciones educativas han incorporado elementos tecnológicos con la finalidad de estimular y actualizar el proceso de enseñanza (docentes) y aprendizaje (estudiantes). Generalmente las variables que se pueden considerar dentro de este ámbito son las relacionadas con:

- El entorno tecnológico institucional
- Posibilidades de acceso
- Usos de internet

Finalmente es importante señalar que cada una de las variables consideradas en todos los contextos anteriores pueden presentar diferente nivel de influencia o significatividad ya sea positiva o negativa en el rendimiento académico. A la hora de intervenir, para una institución educativa es mucho más fácil que lo haga con las variables controlables como las que se ubican en el contexto institucional, pedagógico y académico.

## **2.4 Técnicas estadísticas aplicadas en la medición del rendimiento académico: Evidencia empírica**

Desde hace varias décadas se ha venido investigando como mejorar el rendimiento académico de los estudiantes, que variables influyen sobre él, cuál es la mejor metodología para determinarlo.

El análisis de esta variable multidimensional es y será relevante para la mejora de la calidad educativa, razón por la cual se han dedicado numerosos estudios, tanto desde el punto de vista teórico como empírico. Basados en esta premisa, en la Tabla 4 se presenta un listado de investigaciones relevantes publicadas en revistas de ámbito estadístico y educativo de alto impacto.



**Tabla 4. Metodología estadística aplicada para la medición del rendimiento académico**

AUTOR	TEMA	METODOLOGÍA	MUESTRA	VARIABLES	CONCLUSIÓN	JOURNAL
Marjoribanks (1976)	Academic achievement, intelligence, and creativity: A regression surface analysis	Estimaciones OLS	400 niños de 12 años de la escuela inglesa	Rendimiento académico, medidas de inteligencia y creatividad.	Para algunas áreas la creatividad deja de estar relacionada con el rendimiento después de haber alcanzado un nivel de umbral de la inteligencia. También se encontró que en los altos niveles de habilidad verbal, la capacidad no verbal y la creatividad parecía tener relaciones diferenciales con el rendimiento académico.	Multivariate Behavioral Research
Schram (1996)	A meta-analysis of gender differences in applied statistics achievement	Modelo lineal general, test de homogeneidad, análisis categórico	Estudiantes de pregrado y postgrado matriculados en los cursos de estadística del departamento de psicología, educación y empresas	Puntos en el trayecto en total, grado supuesto, examen final y combinación de exámenes.	Las mujeres superan a los hombres cuando el resultado es en grados, y los hombres se ven favorecidos cuando se utilizan pruebas.	Educational and Behavioral Statistics
Campbell & Campbell (1997)	Faculty/student mentor program: Effects on academic performance and retention	Estimaciones OLS	339 estudiantes universitarios	Género, etnia, rendimiento académico y estado de inscripción.	El logro académico y la retención no están relacionados con el género y la etnia del mentor, el protegido, o el género y etnia partido entre los dos.	Research in Higher Education

AUTOR	TEMA	METODOLOGÍA	MUESTRA	VARIABLES	CONCLUSIÓN	JOURNAL
Zax & Rees (2002)	IQ, academic performance, environment, and earnings	Estimaciones OLS	Estudiantes de las escuelas públicas de Carolina del Norte en los grados 3 a 8	Características de la comunidad, de la escuela, del hogar, de los amigos y de los compañeros.	Aumentos factibles en el rendimiento académico y la educación pueden compensar los efectos de muchos déficits cognitivos y contextuales.	Economics and Statistics
Shin & Raudenbush (2011)	The Causal Effect of Class Size on Academic Achievement: Multivariate Instrumental Variable Estimators With Data Missing at Random	Modelo multivariante de ecuaciones simultáneas de tres niveles con una variable instrumental	79 escuelas de Tennessee con tres tipos de clases: pequeñas (13-17 alumnos), regulares (22 a 25 estudiantes) y regulares con ayudante (22-25 estudiantes)	Del estudiante: puntaje de lectura, puntaje en matemáticas, puntaje en listening y puntaje en habilidades de estudio de palabras  De la clase: tamaño y tipo	La reducción del tamaño de la clase mejora la lectura, las matemáticas y los resultados de las pruebas .No encuentran evidencia de que los efectos causales varían a través de las escuelas.	Educational and Behavioral Statistics
McArdle, Paskus & Boker (2013)	A Multilevel Multivariate Analysis of Academic Performances in College Based on NCAA Student-Athletes	Modelos de regresión multinivel estándar con predicción no lineal	16000 estudiantes de primer año de universidad que provienen de 260 colegios en el periodo 1994-1995	VARIABLES demográficas del estudiante, características del colegio, variables académicas de la secundaria y variables de los estudiantes de primer año.	Los modelos multinivel estándar son útiles para los estudios de validación estándar y los grados de secundaria son los mejores predictores disponibles de los grados de primer año de la universidad.	Multivariate Behavioral Research

AUTOR	TEMA	METODOLOGÍA	MUESTRA	VARIABLES	CONCLUSIÓN	JOURNAL
Otham & Musa (2014)	The Improvement of Students' Academic Performance by Using Social Media through Collaborative Learning in Malaysian Higher Education	Correlación de Pearson	120 estudiantes de pregrado y posgrado en el rango de edad de 18 a 36 años de la Universiti Teknologi Malasia	Interactividad con sus compañeros, interactividad con los maestros, compromiso, percepción de la facilidad de uso, percepción de la utilidad, aprendizaje colaborativo, satisfacción de los estudiantes y rendimiento académico.	Los medios de comunicación social afectan positiva y significativamente el aprendizaje colaborativo . La mayor correlación se presentó entre el rendimiento académico y la satisfacción estudiantil y compromiso.	Asian Social Science
McArdle, Hamagami, Chang & Hishinuma (2014)	Longitudinal dynamic analyses of depression and academic achievement in the Hawaiian High Schools Health Survey using contemporary latent variable change models	Estudio de cohorte secuencial longitudinal con métodos multivariantes	Más de 5000 estudiantes de secundaria en Hawai	La percepción subjetiva de los logros académicos, la percepción subjetiva de la depresión.	La depresión afecta el rendimiento académico y no al contrario.	Structural equation modeling: a multidisciplinary

Fuente: Elaboración propia.

En lo que respecta a evidencia empírica basada en la teoría learning analytics y su relación con el rendimiento académico se han encontrado pocas investigaciones realizadas recientemente, a continuación se detallan las que se han considerado más importantes.

La investigación de Agudo y otros (2012) se centra en establecer tres tipos de interacción de aprendizaje en línea (agentes involucrados, frecuencia de uso y modo de participación) y analizar la relación de sus componentes con el rendimiento académico de los estudiantes. Lo hacen a través de un análisis de regresión múltiple cuyos resultados permitieron concluir : a) Que la clasificación basada en los agentes involucrados en el aprendizaje en línea ofrece una mejor explicación del rendimiento académico, b) Que al menos un componente en cada tipología predice el rendimiento académico; y c) Que la interacción estudiante-profesor, estudiante-estudiante, evaluaciones desarrolladas por el estudiante y una interacción activa, tienen un impacto significativo sobre el rendimiento académico

Agudo y otros (2014) realizan una relación entre el rendimiento académico y dos modalidades de aprendizaje (aprendizaje virtual apoyado con clases presenciales y el aprendizaje netamente en línea). El principal hallazgo indica que hay una relación entre un cierto tipo de interacciones y el rendimiento académico en los cursos en línea, mientras que esta relación no es significativa en el caso de los cursos F2F apoyados por un entorno virtual de aprendizaje.

Se realiza un análisis de regresión lineal múltiple, en la investigación de Yu & Jo (2014) se determinaron como factores controlables importantes para el rendimiento académico de los estudiantes el tiempo total dedicado a estudiar en un LMS(Learning Management System- Sistema de gestión del aprendizaje), la interacción con los compañeros, la regularidad de intervalo aprendizaje en un LMS, y el número de descargas. Se concluye que las variables del modelo explicaron el 33,5% de la varianza en la calificación final.

Tempelaar y otros (2014) estudiaron el poder predictivo de la disposición para el aprendizaje, los resultados de las evaluaciones formativas continuas y otros datos

generados por el rendimiento de los estudiantes para generar retroalimentación informativa. Utilizan una perspectiva dinámica, longitudinal. Las evaluaciones formativas asistidas por la computadora parecen ser el mejor predictor para la detección de los estudiantes de bajo rendimiento académico, mientras que los datos de LMS básicos no predijeron sustancialmente el aprendizaje.

# CAPÍTULO 3

## MODELOS MULTIVARIANTES COMO METODOLOGÍA PARA LA INVESTIGACIÓN EDUCATIVA

---

En la educación universitaria intervienen muchas variables que pueden ser analizadas a través de un enfoque multivariante. En este capítulo se hace una revisión de dos técnicas de Estadística Multivariante aplicables a la investigación educativa, el análisis multinivel y el análisis bivalente logístico. La ventaja fundamental que presentan estas dos metodologías, a diferencia de la regresión clásica, es que los modelos multinivel permiten incluir en una misma ecuación variables independientes de diferentes niveles de agregación y en el caso de los modelos bivariantes dos variables dependientes cualitativas asociadas.

### **3.1 Análisis a través de una regresión multinivel**

Los modelos multinivel han estado aplicándose con mayor fuerza en el campo de la salud y educación desde hace más de dos décadas (Goldstein y otros 1993; Draper, 1995; Goldstein & Spiegelhalter, 1996).

Actualmente se usa en la investigación educativa, dado que los datos educativos se caracterizan por presentar una estructura anidada. Este método de investigación cuantitativa permite analizar de mejor forma las variables que influyen en el ámbito

educativo, estimar separadamente la varianza entre estudiantes de una misma escuela y la varianza entre escuelas, etc. Algunos ejemplos de anidamiento pueden ser:

- Estudiantes *dentro* de escuelas.
- Estudiantes *dentro* de aulas dentro de escuelas y escuelas *dentro* de universidades.
- Aulas *dentro* de universidades *dentro* de regiones.
- Universidades *dentro* de regiones *dentro* de países.

Cuando se presenta una estructura jerárquica las observaciones individuales no son completamente independientes, por lo que, por ejemplo, la correlación intra-clase entre las variables de los estudiantes de una misma universidad será mayor que la correlación de las mismas variables medidas entre los estudiantes de otras universidades. Un análisis multinivel permite analizar no solamente la relación entre los estudiantes sino también analiza los diversos contextos en los que se desenvuelven ya que los individuos pertenecientes a un mismo contexto tenderán a tener intereses, características y un comportamiento común, con respecto a su pertenencia a distintos contextos.

En base a estas consideraciones Abela (2011) plantea que el objetivo de una regresión multinivel es modelizar estadísticamente la influencia de variables contextuales sobre las actitudes o los comportamientos medidos a nivel individual.

### **3.1.1 Definición formal de los modelos multinivel**

Los modelos de regresión multinivel son conocidos como modelos jerárquicos lineales, modelos de coeficientes aleatorios, modelos de componentes de varianza y modelos lineales mixtos (Hox, 1995).

Un modelo multinivel considera la jerarquía de los datos, es decir, cómo una unidad está anidada dentro de otra superior, por eso todos los datos ya no son estimados en un único

modelo de regresión, sino que se realiza una estimación por grupos o niveles de forma simultánea.

Murillo (2008) señala que los modelos multinivel son, en esencia, ampliaciones de los modelos de regresión lineal clásicos, de tal forma que en realidad son varios modelos lineales para cada nivel.

### **3.1.2 Características de los modelos multinivel**

Los aspectos por los que se caracteriza una regresión multinivel son:

- Incluye dos, tres o más unidades de análisis simultáneamente.
- Permite explorar qué otras variables individuales, grupales o contextuales están controlando la conducta de la variable dependiente.
- Contiene efectos fijos y aleatorios.
- Resuelve el problema de la falacia atomista y ecológica.<sup>1</sup>
- Reconoce la estructura anidada de los datos y permite obtener estimaciones insesgadas de las variaciones acaecida en los distintos niveles de la jerarquía (Vallejo y otros 2008).
- Permite incluir información a nivel del individuo y del contexto.
- Presenta estimaciones correctas de los errores estándar y la significancia estadística de los coeficientes.

---

<sup>1</sup> La falacia atomista acontece cuando se realiza una interpretación agregada a partir de datos individuales y la falacia ecológica se provoca cuando se inferencia a nivel individual a partir de información a nivel grupal.



- Provoca que se analice simultáneamente y de manera correcta, los efectos de las variables en los diferentes niveles de agregación.
- Resuelve el problema estadístico de correlación individual en la estimación OLS, analizando una correlación intragrupal a través de una estructura jerárquica.
- Incluye las variables de nivel micro dentro de las de nivel macro.
- Resuelve la limitación del uso de modelos de regresión múltiple que invalidan la hipótesis de independencia cuando se presenta mayor homogeneidad entre individuos de un mismo grupo respecto a individuos de distintos grupos.

### **3.1.3 Métodos de estimación de los modelos multinivel**

Existen algunos procedimientos de estimación en el contexto de los modelos multinivel englobados en métodos de máxima verosimilitud y de estadística Bayesiana.

#### **3.1.3.1 Máxima verosimilitud**

- a) Mínimos Cuadrados Generalizados Iterativos (IGLS): producen estimadores sesgados de los parámetros aleatorios debido a que no tiene en cuenta la varianza muestral de la parte fija del modelo.
- b) Mínimos Cuadrados Generalizados Iterativos Restringidos (RIGLS): producen estimadores insesgados y en casos en donde la muestra es pequeña, este método sería el más adecuado.
- c) Cuasi verosimilitud Marginal (MQL): Se aplican a modelos no lineales con variables dicotómicas. Permiten linealizar la parte fija del modelo multinivel a través de expansiones en series de Taylor.

- d) Cuasi verosimilitud Penalizada (PQL): Este método adicional a lo que hace el MQL linealiza la parte aleatoria del valor ajustado de la variable dependiente. Por lo tanto, sus estimaciones son más precisas que las obtenidas por medio de MQL.

### 3.1.3.2 Estadística Bayesiana

- a) Full Bayes estimation (FB): requiere de una distribución previa de los parámetros aleatorios, se supone que los coeficientes de la parte fija del modelo son "intercambiables".
- b) Empirical Bayes estimation (EM): ignora la distribución previa de los parámetros aleatorios tratándolos como magnitudes conocidas para propósitos de inferencia, cuando se supone normalidad.
- c) Cadena de Markov - Monte Carlo (MCMC): aprovecha las propiedades de las cadenas de Markov en donde la probabilidad de un evento es condicionalmente dependiente de un estado previo. El mecanismo es iterativo y, en cada etapa, de una distribución normal multivariante se calcula la distribución condicional de cada componente respecto a los restantes, la cual se utiliza para generar una variable aleatoria.

### 3.1.4 Ecuaciones básicas de los modelos Lineales Jerárquicos de 2 y 3 niveles

Los modelos multinivel se aplican de acuerdo a la estructura anidada que presente la población en estudio, ésta básicamente suele ser de 2 o 3 niveles. Conforme se aumentan los niveles se supone que aumenta la heterogeneidad.

El modelo de 2 niveles se compone de dos estimaciones en donde  $i = 1, \dots, n_j$  unidades del nivel 1 se encuentran anidados dentro de  $j = 1, \dots, J$  unidades del nivel 2.

**Tabla 5. Ecuaciones básicas para modelos multinivel de 2 niveles**

	Modelos	Notación
<b>Nivel 1</b>	$Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \dots + \beta_{nj}X_{nij} + e_{ij}$	<p><math>Y_{ij}</math> variable dependiente para el caso <math>i</math> en la unidad <math>j</math>.</p> <p><math>\beta_{mj}</math> coeficientes del nivel 1</p> <p><math>X_{mij}</math> variable explicativa <math>n</math> para el caso <math>i</math> en la unidad <math>j</math>.</p> <p><math>e_{ij}</math> efecto aleatorio del nivel 1*</p>
<b>Nivel 2</b>	$\beta_{nj} = \gamma_{n0} + \gamma_{n1}W_{1j} + \dots + \gamma_{np}W_{pj} + u_{nj}$	<p><math>\beta_{mj}</math> variable dependiente, son los coeficientes del nivel 1</p> <p><math>\gamma_{mp}</math> coeficientes del nivel 2</p> <p><math>W_{pj}</math> variable explicativa del nivel 2</p> <p><math>u_{mj}</math> efecto aleatorio del nivel 2**</p>

**Fuente:**Elaboración propia

\*  $e_{ij}$  se distribuye en forma normal con una media 0 y una varianza  $\sigma^2$

\*\*  $u_{nj}$  se distribuye en forma normal multivariantes con media 0 y una matriz de varianzas  $\tau_2$

El modelo de 3 niveles se compone de tres jerarquías en donde hay  $i = 1, \dots, n_{jk}$  unidades del nivel 1 las que se hallan dentro de  $j = 1, \dots, J_k$  unidades del nivel 2, las que a su vez se anidan en  $k = 1, \dots, K$  unidades del nivel 3.

**Tabla 6. Ecuaciones básicas para modelos multinivel de 3 niveles**

	Modelos	Notación
Nivel 1	$Y_{ijk} = \pi_{0jk} + \pi_{1jk} X_{1ijk} + \dots + \pi_{njk} X_{nijk} + e_{ijk}$	<p><math>Y_{ij}</math> variable dependiente para el caso <math>i</math> dentro de la unidad <math>j</math> y la unidad <math>k</math> del nivel 3.</p> <p><math>\pi_{mjk}</math> coeficientes del nivel 1</p> <p><math>X_{mijk}</math> variable explicativa <math>n</math> para el caso <math>i</math> en la unidad <math>j</math> del nivel 2 y en la unidad <math>k</math> del nivel 3.</p> <p><math>e_{ijk}</math> efecto aleatorio del nivel 1*</p>
Nivel 2	$\pi_{njk} = \lambda_{n0k} + \lambda_{n1k} W_{1jk} + \dots + \lambda_{npk} W_{pjk} + u_{njk}$	<p><math>\pi_{mjk}</math> variable dependiente, son los coeficientes del nivel 1</p> <p><math>\lambda_{mpk}</math> coeficientes del nivel 2</p> <p><math>W_{pjk}</math> variable explicativa del nivel 2</p> <p><math>u_{mjk}</math> efecto aleatorio del nivel 2**</p>
Nivel 3	$\lambda_{npk} = \omega_{np0} + \omega_{np1} Z_{1k} + \dots + \omega_{npq} Z_{qk} + r_{npk}$	<p><math>\lambda_{mpk}</math> variable dependiente son los coeficientes del nivel 2</p> <p><math>\omega_{mpq}</math> coeficientes del nivel 3</p> <p><math>Z_{qk}</math> variable explicativa del nivel 3</p> <p><math>r_{mpk}</math> efecto aleatorio del nivel 3***</p>

\*  $e_{ijk}$  se distribuye en forma normal con una varianza  $\sigma^2$ .

\*\*  $u_{njk}$  se distribuye en forma normal multivariante con media 0 y una matriz de varianzas  $\tau_2$

\*\*\*  $r_{npk}$  se distribuye en forma normal multivariante, con una matriz de varianzas  $\tau_3$

### 3.1.5 Proceso de modelización de 2 y 3 niveles

En el análisis de datos multinivel tomando en cuenta la naturaleza jerárquica de los datos se suele desagregar desde el nivel superior hacia el inferior, este procedimiento consiste en adjudicar a individuos el mismo valor de la variable a nivel superior y agregar las variables del nivel inferior hacia el superior. Los datos individuales son agregados en unidades del nivel superior, eliminando toda la variación intragrupal.

a) Modelo nulo (ANOVA de efectos aleatorios)

En esta primera fase no se incluye ninguna variable explicativa, se incluye la variable dependiente, la constante en la parte fija y los efectos aleatorios en los niveles. Este modelo indica la variabilidad de los datos entre los niveles.

b) Modelo de regresión con interceptos aleatorios

Se amplía el modelo nulo incluyendo variables explicativas en la parte fija, si estas variables son significativas posteriormente se incluirán en la parte aleatoria. En esta modelización se conoce cuál es el aporte de cada una de las variables de los niveles a la explicación de la varianza de la variable dependiente.

c) Expansión del modelo con interceptos aleatorios: modelo con pendientes aleatorias

Se genera un modelo para cada nivel incluidas cada una de las variables que han resultado significativas en la parte fija.

d) Modelo final

Se incluyen todas las variables seleccionadas.

A continuación la presentación de las ecuaciones son adaptadas de Bryk & Raudenbush (1992) para un supuesto caso de investigación con datos de estudiantes  $i$ , anidados en aulas  $j$  y estos a su vez en escuelas  $k$ .

### **3.1.5.1 MODELO MULTINIVEL DE 2 NIVELES**

Partiendo de un modelo nulo las expresiones se van expandiendo secuencialmente.

#### **a) Modelo nulo: ANOVA con efectos aleatorios**

Este análisis permite determinar qué porcentaje de la variación en la variable dependiente se debe a los estudiantes o al aula, por lo que como punto de partida no se incluyen variables predictoras en ninguno de los 2 niveles, pero si efectos aleatorios.

$$\text{Nivel 1} \quad Y_{ij} = \beta_{0j} + e_{ij} \quad (3-1)$$

$$\text{Nivel 2} \quad \beta_{0j} = \gamma_{00} + u_{0j} \quad (3-2)$$

En el nivel 1, la variable dependiente  $Y$  para el estudiante  $i$  en el aula  $j$  está en función de la media  $\beta_{0j}$  de  $Y$  en una aula y el error  $e_{ij}$  que indica como  $Y$  individual se desvía de la media de la  $Y$  del aula.

En el nivel 2,  $\gamma_{00}$  representa la media global, es decir, la media que engloba a todos los estudiantes y aulas y el término de error  $u_{0j}$  que muestra como la media de  $Y$  en una aula en particular se desvía de la media global.

Para ambos niveles el efecto aleatorio se distribuye normalmente, para el Nivel 1 con una estimación de la varianza  $\sigma^2$  (para  $e_{ij}$ ) y para el Nivel 2 con una estimación de la varianza  $\tau_2$  (para  $u_{0j}$ ).

### b) Incluyendo variables predictoras del Nivel 2

Se busca predecir la  $Y_{ij}$  media del Nivel 2 a través de características grupales.

$$\text{Nivel 1} \quad Y_{ij} = \beta_{0j} + e_{ij}$$

$$\text{Nivel 2} \quad \beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j} \quad (3-3)$$

La ecuación del Nivel 1 es la (3-1). En la ecuación (3-3) se incluyen variables explicativas del Nivel 2  $W_j$  y ahora  $u_{0j}$  es un residuo condicionado por la variable  $W_j$  por lo tanto la varianza también será residual.

### c) Incluyendo variables predictoras del Nivel 1-ANCOVA con efectos aleatorios

$$\text{Nivel 1} \quad Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij} \quad (3-4)$$

$$\text{Nivel 2} \quad \beta_{0j} = \gamma_{00} + u_{0j} \quad (3-5)$$

$$\beta_{1j} = \gamma_{10} \quad (3-6)$$

La ecuación (3-4) incluye variables explicativas  $X_{ij}$ . La función (3-6) indica que la relación con la  $X_{ij}$  se restringe a que tenga la misma pendiente para las unidades del nivel 2, es decir se fija la varianza de la pendiente en cero.

**a) Variables predictoras del Nivel 1 como coeficientes aleatorios**

$$\text{Nivel 1} \quad Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij} \quad (3-7)$$

$$\text{Nivel 2} \quad \beta_{0j} = \gamma_{00} + u_{0j} \quad (3-8)$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \quad (3-9)$$

Con estas ecuaciones se puede determinar si los coeficientes de las variables independientes muestran variación significativa entre grupos, por lo que se supone que los interceptos del Nivel 1 varían aleatoriamente en la población de las unidades del nivel 2.

En la ecuación (3-8)  $\gamma_{00}$  es la ordenada promedio de las unidades del Nivel 2 y  $u_{0j}$  es el incremento aleatorio específico de la ordenada asociado a la unidad *j del nivel 2*. En la ecuación (3-9)  $\gamma_{10}$  es la pendiente promedio de la regresión de las unidades del Nivel 2 y  $u_{1j}$  es el incremento aleatorio específico de la ordenada asociado a la unidad *j del mismo nivel*.

**b) Interacciones multinivel entre variables del Nivel 1 y del Nivel 2**

$$\text{Nivel 1} \quad Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij}$$

$$\text{Nivel 2} \quad \beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j} \quad (3-10)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_j \quad (3-11)$$

El modelo permite determinar si las variables grupales  $W_j$  explican la variación de la variable dependiente entre grupos.



La ecuación del Nivel 1 es la (3-7) y en las ecuaciones del Nivel 2 (3-10) y la (3-11) incluyen variables independientes, pero  $\beta_{1j}$  no considera una perturbación aleatoria.

Además se genera una interacción multinivel entre las variables contextuales (aula) e individuales (estudiante).

### c) Interacciones multinivel y coeficientes aleatorios del Nivel 1

$$\text{Nivel 1} \quad Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij}$$

$$\text{Nivel 2} \quad \beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + u_{1j} \quad (3-12)$$

Al igual que el modelo anterior se busca determinar si las variables grupales  $W_j$  explican la variación de la variable dependiente entre grupos, la diferencia es que tanto la ordenada como la pendiente se modelan con efectos aleatorios.

### 3.1.5.2 MODELO MULTINIVEL DE 3 NIVELES

#### a) Modelo nulo-ANOVA con efectos aleatorios

En la parte fija del modelo se incluye la constante y en la parte aleatoria se determina la varianza de los residuos en los tres niveles.

$$\text{Nivel 1} \quad Y_{ijk} = \pi_{0jk} + e_{ijk} \quad (3-13)$$

$$\text{Nivel 2} \quad \pi_{0jk} = \lambda_{00} + u_{0jk} \quad (3-14)$$

$$\text{Nivel 3} \quad \lambda_{00k} = \omega_{000} + r_{00k} \quad (3-15)$$

La ecuación (3-13) del Nivel 1 establece la variable dependiente  $Y_{ijk}$  para el estudiante  $i$  en el aula  $j$  de la escuela  $k$  que está en función de la media  $\pi_{0jk}$  del aula  $j$  en la escuela  $k$  y el efecto aleatorio del estudiante como  $e_{ijk}$  que representa la desviación del rendimiento del estudiante  $ijk$  de la media del aula.

En el Nivel 2, en la ecuación (3-14) se encuentra la variable dependiente como  $\pi_{0jk}$  (media de cada aula) que varía aleatoriamente alrededor de la media de alguna escuela,  $\lambda_{00}$  representa la media de la escuela  $k$  y el efecto aleatorio del aula  $u_{0jk}$  que indica la desviación del rendimiento del aula  $jk$  de la media de la escuela.

En el nivel 3 se representa la variabilidad entre escuelas y la ecuación (3-15) incluye como variable dependiente las medias de las escuelas  $\lambda_{00k}$  en función de la media global de las escuelas  $\omega_{000}$  y donde el efecto aleatorio de la escuela  $r_{00k}$  representa la desviación de la media de la escuela  $k$  de la media global.

Los efectos aleatorios en los niveles 1, 2 y 3 se suponen distribuidos en forma normal con una media cero y una estimación de la varianza  $\sigma^2$  (para  $e_{ijk}$ ),  $\tau_2$  (para  $u_{0jk}$ ) y  $\tau_3$  (para  $r_{00k}$ ).

### **b) Incluyendo variables predictoras del Nivel 2 y del Nivel 3**

Se busca predecir  $Y_{ijk}$  media de las aulas y de las escuelas a través de las características grupales de cada nivel. Se incluyen variables explicativas en los niveles 2 y 3.

$$\begin{aligned} \text{Nivel 1} \quad Y_{ijk} &= \pi_{0,jk} + e_{ijk} \\ \text{Nivel 2} \quad \pi_{0,jk} &= \lambda_{00k} + \lambda_{01k}W_{1jk} + u_{0,jk} & (3-16) \\ \text{Nivel 3} \quad \lambda_{00k} &= \omega_{000} + \omega_{001}Z_{1k} + r_{00k} & (3-17) \end{aligned}$$

La ecuación del nivel 1 es la (3-13). Para el nivel 2 la ecuación (3-16) incluye variables explicativas  $W_{1jk}$  que son características de la aulas ,  $\lambda_{00k}$  es la ordenada de la escuela  $k$  cuando se modela el efecto del rendimiento académico medio  $\pi_{0,jk}$  del aula, y  $\lambda_{01k}$  es el intercepto que indica la dirección y el grado de asociación entre las características de las escuelas y  $\pi_{0,jk}$ , finalmente  $u_{0,jk}$  es el efecto aleatorio del Nivel 2.

En el nivel 3 (3-17) la variable explicativa  $Z_{1k}$  es una característica de la escuela,  $\omega_{000}$  es la media global para  $\lambda_{00k}$ ,  $\omega_{001}$  es el intercepto que indica la dirección y el grado de asociación entre las características de las escuelas y  $\lambda_{00k}$ . El efecto aleatorio de este nivel está representado por  $r_{00k}$ .

### c) Incluyendo variables predictoras del Nivel 1 y Nivel 2

Dentro de este paso se pueden presentar dos casos:

Modelo con variables explicativas  $X_{ijk}$  en el Nivel 1 que toma los mismos valores a lo largo de las unidades del Nivel 2 y 3.

$$\text{Nivel 1} \quad Y_{ijk} = \pi_{0,jk} + \pi_{1,jk}X_{ijk} + e_{ijk} \quad (3-18)$$

$$\text{Nivel 2} \quad \pi_{0,jk} = \lambda_{00k} + u_{0,jk} \quad (3-19)$$

$$\pi_{1,jk} = \lambda_{10k} \quad (3-20)$$

$$\text{Nivel 3} \quad \lambda_{00k} = \omega_{000} + r_{00k} \quad (3-21)$$

$$\lambda_{10k} = \omega_{100} \quad (3-22)$$

Modelo con variables explicativas en del Nivel 1  $X_{ijk}$  y del Nivel 2 y constante a lo largo de las escuelas.

$$\text{Nivel 1} \quad Y_{ijk} = \pi_{0jk} + \pi_{1jk} X_{ijk} + e_{ijk}$$

$$\text{Nivel 2} \quad \pi_{0jk} = \lambda_{00k} + u_{0jk}$$

$$\pi_{1jk} = \lambda_{10k} + \lambda_{11k} W_{1jk} + u_{1jk} \quad (3-23)$$

$$\text{Nivel 3} \quad \lambda_{00k} = \omega_{000} + r_{00k}$$

$$\lambda_{10k} = \omega_{100}$$

$$\lambda_{11k} = \omega_{110} \quad (3-24)$$

#### d) Variables predictoras del Nivel 1 y del Nivel 2 con coeficientes aleatorios

El modelo permite determinar si los coeficientes de las variables independientes muestran variación significativa entre aulas y escuelas. Las ecuaciones son las mismas que las del paso anterior salvo que todos los interceptos en cada uno de los niveles son aleatorios. Por lo tanto se pueden dar dos casos:

- Las variables del Nivel 1 ( $X_{ijk}$ ) se centran respecto a la media del grupo  $jk$ , suponiendo que por lo menos uno de los interceptos del Nivel 1 varía aleatoriamente en la población de las unidades del Nivel 2 y 3. No se definen variables en el nivel 2 y 3.
- Las variables del Nivel 2 ( $W_{1jk}$ ) se centran respecto a la media del grupo del Nivel 3, suponiendo que por lo menos uno de los interceptos del Nivel 2 muestra un comportamiento aleatorio entre unidades del Nivel 3.

### e) Interacciones multinivel

$$\text{Nivel 1} \quad Y_{ijk} = \pi_{0jk} + \pi_{1jk} X_{ijk} + e_{ijk} \quad (3-25)$$

$$\begin{aligned} \text{Nivel 2} \quad \pi_{0jk} &= \lambda_{00k} + \lambda_{01k} W_{01k} + u_{0jk} \\ \pi_{1jk} &= \lambda_{10k} + \lambda_{11k} W_{1jk} + u_{1jk} \end{aligned} \quad (3-26)$$

$$\begin{aligned} \text{Nivel 3} \quad \lambda_{00k} &= \omega_{000} + \omega_{001} Z_k + r_{00k} \\ \lambda_{01k} &= \omega_{010} + \omega_{020} Z_k \\ \lambda_{10k} &= \omega_{100} + \omega_{101} Z_k \\ \lambda_{11k} &= \omega_{110} + \omega_{111} Z_k \end{aligned} \quad (3-27)$$

El modelo permite determinar si las variables grupales del Nivel 3 ( $Z_k$ ) explican la variación de la variable  $Y_{ijk}$ . Se adicionan variables explicativas de los 3 niveles ( $X_{ijk}$ ,  $W_{jk}$ ,  $Z_k$ ), cuyos coeficientes como variables independientes en el nivel 3 no consideran una perturbación aleatoria.

### f) Interacciones multinivel con perturbación aleatoria

Comprende las mismas ecuaciones del ítem anterior (3-25) y (3-26), con la diferencia que en los coeficientes del nivel 3 se considera una perturbación aleatoria.

$$\begin{aligned} \lambda_{00k} &= \omega_{000} + \omega_{001} Z_k + r_{00k} \\ \lambda_{01k} &= \omega_{010} + \omega_{020} Z_k + r_{01k} \\ \text{Nivel 3} \quad \lambda_{10k} &= \omega_{100} + \omega_{101} Z_k + r_{10k} \\ \lambda_{11k} &= \omega_{110} + \omega_{111} Z_k + r_{11k} \end{aligned} \quad (3-28)$$

### **3.1.6 Interpretación de resultados**

Hox (2002) establece que básicamente la interpretación de los modelos multinivel depende de la significación de los coeficientes de regresión y de lo bien que el modelo se ajusta a los datos. A continuación se recogen diversos parámetros y conceptos a través de los cuales se puede alcanzar una adecuada interpretación del modelo que finalmente se ajuste al conjunto de datos considerado en el estudio.

#### **a) Correlación intraclase**

Indica si existe una elevada correlación que se presenta cuando la muestra está conformada por datos homogéneos y se mide a través de la proporción de la varianza explicada por la diferencia entre niveles, esto significa que cuanto mayor sea la varianza dentro de los grupos, mayor será la correlación intraclase.

#### **b) Coeficiente fijo y aleatorio**

Los coeficientes fijos son los efectos medios en la población y los coeficientes aleatorios son las varianzas y covarianzas de los niveles. En los modelos multinivel los coeficientes del primer nivel son tratados como aleatorios en el segundo nivel. Es decir la parte fija es general y común a todos los contextos mientras que la parte aleatoria es representada por lo específico de cada contexto, que varía y se estima por medio de la varianza de los distintos niveles.

### **c) Interacción multinivel**

Se estima la interacción entre variables que están medidas en dos o más niveles de una estructura anidada de datos. Si el término de interacción es significativo cuantifica la diferencia en la influencia de la variable contextual sobre la variable individual, según pertenezca a un grupo u otro.

### **d) Análisis de la varianza**

Se puede realizar el análisis a partir de dos medidas básicas de agrupación como son el Coeficiente de Correlación Intraclase (ICC) y el Coeficiente de Participación de la Varianza (VPC) que dan información acerca de la proporción de la varianza en la variable dependiente que se debe a las diferencias entre los niveles. Es decir que si se obtiene un ICC o VPC cercano a cero significa que ese nivel no condiciona a la variable individual.

## **3.1.7 Modelos multinivel en el campo educativo**

El inicio de la aplicación de estos modelos en el campo educativo se debe principalmente al aporte que realizaron Aitkin & Longford (1986) en su investigación "Statistical modelling issues in school effectiveness studies" introdujeron por primera vez el análisis multinivel para determinar la efectividad escolar, demostrando la existencia de errores metodológicos al usar las regresiones tradicionales en investigaciones anteriores y reconociendo la presencia de una estructura jerárquica en la presentación y análisis de datos entre estudiantes y escuelas.

Los modelos multinivel son más aplicables en el campo educativo porque en estas poblaciones las observaciones individuales no son completamente independientes, es decir se presenta una estructura jerárquica, por lo que según Snijders & Bosker (1999) esto implica una dependencia de las observaciones de nivel micro (alumnos) dentro del nivel macro (aulas o centros). Esta dependencia se refiere a que los estudiantes del

mismo grupo comparten el mismo ambiente, mismos profesores, normas, comunicación, etc.

A continuación se mencionan algunos estudios relevantes:

Nuttall y otros (1989) exploran si hay alguna diferencia en la eficacia de 140 escuelas londinenses marcadas por una diferencia social y composición étnica llegando a concluir que la eficacia escolar varía en función del rendimiento relativo de los diferentes subgrupos.

El estudio realizado por Kreft (1995) compara el rendimiento académico de una asignatura entre escuelas privadas y públicas obteniendo como resultado que los modelos multinivel no pueden determinar cuál de los dos sectores son mejores ya que todo depende de cómo son tratados los datos.

Caldas & Bankston (1999) usan una regresión multinivel para analizar la relación entre el rendimiento académico, el status socioeconómico y la estructura familiar individual y de la escuela. Los resultados de esta investigación sugirieron que la estructura familiar de la escuela presentó una relación más fuerte con el rendimiento académico que con el status socioeconómico de la escuela o la composición racial.

Sellström & Bremberg (2006) determinan la influencia del ambiente de la escuela sobre los resultados académicos, consideran aspectos como: hábitos de fumar, bienestar, problemas de comportamiento y rendimiento académico. Después de procesar la información a través de un estudio multinivel longitudinal, concluyen que el hecho de que una institución cuente con una política antitabaco, un buen clima escolar, estudiantes con un nivel socioeconómico promedio alto influye positivamente sobre el rendimiento de los estudiantes.



## 3.2 Análisis con un modelo logístico bivalente

Este análisis permite diferenciar el doble impacto de las covariables sobre las variables de resultado.

La regresión logística bivalente (del inglés, Bivariate Logistic Regression, BLR), es un modelo de regresión no lineal que se aplica cuando existen dos variables discretas dependientes  $Y_1$  y  $Y_2$ . Dale (1986) fue uno de los primeros en considerar este tipo de modelización.

La regresión logística bivalente tiene dos ventajas: (a) el modelado individual de la distribución de probabilidad marginal de dos respuestas, y (b) el modelado de los odds ratio que describe la asociación entre las dos respuestas en relación a varias covariables (Darda, 2009).

### 3.2.1 Tabla de asociación

La Tabla 7 es el resultado de una clasificación cruzada entre  $n$  observaciones y dos variables cualitativas binarias  $Y_1$  y  $Y_2$ .

**Tabla 7. Asociación de 2 variables de respuesta**

		$Y_2$		<b>Probabilidad</b>
		0	1	
$Y_1$	0	$\pi_{00}$	$\pi_{01}$	$1 - \pi_1$
	1	$\pi_{10}$	$\pi_{11}$	$\pi_1$
<b>Probabilidad</b>		$1 - \pi_2$	$\pi_2$	1

**Fuente:** Elaboración propia

Como se observa en la Tabla 7, las variables objetivo o dependientes del estudio pueden describir cuatro posibles resultados:

$$(Y_1=0, Y_2=0)$$

$$(Y_1=0, Y_2=1)$$

$$(Y_1=1, Y_2=0)$$

$$(Y_1=1, Y_2=1)$$

La probabilidad conjunta de un modelo de regresión logística bivariante se modela con tres componentes sistemáticos:  $\Pr(Y_1=1)$ , la  $\Pr(Y_2=1)$  y la odds ratio  $\psi$ , describiendo así la asociación entre las variables dependientes con las covariables.

### 3.2.2 Odds Ratio

La odds ratio permite representar la asociación entre dos variables en una Tabla 2X2 obteniendo de esta forma una medida para la dependencia entre las variables de respuesta, por lo que, fijándonos en la Tabla 7 se puede expresar como:

$$\psi = \frac{\pi_{11}\pi_{00}}{\pi_{10}\pi_{01}} \quad (3-29)$$

La odds ratio ( $\psi$ ) es fácil de interpretar como la relación de las probabilidades de  $Y_1=1$ , dado que,  $Y_2=1$  y las probabilidades de  $Y_1=1$ , dado que,  $Y_2=0$ . Si hay una independencia entre  $Y_1$  e  $Y_2$ ,  $\psi = 1$  (Le Cessie & Van Houwelingen, 1994)

### 3.2.3 Modelización de la regresión logística bivariante

Las ecuaciones para estimar una regresión logística bivariante son tomadas de Kosuke y otros (2008) generadas para un modelo con variables de respuesta discretas, cualitativas y binarias.

Para cada observación se define dos variables dependientes binarias  $Y_1$  y  $Y_2$  las cuales toman valores 0 o 1. El resultado conjunto se modeliza a través de la probabilidad marginal para cada variable dependiente y la odds ratio  $\psi$ . Se define  $Y_{rs} = 1$  cuando  $Y_1 = r$  y  $Y_2 = s$  y es 0 en el caso contrario, donde  $r$  y  $s$  toman un valor de 0 o 1. Estableciendose el modelo como sigue.

a) El componente estocástico es

$$\begin{aligned} Y_{11} &\sim \text{Bernoulli}(y_{11}|\pi_{11}) \\ Y_{10} &\sim \text{Bernoulli}(y_{11}|\pi_{10}) \\ Y_{01} &\sim \text{Bernoulli}(y_{11}|\pi_{01}) \end{aligned}$$

Donde  $\pi_{rs} = \Pr(Y_1 = r, Y_2 = s)$  es la probabilidad conjunta, y  $\pi_{00} = 1 - \pi_{11} - \pi_{10} - \pi_{01}$ .

b) Los componentes sistemáticos que comprende la probabilidad marginal,  $\pi_j = \Pr(Y_j = 1)$  así como la odds ratio  $\psi$ . En donde la odds ratio de un modelo logístico bivariante es  $\psi_{rc} = \pi_{00}\pi_{01}/\pi_{10}\pi_{11}$ .

La probabilidad conjunta para cada uno de los cuatro resultados posibles puede ser modelada a través de tres componentes sistemáticas: las probabilidades marginales  $\pi_j = \Pr[Y_j=1]$ , para  $j=1,2$ , y la odds-ratio  $\psi_{rc}$  que describe la dependencia estadística entre ambas variables objetivo.

Además, si los datos binarios bivariantes son observados u obtenidos en contextos o situaciones experimentales diferentes descritas a través de un conjunto de variables explicativas o predictoras (que denotaremos por  $\mathbf{X}$ ), estas componentes pueden ser modeladas como funciones de subconjuntos (posiblemente diferentes) de variables explicativas incluidas en  $\mathbf{X}$ .

$$\pi_j = \frac{1}{1 + \exp[\beta_j^T \mathbf{X}_j]} \quad j = 1, 2 \quad (3-30)$$

$$\psi = \exp[\beta_3^T \mathbf{X}_3]$$

donde  $\beta_j$  ( $j = 1, 2, 3$ ) son vectores de parámetros y  $\mathbf{X}_j$  ( $j = 1, 2, 3$ ) son subconjuntos de variables explicativas o predictoras (no necesariamente iguales).

- c) Los valores esperados para el modelo logístico bivariante, son las probabilidades conjuntas. Las estimaciones de  $\beta_1$ ,  $\beta_2$  y  $\beta_3$  se sustituyen en los componentes sistemáticos  $\pi_1$ ,  $\pi_2$ ,  $\psi_{rc}$ , por lo tanto para obtener las estimaciones de las probabilidades conjuntas se consideran las ecuaciones siguientes, en las que se incluyen las estimaciones obtenidas para los componentes sistemáticos.

$$\pi_{11} = \begin{cases} \frac{1}{2}(\psi_{rc} - 1)^{-1} - a_{rc} - \sqrt{a_{rc}^2 + b_{rc}} & \text{if } \psi \neq 1, \\ \pi_1 \pi_2 & \text{if } \psi = 1, \end{cases} \quad (3-31)$$

$$\pi_{10} = \pi_1 - \pi_{11}$$

$$\pi_{01} = \pi_2 - \pi_{11}$$

$$\pi_{00} = 1 - \pi_{10} - \pi_{01} - \pi_{11}$$

Donde  $a_{rc} = 1 + (\pi_1 + \pi_2)(\psi_{rc} - 1)$  y  $b_{rc} = -4\psi_{rc}(\psi_{rc} - 1)\pi_1\pi_2$ , y las probabilidades conjuntas para cada observación deben sumar 1.

### 3.2.4 Modelización de la regresión logística bivariante ordinal

Las ecuaciones para estimar una regresión logística bivariante ordinal son tomadas de Enea & Attanasio (2015).

$A_1 \times A_2$  representa una tabla de doble entrada cruzada con la clasificación de dos variables de respuesta ordenadas  $A_1$  y  $A_2$ , respectivamente con  $D_1$  y  $D_2$  categorías,  $\pi$  es un vector subyacente de las probabilidades.

Se define las probabilidades marginales acumuladas (filas/columnas) como

$$u_r = P(A_1 \leq r) = \sum_{i \leq r} \pi_i,$$

$$u_c = P(A_2 \leq c) = \sum_{j \leq c} \pi_j,$$

Siendo las probabilidades acumuladas del primer cuadrante superior izquierdo

$$u_{rc} = P(A_1 \leq r, A_2 \leq c) = \sum_{i \leq r} \sum_{j \leq c} \pi_{ij}$$

Con  $r = 1, l, \dots, D_1, c = 1, l, \dots, D_2$ . Por diferencia, se obtiene

$$P(A_1 \leq r, A_2 > c) = u_r - u_{rc},$$

$$P(A_1 > r, A_2 \leq c) = u_c - u_{rc},$$

$$P(A_1 > r, A_2 > c) = 1 - u_r - u_c + u_{rc}.$$

Al elegir los odds acumulados como las medidas de riesgo ordinal, y la función logit como función de enlace, obtenemos los logits globales fila/columna, definidos como:

$$\log \varphi_{1r} = \text{logit} [P(A_1 \leq r)] = \log(u_r) - \log(1 - u_r), \quad (3-32)$$

$$\log \varphi_{2c} = \text{logit} [P(A_2 \leq c)] = \log(u_c) - \log(1 - u_c), \quad (3-33)$$

$r = 1, l, \dots, D_1 - 1, c = 1, l, \dots, D_2 - 1$ . Al elegir el cross-product de las probabilidades del cuadrante como medida de asociación ordinal, y la función logaritmo natural como función de enlace, el log-odds ratio global (log-GOR), se define como

$$\log \psi_{rc} = \log \frac{P(A_1 \leq r, A_2 \leq c)P(A_1 > r, A_2 > c)}{P(A_1 \leq r, A_2 > c)P(A_1 > r, A_2 \leq c)} = \log \frac{u_{rc}(1 - u_r - u_c + u_{rc})}{(u_r - u_{rc})(u_c - u_{rc})}.$$

Conociendo los parámetros  $u_r, u_c, \psi_{rc}$ , las correspondientes probabilidades acumuladas conjuntas se pueden obtener a través de la siguiente fórmula de inversión

$$u_{rc} = \begin{cases} \frac{1}{2}(\psi_{rc} - 1)^{-1} \left( a_{rc} - \sqrt{a_{rc}^2 + b_{rc}} \right) & \text{if } \psi \neq 1, \\ u_r u_c & \text{if } \psi = 1, \end{cases}$$

Donde  $a_{rc} = 1 + (u_r + u_c)(\psi_{rc} - 1)$  y  $b_{rc} = -4\psi_{rc}(\psi_{rc} - 1)u_r u_c$ . Si las probabilidades acumuladas  $u_r$  y  $u_c$  satisfacen las restricciones  $u_r < u_{r+1}$  para  $r = 1, l, \dots, D_1 - 1$  y  $u_c < u_{c+1}$  para  $c = 1, l, \dots, D_2 - 1$ , y las odds ratios globales (GOR) no dependen de las categorías, es decir  $\psi_{rc} = \psi$ , a continuación, se establece las odds ratio de un modelo logístico bivalente ordinal.

$$\begin{aligned}
\log(\phi_{1r}(x_i)) &= \beta_{10r} + \beta_1' x_i, \\
\log(\phi_{2c}(x_i)) &= \beta_{20c} + \beta_2' x_i, \\
\log(\psi_{rc}(x_i)) &= \beta_{30rc} + \beta_3' x_i,
\end{aligned}
\tag{3-36}$$

Con  $r = 1, l, \dots, D_1 - 1$ ,  $c = 1, l, \dots, D_2 - 1$ , y donde  $\phi_{kr}$ ,  $k = 1, 2$  son odds globales,  $\psi_{rc}$  son odds ratios globales. Los parámetros  $\beta$  son desconocidos y  $x_i$  es el vector de covarianza de longitud  $p$  para la unidad  $i$ , con  $i = 1, \dots, m$ , donde  $m$  es el número observado de configuraciones de respuesta. Nótese que la ecuación (3-36) es un sistema de ecuaciones  $D_1 \times D_2 - 1$ , en las que se supone las covariables para tener un efecto proporcional en las categorías de las variables de respuesta.

### 3.2.5 Modelos bivariantes aplicados a la medición del rendimiento académico

La evidencia empírica aún es muy limitada, por lo que en este apartado se menciona brevemente una única investigación relevante encontrada al respecto.

A través de un modelo bivariado logístico, Enea & Attanasio (2015) analizan el éxito de los estudiantes de una universidad italiana, tomaron en cuenta como “rendimiento cualitativo”, el promedio de calificaciones y como “rendimiento cuantitativo” los créditos universitarios acumulados. Los resultados del análisis mostraron que los estudiantes universitarios que provienen de colegios de tipo clásico o científico tienen un mejor rendimiento que los que provienen de un colegio técnico. Además se determinó que la nota final de secundaria es un predictor significativo y que el nivel de ingresos familiares no influye en el rendimiento académico de los estudiantes.

# **CAPÍTULO 4**

## **METODOLOGÍA DEL MODELO DE ESTUDIO**

---

En el presente apartado se explica la metodología aplicada a un conjunto de datos reales en cuyo caso de estudio el objetivo general es medir desde una perspectiva cualitativa y cuantitativa el rendimiento académico de estudiantes universitarios que desarrollan sus estudios en modalidad abierta y a distancia.

Para este trabajo se usa información del año 2014 de una Universidad Ecuatoriana que cuenta con el mayor número de estudiantes matriculados en modalidad a distancia (más de 30.000 estudiantes) a nivel de todo el Ecuador y es referente para las Universidades Latinoamericanas.

En el proceso de inclusión de variables usadas para la modelización del rendimiento académico se tuvo en cuenta el enfoque de enseñanza centrada en la teoría del “learning analytics”, por lo que se trabaja con datos suministrados por el Entorno Virtual de Aprendizaje, una de las herramientas de apoyo principales en esta modalidad de estudio.

En este capítulo se justifica y se hace una descripción estadística de las variables de respuesta y covariables que se incluyen en ambos modelos, posteriormente se da a



conocer como está constituida la población objeto de estudio, se detalla las fuentes de datos empleadas y, finalmente, en base a las características de las variables y las muestras, se plantea y justifica el porqué de la aplicación de las técnicas multivariantes.

## **4.1 Variables objetivo consideradas**

Las variables se han seleccionado en pro del cumplimiento de los objetivos específicos planteados en este capítulo. Estas variables son de carácter académico, demográfico y tecnológico (en el ámbito tecnológico se trabaja con variables que involucra el enfoque “learning analytics”).

Se toman en cuenta variables individuales del estudiante (nivel inferior), variables del docente y asignatura (nivel intermedio) y variables de la escuela (nivel superior).

Todas las variables se obtienen dentro de la misma universidad, de esta forma, se supone que la correlación promedio (conocida como la correlación intraclase) entre las variables de los alumnos de la misma universidad y del mismo tipo de asignatura (troncal) es mayor que la correlación de las mismas variables medidas entre los alumnos de universidades distintas.

### **4.1.1 Justificación de las variables**

Existe una infinidad de variables inherentes a la medición del rendimiento académico, pero para el presente estudio se incluyeron aquellas variables fáciles de identificar, con un aceptable índice de precisión, de carácter controlable y que se hallan respaldadas por la teoría.

La evidencia empírica contradice la significatividad estadística de algunas variables con respecto al rendimiento académico, sin embargo, con la finalidad de corroborar estas conclusiones, se incluyen en las estimaciones de los modelos.

## A) VARIABLES DE RESPUESTA

**Tabla 8. Variables como indicadores del rendimiento académico**

<b>Variables</b>	<b>Tipo</b>	<b>Representación</b>
Nota final	Cuantitativa	0 a 40 puntos
Grado	Cualitativa	A a E
Créditos acumulados	Cualitativa	1 a 5

**Fuente:** Elaboración propia

**Rendimiento académico:** El objetivo general es medir el rendimiento académico desde un enfoque cualitativo y cuantitativo. En la Tabla 8 se presentan los tres indicadores considerados.

La variable que se toman en cuenta como indicador cuantitativo es la calificación *final* del estudiante que se mide en un rango de 0 a 40 puntos (incluye la sumatoria de los exámenes, trabajos a distancia y otras actividades).

Las calificaciones son consideradas como un indicador preciso y accesible que refleja los logros académicos (Tournon, 1984; Tirados, 1990; Rodríguez, Fita, & Torrado, 2004).

En términos cualitativos el rendimiento académico puede ser medido a través de dos variables de respuesta: Grado (la calificación promedio expresada como A, B, C, D, E) y la acumulación de Créditos universitarios expresada como (1, 2, 3, 4, 5).

El grado o calificación académica corresponde a la calificación media obtenida de todas las asignaturas superadas, la escala expresa una calificación entre 28 a 40 puntos, siendo 28 el mínimo para aprobar. Realizando las equivalencias de calificaciones con el Sistema Europeo de Transferencia de Créditos (ECTS, en sus siglas en inglés) basada en percentiles, la escala de grados que se establece para el presente estudio queda como se recoge en la Tabla 9.

**Tabla 9. Calificación promedio expresada es escala**

<b>GRADO ECTS</b>	<b>DEFINICIÓN</b>		<b>PUNTAJE</b>
A	Sobresaliente	Excelente resultado con escasas insuficiencias	40-39
B	Muy buena	Superior a la media pero con algunos errores.	38-36
C	Buena	Buen trabajo en general con algunos errores notables.	35-33
D	Satisfactorio	Bien pero con carencias significativas.	32-30
E	Suficiente	Se alcanza el criterio mínimo requerido.	29-28

**Fuente:** Elaboración propia basada en el Sistema Europeo de Transferencia de Créditos (ECTS)

La última variable de respuesta es **créditos acumulados**. En recientes estudios se menciona que los créditos universitarios acumulados son un buen indicador para medir el rendimiento académico puesto que estos indican el avance que el estudiante ha tenido en la carrera durante determinado periodo de tiempo (Díaz y otros 2002; Martín y otros 2008).

La acumulación de créditos corresponde a la suma de créditos de las asignaturas superadas por el estudiante y se miden al final del último año de permanencia en una carrera de cinco años. La agrupación por créditos acumulados fue realizada por la universidad para detectar a que grupo pertenecen aquellos estudiantes que deberían estar en último año de la carrera. La escala para las carreras de 5 años (282 créditos) es la recogida en la Tabla 10.

**Tabla 10. Agrupación por créditos universitarios acumulados**

<b>CRÉDITOS</b>	<b>DEFINICIÓN</b>	<b>GRUPO</b>
menos de 50	Tiene menos del 20% de créditos aprobados.	1
de 50 a 109	Tiene entre el 39% y 20% de créditos aprobados	2
de 110 a 159	Tiene entre el 59% y 40% de créditos aprobados	3
de 160 a 219	Tiene entre el 79% y 60% de créditos aprobados	4
de 220 a 282	Tiene el 80% o más de créditos aprobados	5

**Fuente:** Elaboración propia en base al reglamento de la universidad tomada como caso de estudio

## **B) COVARIABLES**

Las variables explicativas o convariables se recogen en la Tabla 11. Pueden ser clasificadas como sigue: características de los estudiantes, que incluyen un grupo de variables sociodemográficas, antecedentes académicos y variables del enfoque “learning analytics”; características del aula, que incluyen variables de la asignatura y del docente; y características de la escuela, tales como área académica, tamaño, tasa de reprobación y ratio alumno-profesor.

Se realiza una revisión de las más importantes y se justifica su inclusión en esta investigación.

### Covariables asociadas al estudiante:

**Edad**, para el caso de estudiantes universitarios en modalidad a distancia generalmente el rango de edad es amplio. Page (1990) señala que la edad puede incidir en un rendimiento académico diferencial. Esta variabilidad en cuanto a edad conlleva, o bien diferencias madurativas de desarrollo, o bien diversidad de valores, actitudes, expectativas y motivación.

**Tabla 11. Variables explicativas**

VARIABLES	TIPO	REPRESENTACIÓN	DIMENSIÓN
Edad	Cuantitativa	-	Sociodemográfica
Género	Cualitativa	M – F	
Tamaño de Región de procedencia	Cualitativa	Pequeña – Grande	
Repite asignatura	Cualitativa	Si – No	Antecedentes académicos
Rinde supletorio	Cualitativa	Si – No	
Tiene beca	Cualitativa	Si – No	
N° de consultas al profesor	Cuantitativa	-	Learning analytics
N° de comentarios	Cuantitativa	-	
N° de accesos al LMS	Cuantitativa	-	
N° accesos al curso de asignatura	Cuantitativa	-	
Tiempo de uso LMS	Cuantitativa	-	
Participación en foros	Cualitativa	Si – No	
Participación en chat	Cualitativa	Si – No	
Participación en video-colaboración	Cualitativa	Si – No	Asignatura
N° de matriculados	Cuantitativa	-	
N° de repetidores	Cuantitativa	-	
N° Créditos	Cuantitativa	4 – 6	
Ciclo	Cuantitativa	1 -10	Docente
Años de experiencia	Cuantitativa	-	
Formación académica	Cualitativa	Grado-Maestría-Doctorado	
Evaluación docente	Cuantitativa	0 – 60	
Tipo de docente	Cualitativa	Tiempo completo -Invitado-Administrativo	Escuela
Área académica	Cualitativa	Administrativa-Biológica-Humanística-Técnica	
Tamaño	Cuantitativa	-	
Tasa de reprobación	Cuantitativa	-	
Ratio alumno profesor	Cuantitativa	-	

Fuente: Elaboración propia

Por ejemplo en el campo de la tecnología, la edad podría estar interviniendo en un mayor o menor uso de estas herramientas ya que los estudiantes de edad avanzada no son “nativos digitales”<sup>2</sup>, es así que el hecho de que un estudiante tenga más edad respecto a la media del grupo puede llegar a afectar negativamente al rendimiento. Duff y otros (2004) consideran que la edad se ha convertido en una variable importante, puesto que en los últimos años la población estudiantil es más heterogénea.

Los resultados de algunos estudios son contradictorios, por tal motivo para establecer adecuadamente relaciones entre el rendimiento y la edad se recomienda incluir otras características (Schram, 1996; Richardson y otros 1999; Sadler-Smith, 1996; Tejedor, 2003; Nasir, 2012; Alhajraf & Alasfour, 2014 ).

**Género**, respecto a esta variable demográfica los resultados de algunos estudios son contradictorios, en algunas investigaciones se indica que no se presenta significancia en la relación entre el rendimiento académico y el género de los alumnos (Borde, 1998; Moral, 2006; Acevedo & Rocha , 2011; Pantoja & Alcaide, 2013) y en otras se concluye que el hombre o la mujer generan un rendimiento diferente significativo (Haist y otros 2000; Gammie & Gammie, 2003; Dayioğlu & Türüt-Aşik, 2007; Sepúlveda y otros 2011).

Sin embargo, con el objeto de corroborar tales afirmaciones en el ámbito de nuestro estudio, la variable se incluye en el análisis.

**Región de procedencia**, los estudios de modalidad a distancia se ofertan no solamente para estudiantes de distintas edades sino también pertenecientes a las 24 provincias de Ecuador, por lo que cada región de residencia de los estudiantes es diferente, lo cual podría ocasionar alguna influencia sobre el rendimiento académico.

---

<sup>2</sup> Esta palabra fue acuñada por Prensky (2001) quien ha denominado como nativos digitales a aquella generación que ha nacido durante las décadas de los años 1980 y 1990 y se ha formado utilizando “lengua digital” de juegos por ordenador, vídeo e Internet.

Los estudiantes en diversas ubicaciones geográficas tienen conocimientos específicos, experiencias previas y modos de vida. Cada uno de ellos juega un papel crítico en cómo aprenden los estudiantes y, por tanto, cómo los profesores guían su instrucción para satisfacer las necesidades educativas de sus alumnos (Ortlieb & Cheek, 2008).

Algunos estudios indican que no hay un efecto significativo (Porto & Di Grescia, 2004; Garzón y otros 2010), mientras que en otros resultados señalan que si existe una influencia sobre el rendimiento académico (Cresswell & Underwood, 2004).

**Becas**, el hecho de que un estudiante sea beneficiario de una beca estudiantil puede convertirse en un motivo para incentivar o presionar al estudiante a mantener un rendimiento académico alto que le permita seguir siendo beneficiario de la beca. Corroborando este criterio se encuentran algunos estudios que determinan que existe una relación positiva y estadísticamente significativa entre las becas y el rendimiento académico (Garzón & otros, 2010; Celis y otros 2005).

### **Variables del enfoque de “learning analytics”**

Existen algunas variables que se pueden considerar en el enfoque del “learning analytics”, las que involucran las variables explícitas (relacionadas con tareas y exámenes) y las que involucran las variables tácitas (relacionadas con las actividades extracurriculares e interacciones en línea) y es en las *interacciones en línea* en las que nos hemos concentrado en la presente investigación.

Las interacciones en línea comprenden datos producidos por el estudiante en un entorno digital para determinar si afectan al rendimiento académico y, más adelante, a través de ellas, predecir y/o asesorar en estrategias de mejora del aprendizaje cooperativo y del rendimiento. En modalidad de estudios a distancia la interacción en línea entre estudiante y profesor es de gran apoyo para el aprendizaje.

Algunas investigaciones que han realizado la predicción del rendimiento académico a través del “learning analytics” en entornos virtuales de aprendizaje, concluyen que se ha presentado una relación significativa entre las actividades realizadas en un entorno virtual y el rendimiento académico (Agudo y otros 2012; Yu & Jo, 2014).

#### Covariables asociadas al docente:

**Años de experiencia y formación académica.** Se supone que si un docente tiene experiencia dando clases de una asignatura, ha adquirido habilidades y destrezas que influirán favorablemente en el rendimiento del alumno. Si además su formación es continua amplía sus conocimientos, lo que debería reflejarse en la acción de un docente que aplica y enseña con estrategias que permiten al estudiante no solamente aprender sino también a obtener un mejor rendimiento académico.

Sin embargo, algunos estudios muestran que estas variables provocan un efecto leve o a veces, ningún efecto sobre el rendimiento académico (Coleman, 1968; Penny & White, 1998) mientras que otros indican que el predictor “años de experiencia” presenta una relación positiva y directa con el rendimiento académico de los estudiantes (Biniaminov & Glasman, 1983).

#### Covariables asociadas a una asignatura:

**Número de estudiantes matriculados.** El tamaño del aula podría influir en el rendimiento académico, dado que la acción docente del profesor se verá afectada por la dispersión de su atención y la disminución de las estrategias docentes personalizadas.

Algunos estudios indican que cuando el tamaño del aula es pequeño los estudiantes obtienen un mejor rendimiento (Turner y otros 1986; Molnar y otros 1999; Pong & Pallas, 2001) y otros, como el estudio realizado por Hoxby (2000) muestran la falta de evidencia de una correlación con el rendimiento académico.



### Covariables asociadas a una escuela:

**Ratio alumno-profesor**, permite conocer cuál es el número de alumnos dependientes de un profesor. Con respecto a la influencia que ejerce sobre el rendimiento académico, existen algunos estudios que concluyen que a medida que aumenta el número de estudiantes por profesor, el rendimiento disminuye Glass & Smith (1978), mientras que otros determinan que no existen efectos estadísticamente significativos (Roig, 1983).

## 4.2 Población objeto de estudio

Se usan dos muestras para determinar el rendimiento académico desde un enfoque cualitativo y cuantitativo.

La primera muestra sirve para hacer el análisis cuantitativo, comprende un ámbito individual, grupal y contextual. Los participantes que la conforman son:

- 23583 estudiantes matriculados en alguna asignatura troncal. Se excluyen de la muestra las asignaturas de carácter genérico, libre configuración, complementarias y formación básica, con la finalidad de mantener una mayor homogeneidad puesto que las materias troncales se caracterizan por proporcionar los contenidos específicos y propios de la carrera y además incluyen en el proceso de enseñanza aprendizaje el uso de aspectos tecnológicos que permitirán obtener las variables involucradas en el learning analytics.
- 468 aulas que engloban a 385 docentes (tiempo completo, invitados y administrativos-docentes) que imparten clases, aproximadamente, de 328 asignaturas de ámbito troncal.

- 19 escuelas que se ofertan en modalidad a distancia.

La segunda muestra se usa para hacer el análisis cualitativo y está conformada por:

- 410 alumnos considerados como estudiantes que deberían estar en el último año de una carrera de 5 años , el periodo de estudios comprende abril 2009 – abril 2014
- 10 carreras que se ofertan en modalidad a distancia cuyo plan de formación comprende 5 años de estudio.

### 4.3 Análisis estadístico descriptivo

Una vez que se ha determinado las muestras y se ha realizado la selección de variables, se puede realizar un análisis estadístico descriptivo para las dos muestras.

Según la selección de variables descritas en el apartado 4.1 se obtienen 27 variables de **la primera muestra**. En la Tabla 12, se recoge un análisis descriptivo de las mismas.

El *rendimiento académico* medido por la nota del estudiante en una escala de 0 a 40 puntos presenta un promedio global para esta investigación de 24 puntos.

De los 23583 estudiantes, más de la mitad de la muestra la conforman mujeres con un promedio de 29 años de edad. La edad de los estudiantes de modalidad a distancia tiende a presentar un rango muy amplio (17-73 años de edad) el cual se puede explicar porque en esta modalidad de estudios existen alumnos que están retomando sus estudios o son individuos que están estudiando otra carrera, etc. La mayor parte de los estudiantes proviene de la Región Sierra, una de las más grandes del Ecuador.

Un bajo número de estudiantes tiene *beca* (1%), dentro de éste grupo se tomó en cuenta becas por nivel de ingresos y méritos académicos.

Las variables que comprenden los *antecedentes académicos* son cualitativas y nos indican que la mayoría de los estudiantes (79%) están tomando por primera vez la asignatura, es decir que apenas la quinta parte de los matriculados toma la materia por segunda o tercera vez y aproximadamente la mitad de los estudiantes se *quedan suspensos* (53%).

Finalmente dentro del nivel 1 también se encuentran las variables de la dimensión del “learning analytics”:

La variable consultas al profesor, considera el número de mensajes enviados al docente a través de la plataforma virtual. Se observa que la interacción es baja ya que el promedio es de 1 mensaje por estudiante.

El número de comentarios, toma en cuenta todos aquellos comentarios hechos en los anuncios o recursos colocados en el entorno virtual. Igual que en la variable anterior se observa una participación muy baja, esto podría indicar que los anuncios o recursos no son atractivos para el estudiante o que el estudiante no está incluyendo dentro de sus actividades académicas la revisión de ellos, entre otras razones.

Las variables que *contemplan la participación en foros, video-colaboración y chat*, son cualitativas ya que para un periodo de estudio se establecen como actividad obligatoria del docente (dos al bimestre). La opción de participar de los estudiantes es opcional, a pesar de que para aquellos que lo hagan reciben un punto adicional a la nota final, esto se podría considerar como un incentivo, sin embargo, los datos muestran que a pesar de esta estrategia más de la mitad de los estudiantes no participan.

**Tabla 12. Análisis descriptivo de las variables utilizadas en el estudio de la influencia de variables individuales, grupales y contextuales sobre el rendimiento académico.**

Descripción		Media	Rango	Desviación estándar		Dimensión			
Variable dependiente	Rendimiento académico	24	0 - 40	9.270		Dimensión			
	Edad	29	17-73	8.189.834					
Variables independientes	Nivel 1 Estudiantes (23583)	Categoría base (%)		Categoría de contraste (%)		Sociodemográficas			
		Género	Hombre	40	Mujer		60		
		Región	Pequeñas	Extranjero	2		Grandes	Costa	25
				Insular	1			Sierra	64
				Oriental	8				
		Repite la asignatura	No	79	Si	21	Antecedentes		
		Rinde supletorio	No	47	Si	53			
		Tiene Beca	No	99	Si	1			
				Media	Rango	Desviación estándar		Learning analytics	
		N° de consultas al profesor	1	0-13	0.348				
	N° de comentarios	1	0-47	2.361					
	N° de accesos LMS	51	0-772	65.213					
	N° de accesos asignatura	1	0-27	0.257					
	Tiempo de uso LMS	19	0-3701	28.801					
			Categoría base (%)		Categoría de contraste (%)				
	Participación en foros	No	62	Si	38				
Participación en video	No	80	Si	20					
Participación en chat.	No	75	Si	25					
		Media	Rango	Desviación estándar		Asignatura			
Nivel 2 Aulas (468)	Número de matriculados	91	20 - 110	22.005					
	Número de repetidores	19	0 - 70	15.042					
	Número de créditos	4	4-6	0.619					
	Ciclo de asignatura	3	1-10	2.331					
	Años de experiencia	4	1-17	3.984					
		Categoría base (%)		Categoría de contraste (%)		Docente			
Formación académica	Maestría	43	Grado	54					
	Doctorado	3							
Tipo de docente	Invitado	7	Tiempo completo	51					
	Administrativo	42							
Nivel 3 Escuelas (19)	Área académica	Biológica	8	Administrativa	37				
		Socio-Humanística	50						
		Técnica	5						
			Media	Rango	Desviación estándar		Global		
Tamaño	121	24 - 201	64.447						
Tasa de reprobación	0.64	0.45 - 0.84	0.0749						
Ratio alumno profesor	36	6-56	11.553						

Fuente: Elaboración propia

Un estudiante de modalidad a distancia accede a la plataforma virtual en promedio 51 veces durante todo el periodo de estudio. *El tiempo de uso de la plataforma*, es representado en minutos acumulados durante todo el bimestre, alcanzando un rango muy amplio que va desde el estudiante que no ha ingresado nunca hasta el que lo ha hecho durante 3700 minutos.

En el nivel 2 (aulas) se contemplan variables del docente y de la asignatura:

- El *número de alumnos matriculados* en una asignatura troncal es, en promedio, de 91 estudiantes y las asignaturas troncales tienen entre 4 y 6 créditos y se dictan desde primero a décimo ciclo.
- Los docentes tienen una media de 4 años de experiencia y un poco más de la mitad posee título de máster o doctorado y trabajan a tiempo completo en la universidad.
- *La evaluación del desempeño docente* realizada por el estudiante considera una calificación de 60 puntos. Para el periodo de estudio, la media es de 45 puntos.

En el modelo de 3 niveles se incluye la jerarquía de las escuelas, las variables tanto del nivel 1 y 2 del ítem anterior se usan en el análisis del modelo de 3 niveles, pero específicamente las variables que representan a este grupo son:

- La variable *área académica* considera las áreas: administrativa, biológica, técnica y socio-humanística, siendo esta última la que tiene la mayor cantidad de estudiantes.
- El tamaño de la escuela es medido por el número de aulas con estudiantes y docentes designados por componente, en promedio cada escuela tiene 122 aulas.

- La tasa de reprobación global es medida por el número total de estudiantes matriculados con respecto a los estudiantes que reprobaron, en promedio la tasa de reprobación por escuela es del 64%.
- El ratio alumno-profesor indica que a nivel de escuela un profesor en promedio tiene 36 estudiantes a su cargo. Los docentes se contabilizan por escuela.

**Tabla 13. Estadística descriptiva de las variables de la segunda muestra**

Descripción		Categoría base (%)		Categoría de contraste (%)		Dimensión		
Variable dependiente	Grado	C (35 a 33)	25	B (38 a 36)	4	Rendimiento académico		
		D (32 a 30)	54					
		E (29 a 28)	17					
	Créditos	2 (50 a 109)	11	5 ( $\geq 220$ )	28			
		3 (110 a 159)	27					
		4 (160 a 219)	34					
Variables independientes	Edad de ingreso	Adultos		Jóvenes	$\leq 39$ años	88	Sociodemográficas	
	Género	Hombre		Mujer		56		
	Tamaño Región	Pequeñas	Extranjero	5	Grandes	Sierra		65
			Oriental	6		Costa		24
	Participación en actividades en línea <sup>3</sup>	Alta		16		Baja	56	Learning analytics
		Media		28				

Fuente: Elaboración propia

En la **segunda muestra** se consideran 6 variables. En la Tabla 13 se presenta la descripción estadística de las variables que se consideran como dependientes e independientes.

En la muestra de 410 estudiantes, se determina que la variable *Grado* está comprendida en cuatro categorías (notable, bien, satisfactorio, suficiente) la categoría sobresaliente (A) no se incluye puesto que ninguno de los estudiantes del grupo de estudio está dentro de

<sup>3</sup> La variable de participación en actividades en línea es considerada para el enfoque de learning analytics y se obtiene luego de hacer un análisis clúster entre el número promedio de veces que un estudiante participó en un chat, en un foro y en una video-colaboración durante el periodo 2009-2014.

este rango (40-39). Con la finalidad de poder realizar el modelo logístico bivariante se transforma la variable en dicotómica, siendo, la variable de contraste Notable (B) y todas las demás, la categoría base (0). Dentro del grupo la categoría D (Satisfactorio 32-30) es la que representa un poco más de la mitad de la muestra (54%).

La segunda variable dependiente es *créditos acumulados*, comprendida en cuatro categorías. Al igual que en la variable dependiente anterior se transforma en una variable dicótoma, siendo el grupo 5 ( $\geq 220$ ) la que se establece como la categoría de comparación y todos los demás (4, 3 y 2) como la categoría base (0). De acuerdo a la estadística descriptiva se detecta que únicamente el 28% de los estudiantes se los podría considerar como estudiantes de último año.

La transformación en variables dicotomas de las variables respuesta podrán tomar otra forma de acuerdo a la asociación que se desee representar en las estimaciones.

Analizando las variables independientes *sociodemográficas* se observa que el porcentaje de mujeres es un poco mayor a la de los hombres y que más del 50% son estudiantes jóvenes. Las edades de los participantes se segmentaron en dos grupos en base a la clasificación realizada por Martín (2005) en el trabajo "Los factores definitorios de los grandes grupos de edad de la población: tipos, subgrupos y umbrales". Dichos grupos son: jóvenes (desde los 20 a 39 años) y adultos (desde los 40 a más de 50 años), en este segundo grupo se agrupa a los denominados adultos medios (desde los 40 a los 49 años) y a los adultos mayores (desde los 50 años).

La variable participación en actividades en línea se incluye en el modelo de regresión como representante del enfoque "learning analytics". Esta variable es el resultado de realizar un **análisis clúster** entre las variables número chats, número de foros y número de video-colaboraciones en las que ha participado el estudiante en promedio durante su periodo de estudio, estas variables son medidas en escala de razón.

El análisis cluster aplicado es de tipo no jerárquico, este tipo de métodos requieren que se defina previamente el número de grupos que se desea obtener y parten de un grupo de semillas aleatorias que en cada iteración son actualizados por el centroide del cluster. El método utilizado es el conocido como método de las k-medias, recomendado para muestras grandes y que se realiza siguiendo las siguientes etapas (Xu & Wunsch, 2009):

1. Partición en K grupos de forma aleatoria o en base a centroides preexistentes.
2. Cálculo de la matriz prototipo del clúster  $M = [ m_1 , \dots , m_K ]$
3. Asignación de cada elemento al clúster más próximo.
4. Recálculo de la matriz del cluster en base a la nueva partición.
5. Se repite los pasos 2 y 3 hasta que el centroide del clúster no cambie

En el caso del análisis discriminante, es utilizado en este apartado para describir la estructura obtenida previamente con el análisis cluster y por lo tanto verificar su nivel de precisión. La pertinencia del análisis discriminante como criterio para validar los resultados del análisis cluster es argumentada por Díaz de Rada (2002). El nivel de discriminación de las variables independientes implica una combinación lineal de estas a través de la siguiente función:

$$\begin{aligned}
 Y_1 &= A_{10} + A_{11}X_1 + A_{12}X_2 + \dots + A_{1n}X_n \\
 &\vdots \\
 Y_j &= A_{j0} + A_{j1}X_1 + A_{j2}X_2 + \dots + A_{jn}X_n
 \end{aligned}$$

Donde cada  $Y$  equivale a la puntuación discriminante de la observación  $j$ ; y  $X_n$  es la ponderación de la variable n-ésima.



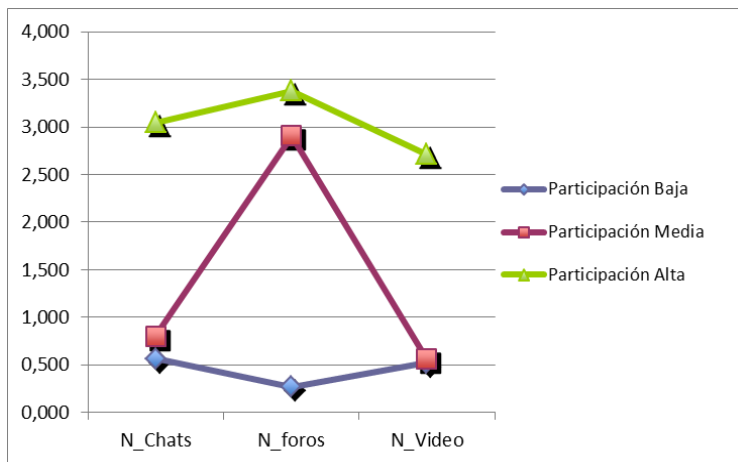
El resultado del análisis discriminante, en cada caso es el nivel de precisión alcanzado, cuyos resultados se presentan en la Tabla 14.

**Tabla 14. Resultados del análisis discriminante**

Clasificaciones	Nivel de precisión
Clasificación en dos grupos	96.60%
Clasificación en tres grupos	99.50%
Clasificación en cuatro grupos	99.80%
Clasificación en cinco grupos	99.30%

Fuente: Elaboración propia

Considerando que los niveles de las clasificaciones para tres, cuatro y cinco grupos son altas y cercanas entre sí, se opta por seleccionar la clasificación que prestaba mayor facilidad de interpretación (la que distribuye a los elementos en tres grupos). Con estos criterios, se selecciona la clasificación que divide a los estudiantes en tres grupos con una precisión del 99,5%, a los que se denomina participación alta, participación media y participación baja.



**Ilustración 5. Centroides de la variable de participación en actividades en línea**

Fuente: Elaboración propia

Los grupos obtenidos (participación alta, participación media, participación baja) presentan valores para los centros finales de los clúster con los que se pueden representar y analizar cada uno de ellos.

Realizando el análisis de estos tres grupos se determina que:

- Los estudiantes con una participación alta son aquellos que se consideran como activos en los chat, foros y sesiones de video-colaboración, este grupo representa el 16%.
- Los estudiantes con una participación media, se caracterizan por tener una participación baja en chats y en sesiones de video-colaboración pero participan activamente en foros, este grupo representa el 28% del total de estudiantes.
- Finalmente el último grupo denominado como los de participación baja, son estudiantes pasivos puesto que su participación en chats, foros y video-colaboración es sumamente leve, este es el grupo más grande de la muestra ya que representan el 56%.

#### **4.4 Fuentes de datos**

Los datos utilizados en la estimación de los modelos se obtuvieron de las siguientes bases de datos y fuentes:

- Para obtener algunas variables que implica el enfoque “learning analytics” se usó el informe proporcionado por el Entorno Virtual de Aprendizaje (Moodle).
- Los datos que comprenden las características del estudiante se obtuvieron del sistema de información académica de la institución educativa (Syllabus Plus).

- Las características de los docentes consideradas como relevantes en este estudio, se obtienen del expediente digital de Recursos Humanos y del sistema de evaluación integral de la institución.
- Las variables correspondientes al nivel de asignatura, se obtuvieron de la planificación académica anual.

Para el caso del modelo multinivel los datos se ordenaron jerárquicamente, de tal forma que las observaciones se agrupen correctamente en cada uno de los niveles de agregación. El tiempo que se tomó en cuenta para levantar los datos de las variables fue el periodo 2014.

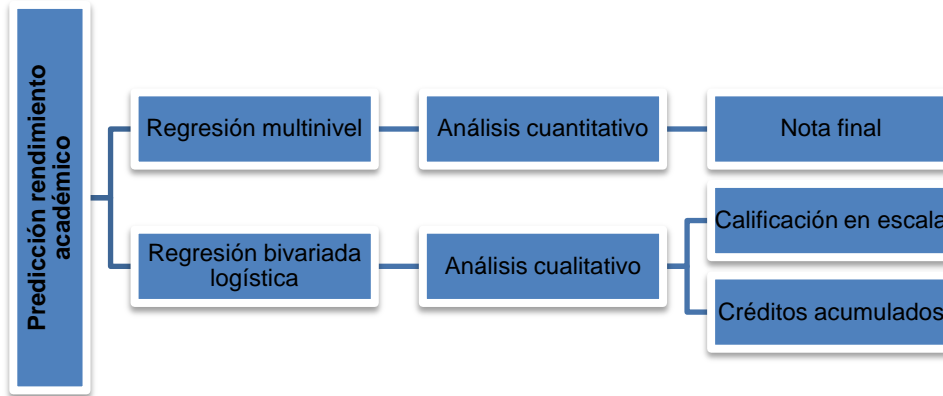
La muestra de estudio para la regresión logística bivariada comprende las carreras de 5 años, por lo tanto el periodo analizado estuvo en el rango abril 2009 - abril 2014. El software que se usó para las estimaciones *multinivel* fue Stata/SE 12.0<sup>4</sup> y para la *regresión bivariada* logística la herramienta RStudio de R Program con el paquete Zelig.

## 4.5 Técnicas multivariantes aplicadas

Dados los objetivos específicos y las características de los datos se aplican los dos modelos multivariantes de los que hemos venido hablando.

---

<sup>4</sup> Existen otros paquetes estadísticos que permiten ejecutar modelos multinivel: MLwiN, HLM, SPSS, R, SAS, S\_Plus, Mplus LISREL, BMDP 5V, AML, EGRET, GENSTAT, GLLAMM, MIXREG, SYSTAT, ASREML, BAYESX, LIMDEP, OSWALD y WinBUGS.

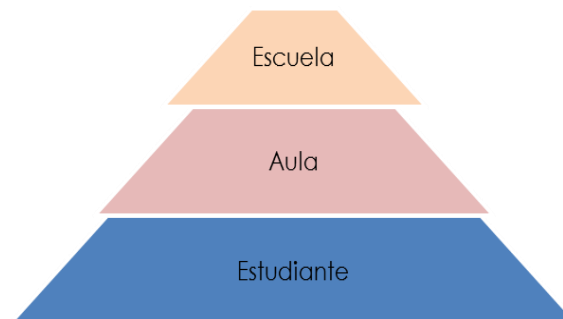


**Ilustración 6. Técnicas multivariantes aplicadas para la predicción del rendimiento académico**

Fuente: Elaboración propia

- a) **Modelo de regresión multinivel**, los datos que se usan para medir el rendimiento académico presentan una estructura anidada o jerárquica de estudiante, aula y escuela. Como se muestra en la Ilustración 7 se trató de un modelo de tres niveles, además se ha tenido presente la premisa de Boado (2013) quien menciona que ignorar la jerarquía de los datos implica asumir graves riesgos teóricos y técnicos, ya que obviar el anidamiento de casos puede llevar a falacias interpretativas de diferente naturaleza. Se decidió aplicar esta técnica para analizar correctamente el rendimiento académico desde diferentes niveles de agregación.

La variable dependiente representada por la nota final se mide en el nivel más bajo y las variables explicativas en todos los niveles.



**Ilustración 7. Estructura jerárquica del modelo**

Fuente: Elaboración propia

- b) **Modelo de regresión logístico bivariante**, lo consideramos como el mejor modelo para medir el rendimiento académico a través de dos variables dependientes cualitativas, las variables de respuesta (Grados y créditos acumulados) que, al ser variables dicotómicas y al estar correlacionadas, no permiten aplicar el modelo de regresión clásico.

Con la regresión logística bivariante se pretende relacionar las covariables con las variables de respuesta asociadas y, al mismo tiempo, la validez de las interacciones puede ser analizada en términos de logaritmos de Odds Ratio.

La asociación se modela entre las dos variables de respuesta (Grado y Créditos) en relación a varias covariables (Edad, género, región y nivel de participación en actividades en línea).

La correlación para la dinámica del rendimiento académico se puede identificar en términos de la relación de las variables de resultado Grado ( $Y_1$ ) y de los Créditos ( $Y_2$ ).

Se realizan dos tipos de asociaciones desde el enfoque del rendimiento académico con una respuesta individual con cuatro probabilidades posibles:

- 1) Asociación conjunta de probabilidades entre la variable Grado y Créditos que se analizan desde dos grupos, los estudiantes que tienen un Grado y créditos sobre la media y los que están debajo de ella.

		<b>Créditos (<math>Y_2</math>)</b>	
		Bajo la media (0)	Sobre la media (1)
<b>Grado (<math>Y_1</math>)</b>	Bajo la media (0)	$\pi_{00}$	$\pi_{01}$
	Sobre la media (1)	$\pi_{10}$	$\pi_{11}$

$\pi_{00}$ : Probabilidad de un estudiante con una calificación y un número de créditos acumulados bajo la media.

$\pi_{01}$ : Probabilidad de un estudiante con una calificación bajo la media que tiene un número de créditos acumulados igual o sobre la media.

$\pi_{10}$ : Probabilidad de un estudiante con una calificación igual o sobre la media que tiene un número de créditos acumulados bajo la media.

$\pi_{11}$ : Probabilidad de un estudiante con una calificación y un número de créditos acumulados igual o sobre la media.

- 2) Asociación conjunta de probabilidades entre la variable Grado y Créditos que representa una asociación entre los mejores estudiantes y los que no han tenido éxito.

		Créditos ( $Y_2$ )	
		No Éxito (0)	Éxito (1)
Grado ( $Y_1$ )	No Éxito (0)	$\pi_{00}$	$\pi_{01}$
	Éxito (1)	$\pi_{10}$	$\pi_{11}$

$\pi_{00}$ : Probabilidad de un estudiante que no ha tenido éxito en sus calificaciones (C -E) y en el número de créditos acumulados (< 220 créditos).

$\pi_{01}$ : Probabilidad de un estudiante que no ha tenido éxito en la calificación (C-E) pero que si ha tenido éxito en cuanto a los créditos acumulados (>220 créditos).

$\pi_{10}$ : Probabilidad de un estudiante que ha tenido éxito en la calificación (A o B) pero que no ha tenido éxito en la acumulación de créditos (< 220 créditos).

$\pi_{11}$ : Probabilidad de un estudiante que ha tenido éxito en la calificación (A o B) y ha superado el número de créditos que lo ubican como estudiante de último año ( $\geq 220$  créditos).

Con la regresión logística bivariante se proporciona la estimación simultánea de la probabilidad marginal del Grado, la probabilidad marginal de los Créditos y el log-odds ratio que describe la asociación conjunta de las probabilidades en la dinámica del rendimiento académico a nivel de respuesta individual.

# CAPÍTULO 5

## ANÁLISIS DE RESULTADOS DEL MODELO MULTINIVEL

---

El objetivo de este capítulo es explicar a través de una regresión multinivel la influencia de variables sociodemográficas, antecedentes académicos, enfoque learning analytics, a “nivel individual”, y las características de la asignatura y del docente, a “nivel contextual”, en el rendimiento académico.

Para lograr este objetivo se realiza un procedimiento “stepwise” hacia adelante, es decir incrementando el número de variables explicativas del nivel 1 y del nivel 2 para ir ampliando la capacidad de explicación y ajuste del modelo, aunque para ello esta secuencia metodológica aumente simultáneamente la complejidad del mismo.

Las estimaciones “stepwise” se desarrollan bajo una especificación lineal debido a las características de los datos y con el apoyo del software Stata/SE 12.0 a través de la funcionalidad `Statistic – Multinivel mixed-effects models`.

El análisis se inicia con el paso 0 en el cual no se incluyen variables explicativas, es decir, se estima un modelo nulo para comprobar la significatividad y luego explicar la varianza, expandiendo el modelo a través de la incorporación de predictores de los dos niveles en la parte fija y aleatoria. En el paso 1 se consideran únicamente predictores del nivel 2, con la finalidad de explicar la variabilidad a través de variables del nivel 2.



Para el paso 2 se ingresan predictores del nivel 1 y estos son los que explican la varianza del rendimiento académico dentro de los grupos. Luego ya en el paso 3 se consideran los resultados anteriores para generar una estimación basada en las variables explicativas de los estudiantes y de las aulas que son estadísticamente significativas y se realizan las interacciones multinivel a nivel del alumno con variables explicativas de las aulas.

Finalmente en el paso 4, a diferencia del paso 3, se incluye en la parte de efectos aleatorios las pendientes significativas del nivel 1.

## 5.1 Ajuste del modelo multinivel con 2 niveles

En este apartado los resultados que se presentan incluyen las estimaciones basadas en dos niveles de análisis, el estudiante (nivel 1) que pertenece a un aula (nivel 2).

En el nivel 1 se encuentran 23583 unidades de análisis (número de observaciones-estudiantes) y 468 unidades en el nivel 2 (número de grupos-aulas).

En el nivel 1 se consideran tres variables sociodemográficas, tres variables sobre los antecedentes académicos y ocho variables del enfoque learning analytics. Para el nivel 2 se incluyen nueve variables, de las cuales, cinco son características de las asignaturas y cuatro de los docentes. Conforme se han ido realizando las regresiones, algunas variables se han suprimido por no ser estadísticamente significativas.

### **Paso 0: Modelo nulo (ANOVA con efectos aleatorios)**

Objetivo : Determinar la variabilidad explicada en cada uno de los dos niveles.

Para el logro de este objetivo se inicia con un análisis ANOVA de efectos aleatorios, que permitirá conocer la parte del rendimiento académico que se debe a diferencias entre aulas y la parte debida a las diferencias entre los estudiantes.

Así el modelo considerado es:

$$R_{_A_{ij}} = \gamma_{00} + u_{0j} + e_{ij}$$

donde

$R_{_A_{ij}}$  es el rendimiento académico representado por el promedio de la nota final de la asignatura troncal de 0 a 40 puntos.

$\gamma_{00}$  es el parámetro que representa el efecto medio global del rendimiento académico.

$e_{ij}$  efecto aleatorio del nivel 1, que representa la variabilidad entre estudiantes

$u_{0j}$  efecto aleatorio del nivel 2, que representa la variabilidad entre aulas

Se supone que los términos aleatorios siguen una distribución normal, esto es

$$\left. \begin{array}{lll} e_{ij} \sim N(0, \sigma^2) & \forall (i, j) & \text{incorreladas entre sí} \\ u_{0j} \sim N(0, \tau^2) & \forall j & \text{incorreladas entre sí} \\ \text{Cov}(u_{0j}, e_{ij}) = 0 & \forall (i, j) & \end{array} \right\} \text{(A)}$$

Los resultados del ajuste del modelo se recogen en la Tabla 15.

**Tabla 15. Regresión modelo nulo 2 niveles**

```

Mixed-effects ML regression      Number of obs      =      23583
Group variable: AULA            Number of groups   =       468

                                Obs per group: min =        1
                                avg =       50.4
                                max =       470

                                Wald chi2(0)    =        .
                                Prob > chi2     =        .

Log likelihood = -83902.204

```

REND_ACADEMICO	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
_cons	25.79167	.2070581	124.56	0.000	25.38584 26.1975

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]
AULA: Identity var(_cons)	16.55076	1.27645	14.22887 19.25154
var(Residual)	69.07423	.6421699	67.827 70.3444

LR test vs. linear regression:  $\chi^2(01) = 4198.15$  Prob  $\geq \chi^2 = 0.0000$

**Fuente:** Elaboración propia con apoyo de Stata/SE 12.0

Promediando entre estudiantes y aulas, la media global del rendimiento académico es de 25.79 puntos y es estadísticamente significativa. Esto permite corroborar que la tasa de reprobación es alta, más aun al observar que ni el mínimo con el que aprueba un estudiante, que es 28 puntos, se encuentra dentro del intervalo de confianza.

Para analizar la significación de la variabilidad provocada entre las aulas, se realiza el test de razón de verosimilitudes para contrastar la hipótesis

$$H_0: \tau_2 = 0$$

En la última fila de la Tabla 15 se recoge los resultados de dicho contraste, a partir de los cuales se debe concluir la significación de dicho componente de la varianza al tener asociado un *p-value* menor de 0.001. Por lo tanto hay evidencia de variación del rendimiento académico entre aulas.

En las estimaciones de los componentes de la varianza correspondiente a los niveles 1 y 2 se obtiene la estimación de la varianza  $\sigma^2$  (para  $e_{ij}$ ) de 69.07 y la estimación de la varianza  $\tau_2$  (para  $u_{0j}$ ) de 16.55.

Con las estimaciones de las varianzas se calcula el coeficiente de correlación intraclase<sup>5</sup> el cual da como resultado 0.19, lo que significa que “el 19% de las variaciones en el rendimiento académico se debe a diferencias de las aulas y la mayor parte de la variación representada por el 81% restante es atribuible a diferencias entre los individuos”.

### **Paso 1: Explicación del intercepto con variables del nivel 2**

Objetivo: Explicar la variabilidad “intraclase” a través de predictores medidos sobre las unidades del nivel 2.

La especificación del modelo en

Nivel 1

$$R_{-}A_{ij} = \beta_{0j} + e_{ij}$$

Nivel 2

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Tasa\ repetidores_j + \gamma_{02}Ciclo_j + \gamma_{03}Tipo\ docente_j + u_{0j}$$

---

<sup>5</sup>  $\rho = \frac{\tau_2}{(\sigma^2 + \tau_2)}$  coeficiente de correlación intraclase que proporciona la variación del rendimiento académico que corresponde a la variación entre las unidades del nivel 2.

Por tanto, en términos generales se construye una sola ecuación que contiene ambos niveles

$$R_{-}A_{ij} = \gamma_{00} + \gamma_{01}Tasa\ repetidores_j + \gamma_{02}Ciclo_j + \gamma_{03}Tipo\ docente_j + u_{0j} + e_{ij}$$

donde

$\gamma_{00}, \gamma_{02}, \gamma_{03}$  son los coeficientes de los predictores del nivel 2

Se mantiene idéntica estructura de covarianzas que el modelo analizado en el paso 0 (A).

Antes de presentar el modelo final para este paso, se realiza una estimación conjunta (Ver Anexo 1) con la finalidad de analizar globalmente qué variables son significativas, luego se hace una eliminación progresiva de todos los predictores considerados en el nivel 2 para no descartar variables significativas, quedando el modelo final ajustado como se recoge en la Tabla 16.

A través del proceso de selección, se han desestimado las variables del nivel 2 (N° matriculados, tasa aprobación, N° créditos) y de los docentes (Experiencia, formación y evaluación) por no resultar sus parámetros significativos.

Por lo tanto todas las variables del modelo final son estadísticamente significativas. En esta estimación el rendimiento académico de las aulas considerando características grupales es de 25.74 puntos. La pendiente de la variable Tasa de repetidores, medida en el intervalo [0,1], nos indica que, un aumento de 10 puntos porcentuales de estudiantes matriculados que repiten por segunda o tercera vez una asignatura troncal provoca, en promedio, una disminución de 1.7 puntos en el rendimiento académico.

**Tabla 16. Estimación del intercepto con variables del nivel 2**

Mixed-effects ML regression	Number of obs =	23583
Group variable: AULA	Number of groups =	468
	Obs per group: min =	1
	avg =	50.4
	max =	470
Deviance = 167570.01	Wald chi2(4) =	318.00
Log likelihood = -83785.003	Prob > chi2 =	0.0000

REND_ACADEMICO	Coef.	Std. Err.	z	P>z	[95% Conf. Interval]
Tasa_Repetidores	-17.36811	1.551223	-11.20	0.000	-20.40845 -14.32777
Ciclo	.5565866	.0634977	8.77	0.000	.4321334 .6810399
<i>Tipo_docente</i>					
Tiempo completo	0	(base)			
Administrativo	1.750859	.602035	2.91	0.004	.5708919 2.930826
Invitado	1.024808	.3258507	3.15	0.002	.3861521 1.663463
_cons	25.74286	.5212646	49.39	0.000	24.7212 26.76452

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]
AULA: Identity			
var(_cons)	8.490554	.74102	7.155611 1.007454
var(Residual)	69.10242	.6424853	67.85458 70.37321

LR test vs. linear regression: chibar2(01) = 1881.28 Prob >= chibar2 = 0.0000

**Fuente:** Elaboración propia con apoyo de Stata/SE 12.0

La influencia generada por el ciclo es positiva, esto implica que la tendencia del puntaje del rendimiento académico del aula suba aproximadamente 0.5 décimas si se avanza un ciclo.

La estimación de los coeficientes asociados a la variable tipo de docente, muestra que, las aulas que cuentan con docentes administrativos, en promedio, aumentan el rendimiento académico del aula en 1.75 puntos más que las aulas que tienen docentes de tiempo completo. Ocurre lo mismo con los docentes invitados, ya que, el resultado indica que las aulas que cuentan con docentes invitados, en promedio, aumentan el rendimiento académico del aula en 1 punto más que las aulas que tienen docentes de tiempo completo.

La varianza explicada<sup>6</sup> por los tres predictores que conforman el nivel 2 es del 49%. Además contrastando la  $H_0: \tau_2 = 0$ , es decir, la no existencia de variación del rendimiento académico entre aulas, se observa que el *p-value* reportado es mayor que 0.001 (vease parte final de la Tabla 16), por lo tanto, se rechaza la hipótesis nula, lo que significa que aún existe una variación significativa del rendimiento académico entre aulas por ser explicada, por lo que se continua con el siguiente paso en el que se incluyen predictores del nivel 1.

## **Paso 2: Significación de las variables explicativas del nivel 1**

Objetivo: Determinar la variabilidad del rendimiento explicada por predictores definidos para el nivel 1.

---

<sup>6</sup>  $p = \frac{\tau_2(\text{Paso 0}) - \tau_2(\text{Paso 1})}{\tau_2(\text{Paso 0})}$  es la proporción de la varianza explicada en el nivel 2.

En este caso las ecuaciones independientes para cada nivel considerando los diez predictores del nivel 1 son

Nivel 1

$$R_{-}A_{ij} = \beta_{0j} + \beta_{1j} Edad_{ij} + \beta_{2j} Rinde\ supletorio_{ij} + \beta_{3j} Re\ pite\ materia_{ij} \\ + \beta_{4j} Beca_{ij} + \beta_{5j} Participa\ chat_{ij} + \beta_{6j} Participa\ foro_{ij} + \beta_{7j} Participa\ video_{ij} \\ + \beta_{8j} N^{\circ}comentarios_{ij} + \beta_{9j} N^{\circ}accesos\ LMS_{ij} + \beta_{10j} N^{\circ}mensajes_{ij} + e_{ij}$$

Nivel 2

$$\beta_{0j} = \gamma_{00} + u_{0j} \\ \vdots \\ \beta_{kj} = \gamma_{k0} \quad \forall k = 1, \dots, 10$$

Es decir, se considera un modelo de regresión sobre las variables del nivel 1 (estudiantes) con interceptos aleatorios provocados por el efecto aula. Con ello se pretende seleccionar las variables del nivel 1 significativas, es decir, con capacidad de explicación de la variable objetivo.

Para convertir las dos ecuaciones en una sola se sustituye la segunda ecuación en la primera y se reordena los términos, quedando como sigue

$$R_{-}A_{ij} = \gamma_{00} + \gamma_{10} Edad_{ij} + \gamma_{20} Rinde\ supletorio_{ij} + \gamma_{30} Re\ pite\ materia_{ij} \\ + \gamma_{40} Beca_{ij} + \gamma_{50} Participa\ chat_{ij} + \gamma_{60} Participa\ foro_{ij} + \gamma_{70} Participa\ video_{ij} \\ + \gamma_{80} N^{\circ}comentarios_{ij} + \gamma_{90} N^{\circ}accesos\ LMS_{ij} + \gamma_{100} N^{\circ}mensajes_{ij} + u_{0j} + e_{ij}$$

donde

$k = 1, \dots, 10$ ,  $\gamma_{k0}$  representa los coeficientes de los predictores del nivel 1.



$\gamma_{00}$  es el coeficiente intercepto fijo o efecto medio global

Se mantiene la misma estructura de covarianza como en la referencia (A).

Previamente a analizar el modelo compuesto por todas las variables consideradas para el nivel 1, se realizan cuatro modelos que contemplan las dimensiones de las variables de los estudiantes por separado; el modelo 1 (variables sociodemográficas), el modelo 2 (antecedentes académicos) el modelo 3 (variables learning analytics) y en el modelo 4 (estimación grupal). Los resultados y el análisis se encuentran en el Anexo 2. La estimación secuencial de estos modelos permite determinar las variables del nivel 1 que son significativas. Tras este procedimiento de selección, el modelo final ajustado se recoge en la Tabla 17.

El “*Random intercept model*” final no incluye en su estimación las variables sociodemográficas (género y región) y las variables del enfoque learning analytics (N° accesos\_curso y tiempo LMS) por no ser estadísticamente significativas.

Los resultados muestran que la edad tiene una relación positiva con el rendimiento académico, aunque muy leve dado que el modelo estima que el aumento de 1 año de edad conduce a un aumento medio de 0.08 décimas en la variable objetivo. La estimación del coeficiente de la rendición de supletorio indica que los estudiantes que se quedan suspensos, en promedio, provocan una disminución de 1.46 puntos en el rendimiento académico. En cambio pasa lo contrario con los estudiantes que están repitiendo la materia, ya que el hecho de que tengan un antecedente de conocerla ocasiona un impacto positivo, es decir, que en promedio el rendimiento académico entre aulas sube en 2.04 puntos.

**Tabla 17. Estimación multinivel con variables del nivel 1**

Mixed-effects ML regression	Number of obs	=	23583
Group variable: AULA	Number of groups	=	468
	Obs per group: min	=	1
	Avg	=	50.4
	Max	=	470
Deviance = 162375.52	Wald chi2(10)	=	6104.61
Log likelihood = -81187.759	Prob > chi2	=	0.0000

REND_ACADEMICO	Coef.	Std. Err.	Z	P>z	[95% Conf. Interval]
Edad	.0866052	.0063238	13.70	0.000	.0742108 .0989996
Rinde_supletorio					
No	0	(base)			
Si	-1.457412	.1009453	-14.44	0.000	-1.655261 -1.259563
Repite_materia					
No	0	(base)			
Si	2.037929	.1263166	16.13	0.000	1.790353 2.285505
Beca					
No	0	(base)			
Si	3.492708	1.495816	2.33	0.020	.5609622 6.424455
N_comentarios	.0826495	.0225055	3.67	0.000	.0385395 .1267596
N_mensajes	.3068819	.151725	2.02	0.043	.0095064 .6042574
N_accesos_LMS	.043388	.0008472	51.21	0.000	.0417276 .0450484
Participa_chat	1.360599	.1623234	8.38	0.000	1.042451 1.678747
Participa_foro					
No	0	(base)			
Si	2.111647	.1209971	17.45	0.000	1.874497 2.348797
Participa_video					
No	0	(base)			
Si	1.370657	.1857309	7.38	0.000	1.006631 1.734683
_cons	19.40119	.274735	70.62	0.000	18.86272 19.93966

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]
AULA: Identity			
var(_cons)	13.09532	1.03324	11.21902 15.28541
var(Residual)	54.87425	.5103762	53.883 55.88374

LR test vs. linear regression: chibar2(01) = 3544.17 Prob >= chibar2 = 0.0000

**Fuente:** Elaboración propia con apoyo de Stata/SE 12.0

La estimación de los coeficientes asociados a las variables de la dimensión *learning analytics* son positivos y significativos, es decir que aquellos estudiantes que han participado en los foros, chat y video colaboración, en promedio, incrementan entre 1 y 2 puntos el rendimiento académico.

Respecto a la significación de la variabilidad provocada entre las aulas, es decir, por las unidades del nivel 2 ( $H_0: \tau_2 = 0$ ), el test LR alcanza un valor de 3544.17 con un *p-value* menor que 0.001, por lo que se puede afirmar la significancia de dicha variabilidad, esto significa, que hay valores distintos del intercepto dependiendo de las unidades del Nivel 2.

En cuanto a la varianza explicada<sup>7</sup> por los predictores del nivel 1, se compara la varianza de esta estimación con la del modelo nulo, de modo que, la varianza explicada por los diez predictores del nivel 1 representa el 21% de la variación en el rendimiento académico.

### **Paso 3: Regresión con interacciones y variables de los niveles 1 y 2**

Objetivo: Estimar las relaciones a nivel del alumno con variables explicativas del nivel 2

Una vez detectada la significación de las variables del nivel 1, es decir, la capacidad de explicación del rendimiento académico, y la significación de la variabilidad de las unidades del Nivel 2 en el comportamiento medio (Intercepto) de la variable objetivo, se plantea ahora el estudio de la relación entre las variables del nivel 1 y la variable objetivo, dependiendo de variables del Nivel 2. Se incluyen posibles interacciones de variables del Nivel 1 con variables del Nivel 2, pero solo se consideran en este caso como efectos fijos.

---

<sup>7</sup>  $p = \frac{\sigma^2(\text{Paso 0}) - \sigma^2(\text{Paso 2})}{\sigma^2(\text{Paso 0})}$  es la proporción de la varianza que en el modelo Random intercept model explica a través de las variables del nivel 1.

En el paso siguiente se ampliará el modelo con pendientes aleatorias. Así el modelo en este paso es:

Nivel 1

$$R_{-}A_{ij} = \beta_{0j} + \beta_{1j} \text{Edad}_{ij} + \beta_{2j} \text{Rinde sup letorio}_{ij} + \beta_{3j} \text{Re pite materia}_{ij} + \\ \beta_{4j} \text{Participa chat}_{ij} + \beta_{5j} \text{Participa foro}_{ij} + \beta_{6j} \text{Participa video}_{ij} + \\ \beta_{7j} \text{N}^{\circ} \text{comentarios}_{ij} + \beta_{8j} \text{N}^{\circ} \text{accesos LMS}_{ij} + e_{ij}$$

Nivel 2

$$\beta_{0j} = \gamma_{00} + \gamma_{01} \text{Tasa repetidores}_j + \gamma_{02} \text{Ciclo}_j + \gamma_{03} \text{Tipo docente}_j + u_{0j} \\ \beta_{1j} = \gamma_{10} + \gamma_{11} \text{Tasa repetidores}_j + \gamma_{12} \text{Ciclo}_j + \gamma_{13} \text{Tipo docente}_j \\ \beta_{2j} = \gamma_{20} + \gamma_{21} \text{Tasa repetidores}_j + \gamma_{22} \text{Ciclo}_j + \gamma_{23} \text{Tipo docente}_j \\ \beta_{3j} = \gamma_{30} + \gamma_{31} \text{Tasa repetidores}_j + \gamma_{32} \text{Ciclo}_j + \gamma_{33} \text{Tipo docente}_j \\ \beta_{4j} = \gamma_{40} + \gamma_{41} \text{Tasa repetidores}_j + \gamma_{42} \text{Ciclo}_j + \gamma_{43} \text{Tipo docente}_j \\ \beta_{5j} = \gamma_{50} + \gamma_{51} \text{Tasa repetidores}_j + \gamma_{52} \text{Ciclo}_j + \gamma_{53} \text{Tipo docente}_j \\ \beta_{6j} = \gamma_{60} + \gamma_{61} \text{Tasa repetidores}_j + \gamma_{62} \text{Ciclo}_j + \gamma_{63} \text{Tipo docente}_j \\ \beta_{7j} = \gamma_{70} + \gamma_{71} \text{Tasa repetidores}_j + \gamma_{72} \text{Ciclo}_j + \gamma_{73} \text{Tipo docente}_j \\ \beta_{8j} = \gamma_{80} + \gamma_{81} \text{Tasa repetidores}_j + \gamma_{82} \text{Ciclo}_j + \gamma_{83} \text{Tipo docente}_j$$

Al sustituir las 9 últimas ecuaciones en la primera, se obtiene la ecuación final reducida

$$\begin{aligned}
 R_{-}A_{ij} = & \gamma_{00} + \gamma_{01} \textit{Tasa repetidores}_j + \gamma_{02} \textit{Ciclo}_j + \gamma_{03} \textit{Tipo docente}_j \\
 & + \gamma_{10} \textit{Edad}_{ij} + \gamma_{11} \textit{Tasa repetidores}_j * \textit{Edad}_{ij} + \gamma_{12} \textit{Ciclo}_j * \textit{Edad}_{ij} \\
 & + \gamma_{13} \textit{Tipo docente}_j * \textit{Edad}_{ij} + \gamma_{20} \textit{Rinde supletorio}_{ij} \\
 & + \gamma_{21} \textit{Tasa repetidores}_j * \textit{Rinde supletorio}_{ij} + \gamma_{22} \textit{Ciclo}_j * \textit{Rinde supletorio}_{ij} \\
 & + \gamma_{23} \textit{Tipo docente}_j * \textit{Rinde supletorio}_{ij} + \gamma_{30} \textit{Re pite materia}_{ij} \\
 & + \gamma_{31} \textit{Tasa repetidores}_j * \textit{Re pite materia}_{ij} + \gamma_{32} \textit{Ciclo}_j * \textit{Re pite materia}_{ij} \\
 & + \gamma_{33} \textit{Tipo docente}_j * \textit{Re pite materia}_{ij} + \gamma_{40} \textit{Participa chat}_{ij} \\
 & + \gamma_{41} \textit{Tasa repetidores}_j * \textit{Participa chat}_{ij} + \gamma_{42} \textit{Ciclo}_j * \textit{Participa chat}_{ij} \\
 & + \gamma_{43} \textit{Tipo docente}_j * \textit{Participa chat}_{ij} + \gamma_{50} \textit{Participa foro}_{ij} \\
 & + \gamma_{51} \textit{Tasa repetidores}_j * \textit{Participa foro}_{ij} + \gamma_{52} \textit{Ciclo}_j * \textit{Participa foro}_{ij} \\
 & + \gamma_{53} \textit{Tipo docente}_j * \textit{Participa foro}_{ij} + \gamma_{60} \textit{Participa video}_{ij} \\
 & + \gamma_{61} \textit{Tasa repetidores}_j * \textit{Participa video}_{ij} + \gamma_{62} \textit{Ciclo}_j * \textit{Participa video}_{ij} \\
 & + \gamma_{63} \textit{Tipo docente}_j * \textit{Participa video}_{ij} + \gamma_{70} \textit{N}^\circ \textit{comentarios}_{ij} \\
 & + \gamma_{71} \textit{Tasa repetidores}_j * \textit{N}^\circ \textit{comentarios}_{ij} + \gamma_{72} \textit{Ciclo}_j * \textit{N}^\circ \textit{comentarios}_{ij} \\
 & + \gamma_{73} \textit{Tipo docente}_j * \textit{N}^\circ \textit{comentarios}_{ij} + \gamma_{80} \textit{N}^\circ \textit{accesos LMS}_{ij} \\
 & + \gamma_{81} \textit{Tasa repetidores}_j * \textit{N}^\circ \textit{accesos LMS}_{ij} + \gamma_{82} \textit{Ciclo}_j * \textit{N}^\circ \textit{accesos LMS}_{ij} \\
 & + \gamma_{83} \textit{Tipo docente}_j * \textit{N}^\circ \textit{accesos LMS}_{ij} + u_{0j} + e_{ij}
 \end{aligned}$$

La estructura de la covarianza en este paso coincide con la referencia (A).

Al ajustar el modelo, los resultados que se obtienen se presentan en la Tabla 18, en las que se han eliminado todas aquellas interacciones no significativas.

La adición de las covariables de ambos niveles ha reducido el tamaño de la componente de la varianza del nivel 2, que ahora es 6.69 en comparación al modelo del paso 2 que fue de 13.10. No obstante, la componente sigue siendo estadísticamente significativa.

**Tabla 18. Estimación de la regresión con interacciones y variables de los niveles 1 y 2**

Mixed-effects ML regression	Number of obs	=	23583
Deviance = 161791.11	Wald chi2(16)	=	6952.46
Log likelihood = -80895.555	Prob > chi2	=	0.0000

REND_ACADEMICO	Coef.	Std. Err.	Z	P>z	[95% Conf. Interval]
Tasa_Repetidores	-20.88847	1.505409	-13.88	0.000	-23.83902 -17.93792
Ciclo	.7806166	.0666854	11.71	0.000	.6499157 .9113176
Tipo_docente					
Tiempo completo	0	(base)			
Administrativo	1.357542	.534204	2.54	0.011	.3105209 2.404562
Invitado	.8822802	.2893487	3.05	0.002	.3151672 1.449393
Edad	.0839426	.0062739	13.38	0.000	.0716461 .0962391
Rinde_supletorio	-.4802798	.1629353	-2.95	0.003	-.7996271 -.1609325
Rinde_supletorio*Ciclo	-.3591831	.0432867	-8.30	0.000	-.4440235 -.2743426
Repite_materia	2.968676	.2122925	13.98	0.000	2.55259 3.384761
Repite_materia*Ciclo	-.2754914	.0579427	-4.75	0.000	-.389057 -.1619258
Participa_chat	1.289819	.1599049	8.07	0.000	.976411 1.603227
Participa_foro	2.089442	.1195308	17.48	0.000	1.855166 2.323718
Participa_video	1.274524	.1819438	7.01	0.000	.9179206 1.631127
N_comentarios	.0864688	.0222376	3.89	0.000	.0428839 .1300538
N_accesos_LMS	.044745	.0024998	17.90	0.000	.0398455 .0496444
N_accesos_LMS*Tasa_Repetidores	.0685623	.0085887	7.98	0.000	.0517288 .0853957
N_accesos_LMS*Ciclo	-.0043281	.0003332	-12.99	0.000	-.0049811 -.0036751
_cons	20.17987	.5269296	38.30	0.000	19.1471 21.21263

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]
AULA: Identity			
var(_cons)	6.692232	.5897355	5.630683 7.953914
var(Residual)	54.07873	.5029103	53.10197 55.07345

LR test vs. linear regression: chibar2(01) = 1824.48 Prob >= chibar2 = 0.0000

**Fuente:** Elaboración propia con apoyo de Stata/SE 12.0

Las interacciones multinivel entre las variables del aula y del estudiante que resultan significativas, ejercen una influencia leve, dado el valor de las estimaciones de los parámetros correspondientes.

En relación a las variables del Nivel 2, se puede señalar que la variable tasa de repetidores resulta negativa para el rendimiento académico, ligeramente matizada por su interacción con el N° de accesos al LMS. La variable ciclo ejerce un efecto positivo, aunque algo matizado por su interacción con la variables rinde supletorio, repite materia y número de accesos al LMS. Finalmente la variable tipo de docente resulta también significativa dado que existe una influencia positiva de los docentes que no tienen dedicación a tiempo completo respecto de estos.

En relación a las variables del Nivel 1 cabe indicar que la edad, repetir materia y rendir supletorio tienen un efecto ligeramente positivo y ligeramente matizado por algunas interacciones con variables del Nivel 2, así como todas las variables incluidas en la dimensión de *learning analytics*.

La varianza explicada<sup>8</sup> es del 60% si se compara con las varianzas de las estimaciones anteriores esta es sumamente alta, valor justificado por la inclusión de predictores de los 2 niveles.

#### **Paso 4: Variabilidad en los coeficientes de los predictores del nivel 1**

Objetivo: Determinar si las pendientes de las variables explicativas del nivel 1 muestran componentes de variación significativos entre aulas.

---

<sup>8</sup>  $p = \frac{\tau_2(\text{Paso 0}) - \tau_2(\text{Paso 3})}{\tau_2(\text{Paso 0})}$  es la proporción de la varianza explicada por la estimación de la regresión con interacciones y variables de los Niveles 1 y 2.

Adicional a lo que se realizó en el paso anterior, se incluyen en el modelo las pendientes del nivel 1 en la parte de efectos aleatorios. Se realiza una estimación secuencial y a través del AIC<sup>9</sup> (Akaike information criterion) y la razón de verosimilitudes se determina cuales son estadísticamente significativas (Ver resultados en Anexo 3).

La diferencia con las ecuaciones del paso 4 es que ahora se integran las pendientes en los efectos aleatorios, por lo que las ecuaciones independientes quedan como sigue

Nivel 1

$$R_{-}A_{ij} = \beta_{0j} + \beta_{1j}Edad_{ij} + \beta_{2j}Rinde\ supletorio_{ij} + \beta_{3j}Re\ pite\ materia_{ij} + \beta_{4j}Participa\ chat_{ij} + \beta_{5j}Participa\ foro_{ij} + \beta_{6j}Participa\ video_{ij} + \beta_{7j}N^{\circ}comentarios_{ij} + \beta_{8j}N^{\circ}accesos\ LMS_{ij} + e_{ij}$$

Nivel 2

$$\begin{aligned}\beta_{0j} &= \gamma_{00} + \gamma_{01}Tasa\ repetidores_j + \gamma_{02}Ciclo_j + \gamma_{03}Tipo\ docente_j + u_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}Tasa\ repetidores_j + \gamma_{12}Ciclo_j + \gamma_{13}Tipo\ docente_j + u_{1j} \\ \beta_{2j} &= \gamma_{20} + \gamma_{21}Tasa\ repetidores_j + \gamma_{22}Ciclo_j + \gamma_{23}Tipo\ docente_j + u_{2j} \\ \beta_{3j} &= \gamma_{30} + \gamma_{31}Tasa\ repetidores_j + \gamma_{32}Ciclo_j + \gamma_{33}Tipo\ docente_j + u_{3j} \\ \beta_{4j} &= \gamma_{40} + \gamma_{41}Tasa\ repetidores_j + \gamma_{42}Ciclo_j + \gamma_{43}Tipo\ docente_j + u_{4j} \\ \beta_{5j} &= \gamma_{50} + \gamma_{51}Tasa\ repetidores_j + \gamma_{52}Ciclo_j + \gamma_{53}Tipo\ docente_j + u_{5j} \\ \beta_{6j} &= \gamma_{60} + \gamma_{61}Tasa\ repetidores_j + \gamma_{62}Ciclo_j + \gamma_{63}Tipo\ docente_j + u_{6j} \\ \beta_{7j} &= \gamma_{70} + \gamma_{71}Tasa\ repetidores_j + \gamma_{72}Ciclo_j + \gamma_{73}Tipo\ docente_j + u_{7j} \\ \beta_{8j} &= \gamma_{80} + \gamma_{81}Tasa\ repetidores_j + \gamma_{82}Ciclo_j + \gamma_{83}Tipo\ docente_j + u_{8j}\end{aligned}$$

---

<sup>9</sup> AIC = 2k-2ln(L) estima la calidad de cada modelo, en relación con cada uno de los otros modelos. Por lo tanto, AIC proporciona un medio para la selección del modelo.



En este modelo las estructura de varianzas y covarianzas es:

$$e_{ij} \sim N(0, \sigma^2) \quad \forall (i, j) \text{ incorreladas entre sí}$$

Para  $k = 0, 1, 2, \dots, 8$ :

$$u_{kj} \sim N(0, \tau_k^2) \quad \forall j \text{ incorreladas entre sí}$$

$$\text{Cov}(u_{kj}, e_{ij}) = 0 \quad \forall (i, j)$$

Para  $k, h = 0, 1, 2, \dots, 8$ :

$$\text{Cov}(u_{kj}, u_{hj}) = \Sigma_{kj} \quad \forall j$$

Integrando las ecuaciones de los niveles 1 y 2 y realizando la regresión, los resultados que se obtienen se exponen en la Tabla 19.

Los resultados muestran que después de incluir las interacciones, el componente de la varianza de las pendientes de las variables explicativas del nivel 1 muestra una variación leve pero significativa entre aulas.

La varianza explicada<sup>10</sup> incluyendo las pendientes del nivel 1 en los efectos aleatorios es del 55%.

---

<sup>10</sup>  $p = \frac{\tau_2(\text{Paso 0}) - \tau_2(\text{Paso 4})}{\tau_2(\text{Paso 0})}$  es la proporción de la varianza explicada por la estimación de la regresión con

interacciones, variables de los Niveles 1 y 2 y las pendientes del nivel 1 en los efectos aleatorios.

**Tabla 19. Estimación de la variabilidad en los coeficientes de los predictores del nivel 1**

Mixed-effects ML regression                      Number of obs =                      23583  
 Deviance = 161611.82                      Wald chi2(16) =                      6146.69  
 Log likelihood = -80805.908                      Prob > chi2 =                      0.0000

REND_ACADEMICO	Coef.	Std. Err.	Z	P>z	[95% Conf. Interval]
Tasa_Repetidores	-21.2218	1.609523	-13.19	0.000	-24.3764 -18.06719
Ciclo	.7623473	.0697031	10.94	0.000	.6257318 .8989628
Tipo_docente					
Tiempo completo	0	(base)			
Administrativo	1.510935	.5678153	2.66	0.008	.3980378 2.623833
Invitado	.9573145	.3116882	3.07	0.002	.3464169 1.568212
Edad	.0837473	.0062392	13.42	0.000	.0715186 .095976
Rinde_supletorio	-.5267685	.2636077	-2.00	0.046	-1.04343 -.0101069
Rinde_supletorio*Ciclo	-.3366908	.0581579	-5.79	0.000	-.4506782 -.2227034
Repite_materia	2.808255	.2476235	11.34	0.000	2.322922 3.293588
Repite_materia*Ciclo	-.2486661	.0633913	-3.92	0.000	-.3729107 -.1244215
Participa_chat	1.313279	.1692679	7.76	0.000	.98152 1.645038
Participa_foro	2.057453	.1272299	16.17	0.000	1.808087 2.306819
Participa_video	1.31303	.1937892	6.78	0.000	.9332105 1.69285
N_comentarios	.091933	.0222558	4.13	0.000	.0483125 .1355535
N_accesos_LMS	.0438462	.0025096	17.47	0.000	.0389275 .048765
N_accesos_LMS*Tasa_Repetidores	.0696397	.0086301	8.07	0.000	.052725 .0865544
N_accesos_LMS*Ciclo	-.0042219	.0003348	-12.61	0.000	-.004878 -.0035658
_cons	2.025.424	.5565587	36.39	0.000	191.634 2.134.507

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]
AULA: Independent			
var(Rinde_supletorio)	3.243811	.4901397	2.412.343 4.361863
var(Repite_materia)	.8369861	.4001639	.3279125 2.13638
var(Partica_chat)	.4351275	.3708907	.0818607 2.312903
var(Participa_foro)	.3788677	.2529417	.1023773 1.402075
var(Participa_video)	.5905212	.4444584	.135075 2.581642
var(_cons)	7.385529	.6798355	6.166359 8.845746
var(Residual)	52.75249	.5014936	51.77868 53.74461

LR test vs. linear regression:    chi2(6) = 2003.78    Prob > chi2 = 0.00

**Fuente:** Elaboración propia con apoyo de Stata/SE 12.0

### 5.1.1 Ajuste global

Finalmente, además de analizar la aportación de los predictores que resultaron significativos, se determina cuál de los modelos aporta mayor información significativa en comparación con el modelo nulo (paso 0).

**Tabla 20. Comprobación de ajuste de 2 niveles**

Paso	Log likelihood	Deviance	p-value	Varianza explicada %	AIC
0	-83902.204	167804.41	0.0000	19	167810.4
1	-83785.003	167570.01	0.0000	49	167584.0
2	-81187.759	162375.52	0.0000	21	162401.5
3	-80895.555	161791.11	0.0000	59	161829.1
4	-80805.908	161611.82	0.0000	55	161659.8

**Fuente:** Elaboración propia con apoyo de Stata/SE 12.0

A través de la diferencia de los estadísticos de desviación (*deviances*<sup>11</sup>) se detecta si uno de los modelos de la Tabla 20, mejora la estimación respecto al modelo nulo.

Analizando todas las estimaciones para dos niveles (estudiantes y aulas), las estimaciones que explican un mayor porcentaje de la varianza son las del paso 3 y 4, sin embargo, la estimación que mejor se ajusta es la del paso 4, por lo que es este modelo el que se considera como modelo final definitivo para dar respuesta al objetivo prefijado.

+

---

<sup>11</sup>  $D_1 = -2 \log \text{likelihood} (L_1)$ , donde  $D_1$  es el estadístico desviación del modelo nulo y  $L_1$  es el valor máximo del estimador de verosimilitud bajo la hipótesis nula (modelo más restringido). Para la hipótesis alternativa, se calcula un  $D_2$  para el modelo 2 y así sucesivamente para el resto de modelos; luego se toma la diferencia entre las desviaciones ( $D_1 - D_2$ ) para comparar los modelos, el cual se distribuye según una ley  $\chi^2$  con los g.l dados por la diferencia de parámetros estimados en uno y otro modelo.

## 5.2 Ajuste del modelo multinivel con 3 niveles

En este apartado los resultados que se presentan incluyen las estimaciones basadas en tres niveles de análisis, el nivel 1 los estudiantes, el nivel 2 las aulas y el nivel 3 las escuelas. Al existir una estructura jerárquica, el nivel de los estudiantes se halla dentro del nivel de las aulas que, a su vez, se encuentra dentro del nivel de las escuelas.

Las unidades y variables consideradas para el nivel 1 y 2 son las mismas mencionadas en el apartado 5.1, se añade el nivel 3 que comprende 19 unidades de análisis (número de escuelas). Para este nivel se incluye: el área académica, el tamaño medido por número de paralelos, la tasa de reprobación global y el ratio alumno profesor.

Lo que se busca es determinar cómo las escuelas, al ser el ente superior en esta jerarquía, afectan a la variación o medición del rendimiento académico y sobre todo si las nuevas estimaciones se ajustan mejor que las anteriores.

### **Paso 0: Modelo nulo (ANOVA con efectos aleatorios)**

Objetivo: Conocer la variabilidad explicada por cada nivel de la jerarquía.

Se estima los resultados a través del modelo nulo, con efectos aleatorios de tres niveles que se recoge en la Tabla 21.

**Tabla 21. Estimación ANOVA con efectos aleatorios de 3 niveles**

Mixed-effects ML regression Number of obs = 23583

Group Variable	No. of Groups	Observations per Group		
		Minimum	Average	Maximum
ESCUELA	19	127	1241.2	5074
AULA	468	1	50.4	470

Log likelihood = -83895.04 Wald chi2(0) = .  
Prob > chi2 = .

REND_ACADEMICO	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
_cons	25.74426	.3294056	78.15	0.000	25.09863	26.38988

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
ESCUELA: Identity var(_cons)	1.193631	.6157485	.4342783	3.280744
AULA: Identity var(_cons)	15.10706	1.213427	12.90654	17.68276
var(Residual)	69.08471	.642341	67.83714	70.35521

LR test vs. linear regression: chi2(2) = 4212.48 Prob > chi2 = 0.0000

**Fuente:** Elaboración propia con apoyo de Stata/SE 12.0

La parte fija del modelo muestra que 25.74 es el rendimiento académico medio de las escuelas, esta media es similar a la obtenida en el análisis de 2 niveles (5.1.1). En el cuadro de efectos aleatorios hay tres estimaciones, cada una cuantifica la desviación de la media en cada nivel de la jerarquía. Los elementos que se obtienen son la estimación de la varianza del nivel 1,  $\sigma^2$  (69.085), la estimación de la varianza del nivel 2,  $\tau_2$  (15.107) y una estimación para la varianza del nivel 3,  $\tau_3$  (1.194). Además el test LR indica que todas las varianzas son significativamente no nulas, esto significa que, los estudiantes, las aulas y las escuelas provocan variabilidad significativa en el rendimiento académico alcanzado.

Por otro lado, para determinar si la mayor parte de la variación se da a nivel del estudiante, de las aulas o de las escuelas se estiman los coeficientes de correlación intra-clase:

Proporción de la varianza dentro de las aulas nivel del estudiante

$$\rho_1 = \frac{\sigma^2}{\sigma^2 + \tau_2 + \tau_3} = 0.809$$

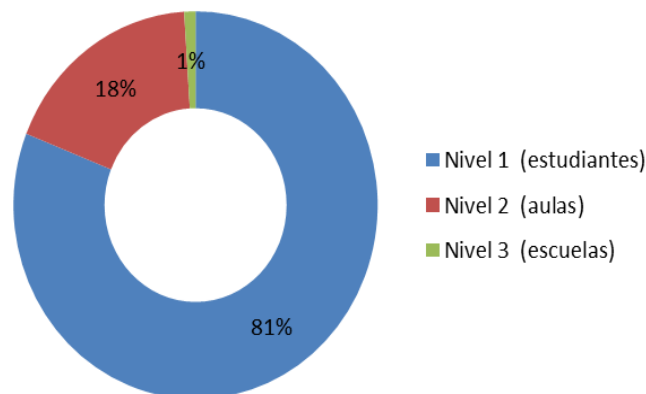
Proporción de la varianza entre las aulas dentro de las escuelas

$$\rho_1 = \frac{\tau_2}{\sigma^2 + \tau_2 + \tau_3} = 0.177$$

Proporción de la varianza entre escuelas

$$\rho_1 = \frac{\tau_3}{\sigma^2 + \tau_2 + \tau_3} = 0.014$$

Como se observa en la Ilustración 8 el porcentaje de variación del rendimiento académico disminuye a medida que se sube a los niveles jerárquicos superiores, es decir la mayor parte de la variación se encuentra a nivel del estudiante (dentro de las aulas) y la menor proporción de la varianza se encuentra entre escuelas.



**Ilustración 8. Varianza explicada por nivel**

Fuente: Elaboración propia

Dados los resultados obtenidos, ante la escasa influencia del nivel 3 sobre el rendimiento académico, y con objeto de no añadir complejidad innecesaria en el modelo, se mantiene el modelo final obtenido previamente con 2 niveles. Y, en consecuencia, las conclusiones que a través del mismo se han obtenido.

# CAPÍTULO 6

## ANÁLISIS DE RESULTADOS DEL MODELO LOGÍSTICO BIVARIANTE

---

El objetivo general de este capítulo es analizar el rendimiento académico universitario a través de dos indicadores, el grado o calificación académica y los créditos universitarios acumulados. El análisis se desarrolla usando una muestra de 410 estudiantes matriculados en una carrera de 5 años equivalente a 282 créditos, cuyo tiempo de estudio comprende el periodo abril 2009- abril 2014. Se asume que al finalizar este periodo debieron terminar su fase de estudio.

Las variables o indicadores de rendimiento académico universitario son:

- *Grado o calificación académica*, que corresponde a la calificación media obtenida en las asignaturas superadas, clasificadas según la escala establecida por el Sistema Europeo de Transferencia de Créditos (ECTS, en sus siglas en inglés):

A: Sobresaliente  
B: Muy buena  
C: Buena  
D: Satisfactorio  
E: Suficiente

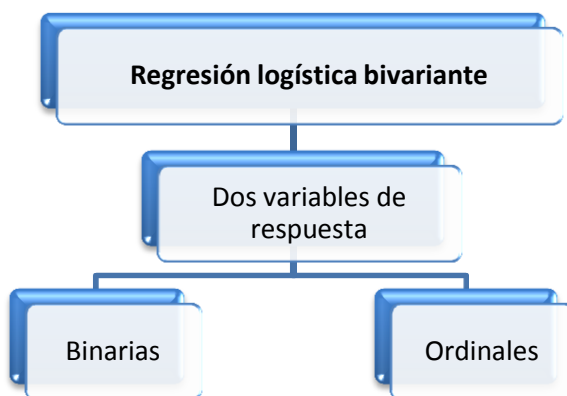


- *Créditos acumulados* que corresponden a la suma de créditos de las asignaturas superadas por el estudiante en el periodo de tiempo de duración de una carrera, clasificados según la escala establecida por la universidad:

1:	menos de 50 créditos
2:	de 50 a 109 créditos
3:	de 110 a 159 créditos
4:	de 160 a 219 créditos
5:	de 220 a 282 créditos

Las variables explicativas que se consideran relevantes en estudiantes cuya modalidad de estudios es a distancia son: edad de ingreso<sup>12</sup> (expresada de acuerdo al grupo de edad como jóvenes y adultos), género (mujer – hombre), tamaño de la región de procedencia (grande – pequeña) y nivel de participación en actividades en línea<sup>13</sup> (alta, media o baja).

Considerando las dos variables de respuesta como estadísticamente asociadas, el modelo más adecuado para el análisis de los datos es la regresión logística bivariante, por lo que este estudio se desarrolla bajo dos enfoques.



**Ilustración 9. Enfoques de la regresión bivariante**

**Fuente:** Elaboración propia

<sup>12</sup> La edad se segmentó en dos grupos en base a la clasificación realizada por Martín (2005).

<sup>13</sup> La variable participación en actividades en línea se incluye en la regresión como representante del enfoque “learning analytics”. Esta variable es el resultado de realizar un análisis clúster entre las variables número de chats, número de foros y número de video-colaboraciones en las que ha participado el estudiante, en promedio, durante su periodo de estudio.

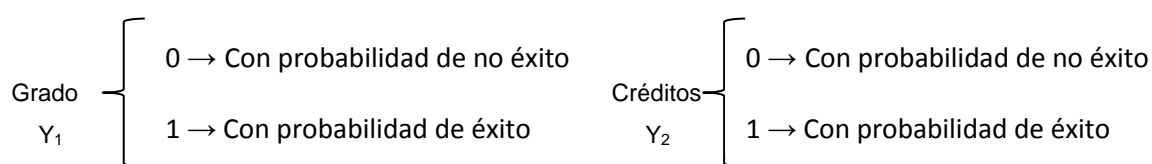
Para realizar la estimación logística bivalente con dos variables de respuesta dicótomas se utiliza la herramienta Rstudio de R Program<sup>14</sup> con el paquete Zelig<sup>15</sup> (Ver Anexo 4).

La regresión logística bivalente ordinal es estimada con la versión beta del pblm package: semiparametric regression for bivariate categorical responses in R (Enea y otros 2014).

## 6.1 Objetivos de la regresión logística bivalente

Inicialmente se busca determinar cuál es la influencia de las variables explicativas sobre un rendimiento académico óptimo y sobre un rendimiento académico positivo. Para lograr este objetivo se debe estimar una regresión logística bivalente con dos variables de respuesta binarias, siendo necesario convertir las variables de respuesta en dicótomas.

### Asociación Créditos y Grados – Rendimiento académico óptimo



---

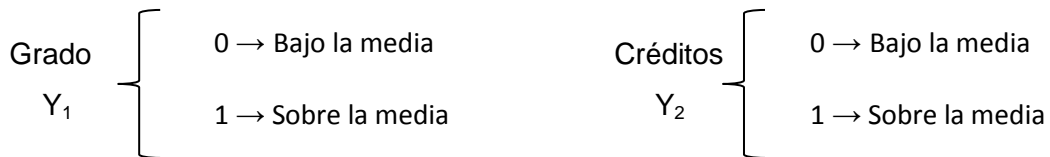
<sup>14</sup> R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <http://www.R-project.org/>.

<sup>15</sup> Zelig: Everyone's Statistical Software. M. Owen, K. Imai, G. King, O. Lau <https://cran.rproject.org/web/packages/Zelig/index.html>

		CRÉDITOS				
		220-282	160-219	110 - 159	50 - 109	<50
		5	4	3	2	1
GRADOS	40-39 A	Éxito	Éxito parcial (Grados)	Éxito parcial (Grados)	Éxito parcial (Grados)	Éxito parcial (Grados)
	38-36 B	Éxito	Éxito parcial (Grados)	Éxito parcial (Grados)	Éxito parcial (Grados)	Éxito parcial (Grados)
	35-33 C	Éxito parcial (Créditos)	No éxito	No éxito	No éxito	No éxito
	32-30 D	Éxito parcial (Créditos)	No éxito	No éxito	No éxito	No éxito
	29-28 E	Éxito parcial (Créditos)	No éxito	No éxito	No éxito	No éxito

Como *rendimiento académico óptimo* se consideran los casos en los que los estudiantes han tenido éxito, es decir, aquellos estudiantes que han alcanzado una acumulación de 220 a 282 créditos que los ubica en el último año de la carrera y además obtienen un grado entre A (Sobresaliente) y B (Muy buena), por lo tanto, aquellos estudiantes que no han tenido éxito son aquellos que obtienen un grado entre C y E y que se quedan mucho más tiempo que la duración legal de la carrera.

### Asociación Créditos y Grados – Rendimiento académico positivo



		CRÉDITOS				
		220-282	182-219	110 - 181	50 - 109	<50
		5	4	3	2	1
GRADOS	40-39 A	Sobre la media	Sobre la media	Sobre la media parcial (Grados)	Sobre la media parcial (Grados)	Sobre la media parcial (Grados)
	38-36 B	Sobre la media	Sobre la media	Sobre la media parcial (Grados)	Sobre la media parcial (Grados)	Sobre la media parcial (Grados)
	35-33 C	Sobre la media	Sobre la media	Sobre la media parcial (Grados)	Sobre la media parcial (Grados)	Sobre la media parcial (Grados)
	32-30 D	Sobre la media	Sobre la media	Sobre la media parcial (Grados)	Sobre la media parcial (Grados)	Sobre la media parcial (Grados)
	29-28 E	Sobre la media parcial (Créditos)	Sobre la media parcial (Créditos)	Bajo la media	Bajo la media	Bajo la media

Como *rendimiento académico positivo* se consideran aquellos casos en que los estudiantes obtienen un grado y una acumulación de créditos superior a la media del grupo estudiado, por lo que, el grupo que se ubica *sobre la media* tiene un grado entre A y D y una acumulación de créditos  $\geq 182$ .

Finalmente se realiza un análisis logístico con las dos *variables de respuesta ordinal*, cuyo objetivo específico es determinar cuál es la probabilidad de ocurrencia más alta (Créditos y grados bajos, créditos bajos y grados altos, créditos altos y grados bajos, créditos y grados altos).

## 6.2 Estudio de la relación entre las variables de respuesta

En las Tablas de asociación 22, 23 y 24 no se consideran covariables, se realiza una clasificación cruzada que muestra las respuestas de distribución conjunta entre Grados y Créditos y a través del test de chi2 se comprueba la existencia de asociación entre estas dos variables de respuesta.

Se considera la siguiente notación para las probabilidades conjuntas de las variables dependientes binarias:

$$\Pr (\text{Grados} = 0, \text{Créditos} = 0) \sim \pi_{00}$$

$$\Pr (\text{Grados} = 0, \text{Créditos} = 1) \sim \pi_{01}$$

$$\Pr (\text{Grados} = 1, \text{Créditos} = 0) \sim \pi_{10}$$

$$\Pr (\text{Grados} = 1, \text{Créditos} = 1) \sim \pi_{11}$$

**Tabla 22. Asociación entre Grados y Créditos, rendimiento académico óptimo**

			CRÉDITOS		TOTAL
			1 a 4 No éxito (0)	5 Éxito (1)	
<b>GRADOS</b>	C-E	No éxito (0)	293 (0.71)	102 (0.25)	395 (0.96)
	B-A	Éxito (1)	3 (0.01)	12 (0.03)	15 (0.04)
TOTAL			296 (0.72)	114 (0.28)	410 (1.00)

Pearson  $\chi^2(1) = 21.1305$  Pr = 0.000. Nota: entre paréntesis las proporciones

**Fuente:** Elaboración propia

Desde el enfoque de si un estudiante tiene éxito académico, los datos nos muestran un dato alarmante que debe ser analizado por la institución educativa, dado que, más de la mitad de los estudiantes (71%) no han tenido éxito puesto que no han alcanzado el número de créditos y el grado considerados como el mejor, lo cual se podría considerar como un rendimiento académico bajo. Las proporciones que indican los casos de éxito muestran que es mayor en cuanto a la acumulación de créditos (3%) pero igual en ambos casos sigue siendo bajo. Además, resulta significativa la asociación entre ambas variables, como lo indica el test chi-cuadrado ( $p \text{ value} < 0.0001$ ).

**Tabla 23. Asociación entre Grados y Créditos, rendimiento académico positivo**

			CRÉDITOS		TOTAL
			< 182 Bajo la media (0)	$\geq 182$ Sobre la media (1)	
<b>GRADOS</b>	< 31	Bajo la media (0)	144 (0.35)	44 (0.11)	188 (0.46)
	$\geq 31$	Sobre la media (1)	61 (0.15)	161 (0.39)	222 (0.54)
TOTAL			205 (0.50)	205 (0.50)	410 (1.00)

Pearson  $\chi^2(1) = 98.2365$  Pr = 0.000. Nota: entre paréntesis las proporciones

**Fuente:** Elaboración propia

La Tabla 23 indica que la mitad de la muestra tiene el número de créditos acumulados sobre la media (50%), mientras que la distancia entre el grupo que tiene el Grado y los Créditos bajo la media es de apenas 4% con respecto al grupo que tiene los Grados y Créditos sobre la media. Además, como cabía esperar, la asociación entre ambas variables es significativa según lo indicado por el test chi-cuadrado ( $p \text{ value} < 0.0001$ ).

Considerando las variables de respuesta como ordinales, se presenta una tabla  $A_1 \times A_2$  de doble entrada, cruzada con la clasificación de dos variables de respuesta ordinales  $A_1$  y  $A_2$ , respectivamente con  $D_1$  y  $D_2$  categorías.

Se definen las probabilidades marginales acumuladas (filas/columnas) como

$$u_r = \Pr(A_1 \leq r) = \sum_{i \leq r} \pi_i,$$

$$u_c = \Pr(A_2 \leq c) = \sum_{j \leq c} \pi_j,$$

Así, las probabilidades acumuladas del primer cuadrante superior izquierdo se obtienen a través de la expresión

$$u_{rc} = \Pr(A_1 \leq r, A_2 \leq c) = \sum_{i \leq r} \sum_{j \leq c} \pi_{ij}$$

**Tabla 24. Asociación entre Grados y Créditos como variables de respuesta ordinales**

			CRÉDITOS <sup>16</sup>				TOTAL
			50 - 109	110 - 159	160-219	≥220	
			2	3	4	5	
GRADOS	29-28	E	24 (0.06)	36 (0.09)	8 (0.02)	0	68 (0.17)
	32-30	D	17 (0.04)	61 (0.15)	104 (0.25)	40 (0.10)	222 (0.54)
	35-33	C	5 (0.01)	12 (0.03)	26 (0.06)	62 (0.15)	105 (0.26)
	38-36	B	0	0	3 (0.01)	12 (0.03)	15 (0.03)
TOTAL			46 (0.11)	109 (0.27)	141 (0.34)	114 (0.28)	410 (0.01)

Pearson  $\chi^2(9) = 176.1037$  Pr = 0.000. Nota: entre paréntesis las proporciones

**Fuente:** Elaboración propia

Dado que el valor del estadístico chi-cuadrado es significativo en las tres Tablas de asociación, se afirma que existe una relación entre las dos variables de respuesta y por lo tanto se procede a realizar el análisis logístico bivariante bajo los enfoques propuestos en la Ilustración 9.

### 6.3 Análisis logístico bivariante con dos variables de respuesta binarias

El modelo logístico bivariante con dos variables binarias consiste en modelizar las probabilidades de ausencia/presencia (0-1) de ambas variables de respuesta a través de un conjunto de covariables. El modelo trata de explicar el comportamiento conjunto de las variables, es decir, sus probabilidades conjuntas  $\pi_{hk} = \Pr[Y_1 = h, Y_2 = k]$  ( $h, k = 0, 1$ ) a través de la colección de covariables. Dada la relación biunívoca entre los parámetros

---

<sup>16</sup> No se incluye el Grado A (sobresaliente) y en Créditos la escala 1 por no existir en la población de estudio.

$(\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11})$  y  $(\pi_1, \pi_2, \psi)$  donde  $\pi_j = \Pr[Y_j = 1] \quad j = 1, 2$  y  $\psi = \frac{\pi_{11}\pi_{00}}{\pi_{10}\pi_{01}}$ , el modelo puede ser parametrizado según la terna  $(\pi_1, \pi_2, \psi)$ .

Las covariables seleccionadas (edad, género, tamaño de región y participación en actividades en línea) se suponen relacionadas con las dos variables binarias de respuesta Grados y Créditos (en el Anexo 5 se representa gráficamente la relación entre covariables y variables de respuesta), por lo que, realizando la modelización conjunta las probabilidades marginales vienen dadas por el siguiente modelo de regresión:

$$Y_j | X = x \sim Be(\pi_j(x)) \quad j = 1, 2$$

$$\log \left[ \frac{\pi_j(x)}{1 - \pi_j(x)} \right] = \beta_j^T X$$

$$\log \psi = \beta_3^T X$$

En el caso que nos ocupa

$$\log \left[ \frac{\pi_j(x)}{1 - \pi_j(x)} \right] = \alpha_j + \beta_{j1} \text{Edad} + \beta_{j2} \text{Género} + \beta_{j3} \text{Re gió n} + \beta_{j4} \text{Participación}$$

$$\log \psi = \alpha_3 + \beta_{31} \text{Edad} + \beta_{32} \text{Género} + \beta_{33} \text{Re gió n} + \beta_{34} \text{Participación}$$

donde

$j = 1, 2$  es decir representa las probabilidades marginales de Grados y Créditos

$\psi$  es la Odds Ratio entre  $Y_1, Y_2$

$\beta_j$  son los coeficientes de las covariables



Al realizar las estimaciones simultáneas, los resultados que se obtienen tanto para los casos de rendimiento académico óptimo como para los casos de rendimiento académico positivo son los que se recogen a continuación.

a) **Análisis de casos de rendimiento académico óptimo**

Se excluyen las variables género y tamaño de región por no resultar estadísticamente significativas (en el Anexo 6 se presenta la estimación incluyendo éstas covariables).

En la Tabla 25 se observa que la variable edad de ingreso es significativa para ambas variables de respuesta y las Odds Ratio de ambos coeficientes (0.22 y 0.38 respectivamente) son menores a la unidad, las “Odds Ratios” entre  $Y_j$ , Edad ( $k=1$ ) y Participación ( $k=2$ ) indican que los estudiantes jóvenes tienen menos ventaja (entendiendo “ventaja” como la razón entre las probabilidades de éxito y fracaso) de conseguir un rendimiento académico óptimo con respecto a los estudiantes en edad adulta, es decir, existe menos probabilidad de que obtengan un Grado entre A o B y una acumulación de créditos mayor a  $\geq 220$ .

La variable participación en actividades en línea se presenta significativa solo para el caso de la acumulación de créditos, la Odds Ratio resultante es 0.465, la cual es menor que la unidad, lo que significa que los estudiantes que participan poco en las actividades en línea (chat, foro y video-colaboración) tienen menos ventaja de conseguir un Grado entre A o B y una acumulación de créditos mayor a  $\geq 220$ , en comparación con los estudiantes que participan medianamente o de forma activa, quienes se asume tienen más ventaja de obtener un rendimiento académico óptimo.

**Tabla 25. Estimación logística bivariante con dos variables de respuesta binarias**

Log-likelihood: -278.3545

Deviance: 556.7089

Coefficients	Estimate	Std.Error	z value	Pr(> z )	Odds Ratio
<b>GRADOS</b>					
<i>Edad ingreso</i>					
Adulto	0 base				
Joven	-1.50857	0.55016	-2.742	0.006106 **	<b>0.221226</b>
<i>Participación</i>					
Alta	0 base				
Media	0.31208	0.74378	0.420	0.674794	1.366264
Baja	-0.40871	0.74416	-0.549	0.582852	0.664507
Intercept	-2.04283	0.69239	-2.950	0.003173 **	
<b>CRÉDITOS</b>					
<i>Edad ingreso</i>					
Adulto	0 base				
Joven	-0.96636	0.30599	-3.158	0.001588 **	<b>0.380465</b>
<i>Participación</i>					
Alta	0 base				
Media	0.47743	0.32757	1.457	0.144988	1.611926
Baja	-0.76402	0.31704	-2.410	0.015959 *	<b>0.465790</b>
Intercept	0.08653	0.36264	0.239	0.811399	

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

**Fuente:** Elaboración propia con apoyo de Rstudio y el paquete Zelig

Con las estimaciones obtenidas en la Tabla 25 se puede determinar las estimaciones de las probabilidades acumuladas conjuntas, es decir, la probabilidad de que sucedan o no

los casos de éxito, dadas las covariables, estas probabilidades se obtienen a través de la fórmula tomada de Kosuke y otros (2008):

$$\pi_{11} = \begin{cases} \frac{1}{2}(\psi - 1)^{-1} - a - \sqrt{a^2 + b} & \text{if } \psi \neq 1 \\ \pi_1 \pi_2 & \text{if } \psi = 1 \end{cases} \quad (\text{PC})$$

$$\pi_{10} = \pi_1 - \pi_{11}$$

$$\pi_{01} = \pi_2 - \pi_{11}$$

$$\pi_{00} = 1 - \pi_{10} - \pi_{01} - \pi_{11}$$

Donde

$\pi_{00}, \dots, \pi_{11}$  son las probabilidades individuales

$$a = 1 + (\pi_1 + \pi_2)(\psi - 1)$$

$$b = -4\psi(\psi - 1)\pi_1\pi_2$$

En la Tabla 26 se incluye las estimaciones de las probabilidades conjuntas marginales a través de:

$$\tilde{\pi}_{kh} = \frac{1}{n} \sum_{i=1}^n \pi_{kh}(X_i)$$

Donde

$$k, h = 0, 1$$

$X_i$  son los valores de las covariables en la muestra de individuos

Por lo que, análogamente una estimación conjunta de  $\psi$  sería  $\tilde{\psi} = \frac{\tilde{\pi}_{11} \tilde{\pi}_{00}}{\tilde{\pi}_{10} \tilde{\pi}_{01}}$

**Tabla 26. Probabilidades conjuntas -rendimiento académico óptimo**

Probabilities	mean	sd	50%	2.5%	97.5%
Pr(Y1=0, Y2=0)	0.808	0.027	0.810	0.751	0.856
Pr(Y1=0, Y2=1)	0.166	0.026	0.165	0.119	0.222
Pr(Y1=1, Y2=0)	0.009	0.006	0.008	0.002	0.024
Pr(Y1=1, Y2=1)	0.017	0.008	0.015	0.006	0.037

**Fuente:** Elaboración propia con apoyo de Rstudio y el paquete Zelig

Los resultados recogidos en la Tabla 26, nos indican que dadas la covariables “edad de ingreso y participación en actividades en línea”, las probabilidades conjuntas que se presentan son las siguientes:

$\tilde{\pi}_{00}$ : La probabilidad de que un estudiante no tenga éxito, es decir, que obtenga una calificación entre entre C y E y un número inferior a 220 créditos acumulados es del 81%.

$\tilde{\pi}_{01}$ : La probabilidad de que un estudiante no tenga éxito en su calificación (obtiene entre C y E) pero que si tenga éxito en su acumulación de créditos (obtiene más de 220 créditos) es del 17%.

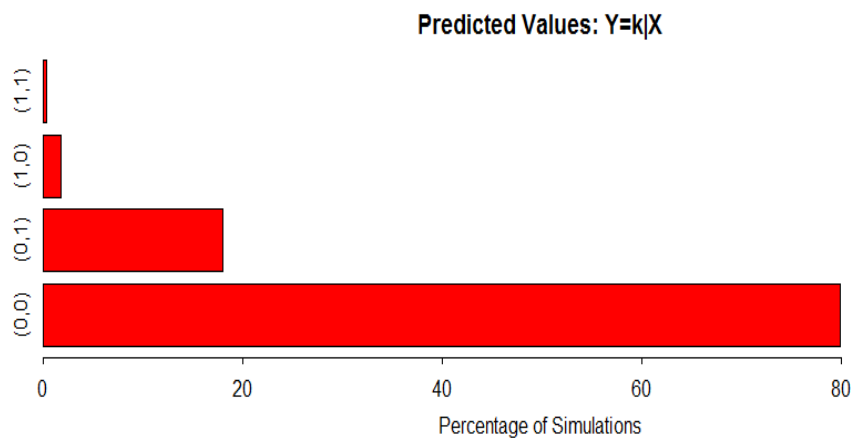
$\tilde{\pi}_{10}$ : La probabilidad de que un estudiante tenga éxito en su calificación (obtiene entre A y B) pero que no tenga éxito en la acumulación de créditos (obtiene menos de 220 créditos) es cerca del 1%.

$\tilde{\pi}_{11}$ : La probabilidad de que un estudiante tenga éxito total, es decir que, obtenga una calificación entre A y B y una acumulación > a 220 créditos, es cerca del 2%.

En la Ilustración 10 se representa los valores ajustados de las probabilidades, en donde se observa con claridad que la probabilidad de que ocurra (Y1=0, Y2=0), es decir que los estudiantes no tengan éxito es la más alta de todas.

**Tabla 27. Valores ajustados - rendimiento académico positivo**

Probabilities	No ocurra	Ocurra
(Y1=0, Y2=0)	0.201	0.799
(Y1=0, Y2=1)	0.820	0.180
(Y1=1, Y2=0)	0.982	0.018
(Y1=1, Y2=1)	0.997	0.003



**Ilustración 10. Valores ajustados –Casos de éxito**

**Fuente:** Elaboración propia con apoyo de Rstudio- paquete Zelig

**b) Análisis de casos de rendimiento académico positivo**

Como se recogía anteriormente, para realizar este análisis se consideran aquellos casos en que los estudiantes obtienen un grado y una acumulación de créditos superior a la media del grupo estudiado, por lo que, el grupo que se ubica sobre la media tienen un grado entre A y D y cuya acumulación es  $\geq 182$  créditos.

Se realiza una estimación conjunta con todas las covariables (ver Anexo 7) en las que se determina cuáles son las variables significativas que ajustan mejor el modelo y permiten generar la estimación final que se presenta en la Tabla 28.

Analizando los coeficientes significativos, los resultados muestran que el tamaño de la región de procedencia se relaciona únicamente con la variable de respuesta Grados y la edad está relacionada solo con la variable de respuesta Créditos.

La variable tamaño de la región presenta un coeficiente positivo, su Odds Ratio es mayor a la unidad (2.30) lo cual indica que a igualdad de todas las demás condiciones, un estudiante que proviene de una región grande tiene 2.30 veces más “ventaja” de obtener un grado sobre la media, en comparación con los estudiantes que provienen de regiones pequeñas. Entendiendo “ventaja” como la razón entre las probabilidades de éxito y fracaso.

La variable edad se relaciona únicamente con Créditos y resulta su coeficiente es negativo siendo la Odds Ratio menor a la unidad (0.35) lo que indica que los estudiantes jóvenes tienen menos ventaja en comparación con los adultos de acumular créditos sobre la media.

**Tabla 28. Estimación logística bivariante con dos variables de respuesta binarias**

Log-likelihood: -505.8385

Deviance: 1011.677

Coefficients	Estimate	Std.Error	z value	Pr(> z )	Odds Ratio
<b>GRADOS</b>					
Edad					
Adulto	0 base				
Joven	-0.4544	0.3005	-1.512	0.13054	0.634829
Región					
Pequeña	0 base				
Grande	0.8337	0.3297	2.528	0.01146 *	<b>2,301819</b>
Intercept	-0.1828	0.4136	-0.442	0.65850	0,832935
<b>CRÉDITOS</b>					
Edad					
Adulto	0 base				
Joven	-1.0487	0.3150	-3.330	0.00087 ***	<b>0.350392</b>
Región					
Pequeña	0 base				
Grande	0.1898	0.3212	0.591	0.55465	1,209007
(Intercept)	0.7450	0.4191	1.778	0.07543 .	2,106441

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 '.' 1

**Fuente:** Elaboración propia con apoyo de Rstudio y el paquete Zelig

Al realizar la respectiva simulación aplicando la fórmula de la referencia (PC) los resultados que se recogen en la Tabla 29.

**Tabla 29. Probabilidades conjuntas -rendimiento académico positivo**

Probabilities	mean	sd	50%	2.5%	97.5%
Pr(Y1=0, Y2=0)	0.351	0.023	0.351	0.305	0.397
Pr(Y1=0, Y2=1)	0.108	0.016	0.107	0.079	0.141
Pr(Y1=1, Y2=0)	0.147	0.018	0.147	0.113	0.182
Pr(Y1=1, Y2=1)	0.395	0.024	0.394	0.351	0.443

**Fuente:** Elaboración propia con apoyo de Rstudio y el paquete Zelig

En la Tabla 29 se observa que al incluir las variables “edad de ingreso y tamaño de la región” asociadas con las dos variables de respuesta, las estimaciones de las probabilidades conjuntas que se obtienen indican lo siguiente:

$\tilde{\pi}_{00}$ : La probabilidad de que un estudiante obtenga un grado y una acumulación de créditos bajo la media grupal, es del 35%.

$\tilde{\pi}_{01}$ : La probabilidad de que un estudiante obtenga un grado bajo la media grupal y una acumulación de créditos sobre las media, es del 11%.

$\tilde{\pi}_{10}$ : La probabilidad de que un estudiante obtenga un grado sobre la media grupal y una acumulación de créditos bajo la media, es del 15%.

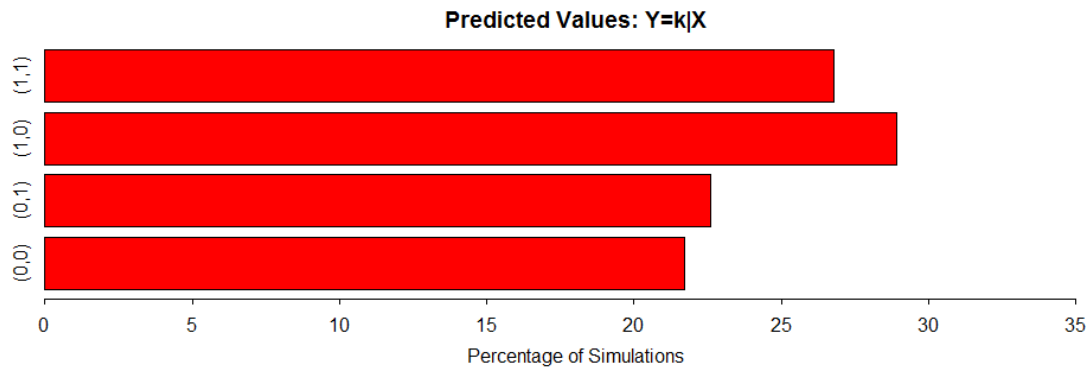
$\tilde{\pi}_{11}$ : La probabilidad de que un estudiante obtenga un grado y una acumulación de créditos sobre la media grupal, es del 40%.

En la ilustración 11 se presenta los valores ajustados de la Tabla 30, la cual muestra que no existe mucha diferencia porcentual entre el valor sobre y bajo la media, siendo la primera probabilidad la más alta (Y1=0, Y2=0).



**Tabla 30. Valores ajustados -rendimiento**

Probabilities	0	1
(Y1=0, Y2=0)	0.783	0.217
(Y1=0, Y2=1)	0.774	0.226
(Y1=1, Y2=0)	0.711	0.289
(Y1=1, Y2=1)	0.732	0.268



**Ilustración 11. Valores ajustados –Casos sobre la media**

**Fuente:** Elaboración propia con apoyo de Rstudio- paquete Zelig

## 6.4 Análisis logístico bivalente con dos variables de respuesta ordinales

Se estima la asociación entre las dos variables de respuesta y las covariables edad e interacción por ser significativas y las que mejor ajustan el modelo.

El modelo logístico bivalente que se estima en la Tabla 31 se basa en el modelo propuesto por Dale (1986):

$$\log(\psi) = \alpha_{ia} + \beta_{ja} + \delta_{ij} - \gamma^T \mathbf{x}$$

$i = 1, \dots, r - 1; j = 1, \dots, c - 1; \mathbf{a} = \text{asociación};$  con restricciones simples, tales como  $\alpha_{ri,a} = \beta_{c-1,a} = 0; \delta_{i,c-1} = 0, i = 1, \dots, r - 1; \delta_{r-1,j} = 0, j = 1, \dots, c - 1.$  Entonces  $\gamma = 0$  indica que la asociación no depende de las categorías de la variable Y o X.

Relacionando la covariable edad con la variable dependiente grados, se observa que los estudiantes jóvenes tienen una probabilidad de obtener un mejor grado 1.8 veces más que los estudiantes en edad adulta. En lo que respecta a los créditos el comportamiento es similar, esto significa que los estudiantes jóvenes tienen la probabilidad de acumular créditos 2.4 veces más que los estudiantes de edad adulta.

La participación baja en actividades en línea presenta una asociación significativa solo con la variable de respuesta Créditos, al observar la Odds Ratio menor a la unidad, se puede decir que, aquellos estudiantes que tienen una participación baja, tienen una ventaja menor, es decir 1.6 veces menos (1/0.636) de acumular créditos en comparación a los que si participan activamente.

Al final de la Tabla 31 se observa el log-global odds ratios (Log-GORs) entre Grados y créditos, dada las covariables edad y participación, cuyo valor es mayor a la unidad, lo cual indica que dada las covariables significativas las variables de respuesta Grados y Créditos tienen la probabilidad de ser 2.23 veces mejores que aquellos estudiantes de edad adulta o estudiantes que hayan tenido una participación media o alta, este último resultado inesperado se puede justificar porque tal vez los estudiantes que participan en actividades en línea de forma pasiva están usando otras estrategias de estudio, como las tutorías telefónicas o desarrollo de autoevaluaciones.

**Tabla 31. Estimación de la regresión logística bivariante ordinal**

Log-likelihood : -896.338      Deviance : 1792.677      AIC : 1818.677

Coefficients	beta	se	z	p.value	Odds Ratio
<b>GRADOS</b>					
Intercept [1]	-2.31599	0.3475	-6.6655	0.00000 ***	
Intercept [2]	0.22399	0.3249	0.6894	0.49056	
Intercept [3]	2.63155	0.3978	6.6155	0.00000 ***	
<i>Edad</i>					
Adultos	0	base			
Jóvenes	0.60254	0.2722	2.2136	0.02685 *	<b>1.826752</b>
<i>Participación</i>					
Alta	0	base			
Media	-0.11884	0.2919	-0.4071	0.68392	0.887949
Baja	0.37975	0.2657	1.4291	0.15299	1.461919
<b>CRÉDITOS</b>					
Intercept [1]	-3.24217	0.3572	-9.0756	0.00000 ***	
Intercept [2]	-1.61425	0.3276	-4.9275	0.00000 ***	
Intercept [3]	-0.05829	0.3175	-0.1836	0.85431	
<i>Edad</i>					
Adultos	0	base			
Joven	0.86136	0.2679	3.2149	0.00130 **	<b>2.366377</b>
<i>Participación</i>					
Alta	0	base			
Media	-0.29345	0.2803	-1.0470	0.29509	0,745687
Baja	0.63612	0.2531	2.5133	0.01196 *	<b>0.636125</b>
<b>Log-GORs entre Grados y Créditos</b>					
Intercept	2.23325	0.1847	12.0907	0.00000 ***	

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' 1

**Fuente:** Elaboración propia con apoyo de Rstudio y el paquete Zelig

Luego realizando una clasificación cruzada se puede obtener las Odds Ratios globales entre la asociación de las dos variables de respuesta.

**Tabla 32. Log-Odds Ratios globales**

Coefficients	beta	se	z	p.value
Intercept [1] [1]	1.9622	0.3199	6.134	0.00000 ***
Intercept [2] [1]	2.5975	0.2684	9.676	0.00000 ***
Intercept [3] [1]	3.2328	0.4460	7.249	0.00000 ***
Intercept [1] [2]	1.4622	0.4214	3.470	0.00052 ***
Intercept [2] [2]	1.9511	0.2589	7.536	0.00000 ***
Intercept [3] [2]	2.4400	0.2358	10.350	0.00000 ***
Intercept [1] [3]	0.9623	0.7855	1.225	0.22057
Intercept [2] [3]	1.3048	0.4973	2.624	0.00869 **
Intercept [3] [3]	1.6474	0.4722	3.488	0.00049 ***

0 '\*\*\*\*' 0.001 '\*\*\*' 0.01 '\*\*' 0.05 '.' 0.1 ' ' 1

**Fuente:** Elaboración propia con apoyo de Rstudio y el paquete Zelig

Se observa que la asociación entre las dos variables de respuesta es positiva, es decir todos las Odds Ratios son mayores que la unidad, a excepción de la interacción [1] [3], pero esta no es significativa.

**Tabla 33. Asociación Grados y Créditos con log-GORs**

			CRÉDITOS			
			50 – 109	110 - 159	160-219	≥220
			1	2	3	4
<b>GRADOS</b>	29-28	E	1.96	1.46	0.96	
	32-30	D	2.59	1.95	1.30	base
	35-33	C	3.23	2.44	1.65	
	38-36	B		base		

En la Tabla 33 se muestra que la clasificación cruzada del los estudiantes con las dos variables de respuesta y su correspondientes Odds Ratio globales dan valores positivos, indicando la existencia de una relación directa. Los valores más altos indican que, la

probabilidad de que un estudiante obtenga un grado C y que acumule créditos entre 50 y 109 es 3.23 veces más posible que obtener un grado y acumulación de créditos considerados como los óptimos, algo similar sucede con las demás interacciones.

# CAPÍTULO 7

## DISCUSIÓN Y FUTURAS INVESTIGACIONES

---

El objetivo central de esta tesis, recogido en el preámbulo de la misma, es construir un modelo para determinar el rendimiento académico de los estudiantes basado en *learning analytics*, mediante el uso de técnicas estadísticas multivariantes. Ello conlleva la adaptación de dichas técnicas al contexto de la investigación así como la estructuración de las mismas recogiendo sus últimos desarrollos teóricos y la descripción del software estadístico adecuado para su aplicación.

Dicho objetivo central quedaba configurado a través de diversos objetivos específicos referidos a la modelización del rendimiento académico universitario (medido a través de diversos indicadores) respecto a un conjunto de variables predictoras o explicativas medidas en distintos niveles (individuales, grupales o contextuales), tratando de determinar el aporte de dichas variables sobre la conducta de los indicadores del rendimiento, y de identificar la naturaleza de la relación entre las variables del enfoque *learning analytics* y el citado rendimiento académico de los estudiantes.

Las variables incluidas en la presente investigación permiten identificar cual es la influencia que ejercen sobre el rendimiento académico, estas estimaciones pueden permitir a una institución educativa mejorar la focalización de las intervenciones y los servicios de apoyo a estudiantes en riesgo de problemas académicos.

Los resultados obtenidos dan respuesta a las hipótesis y objetivos planteados, además este trabajo es un punto de partida para futuras investigaciones que consideren que el ámbito tecnológico se está convirtiendo en una de las mejores herramientas de enseñanza aprendizaje, sobre todo en educación a distancia

## 7.1 Del análisis multinivel

La discusión de resultados se realiza sobre la estimación de la Tabla 19, es el modelo que mejor se ajusta de acuerdo al estadístico desviación (Deviance) y al criterio de información de Akaike (AIC), por lo tanto se puede decir que:

En la parte de estimaciones de efectos fijos y aleatorios se observa que todos los predictores a nivel del estudiante y del aula son estadísticamente significativos, este resultado se obtiene al haber realizado un procedimiento “stepwise” hacia adelante que permitió eliminar secuencialmente las variables que no tenían significación estadística.

Dentro de estas variables no significativas se encuentra la variable *género*, corroborando los resultados presentados en las investigaciones de Moral (2006); Acevedo & Rocha (2011) y Pantoja & Alcaide (2013) coinciden en concluir que no existen diferencias entre hombres y mujeres con respecto al rendimiento académico.

Otra variable considerada como característica del estudiante es la región de procedencia, la cual no presenta significancia en su relación con el rendimiento académico, estos resultados están acorde con los obtenidos por Di Gresia & Porto (2004) y Garzón y otros (2010).

En cuanto a las *becas* por nivel de ingresos o méritos académicos, a pesar de que Garzón y otros (2010) y Celis y otros (2005) determinan que existe una relación positiva y

estadísticamente significativa entre las becas y el rendimiento académico, en nuestros resultados finales no es significativa estadísticamente. Esto puede estar causado por el hecho de que el porcentaje de becas asignadas en nuestro caso de estudio es muy bajo (1%), lo cual implica que no se provoque la influencia esperada sobre el rendimiento académico.

El modelo final involucra:

- Tres covariables del Nivel 2: tasa de repetidores, ciclo y tipo de docente.
- Ocho variables del Nivel 1: edad, rinde supletorio, repite materia, participa en chat, participa en foro, participa en videocolaboración, N° comentarios, N° accesos al LMS
- Cuatro interacciones multinivel
- Y la varianza de cinco pendientes del Nivel 1.

El coeficiente de la variable tasa de repetidores medida en el intervalo  $[0,1]$ , nos indica que un aumento en 10 puntos porcentuales de estudiantes matriculados por segunda o tercera vez en una asignatura troncal, ocasiona una disminución de 2.1 puntos en el rendimiento académico. Esto significa que a pesar de que se asume que los estudiantes tienen más experiencia que los estudiantes nuevos en la materia, no obtienen una mejor nota, lo cual podría estar ligado a la metodología de enseñanza o a los instrumentos de evaluación.

Otra variable del nivel 2 es la variable ciclo. Los resultados indican que cuando la asignatura se encuentra en un ciclo superior el rendimiento académico incrementa en 0.8 décimas. Esto se puede esperar ya que se considera que conforme un estudiante avanza a ciclos superiores tiene más conocimientos y en cierta forma ha adquirido madurez académica.

La pendiente de la variable tipo de docente influye positivamente sobre el rendimiento académico, ya que, este tiende a subir aproximadamente 1 punto si el docente es



administrativo o invitado. Estos resultados se pueden explicar posiblemente por dos razones: la primera sería que los docentes a tiempo completo son más estrictos y la segunda puede ser que estos docentes tienen más créditos o asignaturas a su cargo en comparación a los docentes invitados o administrativos. Estos en sí son dos supuestos, que se deberían de verificar en base a otros aspectos.

En cuanto a la edad, los resultados indican que por un año más de edad que tenga el estudiante, el puntaje del rendimiento académico subirá en 0.08 décimas. El comportamiento de estos resultados coinciden con los planteados por Nasir (2012); Alhajraf & Alasfour (2014), quienes encontraron que la edad tiene una relación positiva y significativa con el rendimiento académico de los estudiantes universitarios.

Los coeficiente de las pendientes de la variable rinde supletorio y su interacción con el ciclo indican que si un estudiante se queda suspenso y está en un ciclo superior el rendimiento académico disminuirá en 0.86 décimas (resultante de la suma de los coeficientes -0.52677 y -0.33669 recogidos en la Tabla 19). Mientras que analizando los resultados de la variable repite materia y su interacción con el ciclo nos muestra que si un estudiante repite la materia y está en un ciclo superior, el rendimiento académico en promedio subirá en 2.6 décimas (resultante de la suma de los coeficientes 2.80826 y -0.24867 recogidos en la Tabla 19).

Todas las variables del enfoque “learning analytics” tienen una relación positiva con el rendimiento académico, siendo la participación en chat, foro y video-colaboración las que ocasionan el mayor impacto ya que provocan un incremento de entre 1 y 2 puntos en el rendimiento académico, afirmando de esta forma que si existe una relación significativa con el rendimiento académico tal como lo plantean Agudo y otros (2012), Yu & Jo (2014). La variable N° accesos al LMS interacciona con la tasa de repetidores y el ciclo de la asignatura, lo cual indica que ocasiona un incremento de cerca de 0.11 décimas en el rendimiento académico (resultado de la suma de los coeficientes 0.04385, 0.06964 y -0.00422 recogidos en la Tabla 19).

## 7.2 Del análisis logístico bivariante

En las tres modelizaciones logísticas bivariantes que se han ajustado, las covariables que sobresalen son la edad de ingreso a la universidad y la participación activa en línea.

La variable género no es estadísticamente significativa en ninguna de las estimaciones, esto va acorde con las conclusiones de Moral (2006), Acevedo & Rocha (2011), Pantoja & Alcaide (2013) que indican que no existe relación entre género y rendimiento académico.

En cuanto a la región de procedencia no se presenta ningún efecto en el rendimiento académico.

Para el caso del estudio del rendimiento académico óptimo, la variable edad indica que los estudiantes jóvenes tienen menos ventaja (entendiendo “ventaja” como la razón entre las probabilidades de éxito y fracaso) de conseguir un rendimiento académico óptimo con respecto a los estudiantes en edad adulta. Este resultado coincide con los planteados por Bourner & Hamed (1987) y por Richardson (1995), que señalan que los estudiantes adultos tienden a aproximarse a un aprendizaje más óptimo debido a su persistencia.

La variable participación en actividades en línea muestra que los estudiantes que participan poco en las actividades en línea (chat, foro y video-colaboración) tienen menos ventaja de conseguir un rendimiento académico óptimo, en comparación con los estudiantes que participan medianamente o de forma activa.

Para el caso del análisis del rendimiento académico positivo, la variable edad tiene el mismo comportamiento que para el rendimiento académico óptimo, la diferencia es que ahora la variable tamaño de región de procedencia del estudiante es significativa para el GRADO, cuyos resultados muestran que los estudiantes que provienen de una región grande tienen más ventaja de obtener un grado sobre la media. Estos resultados se pueden justificar a través de la conclusión de Ortlieb & Cheek, 2008 que consideran que

los estudiantes en diversas ubicaciones geográficas tienen conocimientos específicos, experiencias previas y modos de vida las cuales juegan un papel crítico en su proceso de aprendizaje. Resultados de análisis similares conducen a discrepancias al respecto, dado que algunos estudios indican que no hay un efecto significativo (Porto & Di Gresia, 2004; Garzón y otros 2010), mientras que en otros resultados señalan que si existe una influencia sobre el rendimiento académico (Cresswell & Underwood, 2004).

La diferencia entre las probabilidades que resultan de la asociación entre las dos variables de respuesta no está muy alejada de las probabilidades resultantes incluyendo covariables, por ejemplo la probabilidad de no éxito solo con variables de respuesta es del 71% y luego de incluir covariables esta sube al 81%, lo cual da cabida a futuras investigaciones en las que se pueda incluir más variables de otras dimensiones.

Finalmente, se debe concluir que los modelos logísticos bivariantes permiten obtener conclusiones en el contexto del análisis del rendimiento académico, proporcionando información relevante sobre las variables asociadas a la mejora del mismo. Especial hincapié se debe realizar sobre el modelo bivalente logístico ordinal, recientemente actualizado por Enea & Attanasio (2015).

### **7.3 Futuras investigaciones**

Tras el desarrollo de la investigación recogida en esta memoria de tesis, se abren nuevas líneas y expectativas de investigación para el futuro. Una de ellas centrada en el rendimiento académico universitario y la segunda centrada en la mejora y la adaptación de técnicas estadísticas a este contexto, especialmente el modelo de regresión bivalente logístico ordinal.

La investigación sobre el rendimiento académico es amplia y muy actual en el ámbito universitario, dada la necesidad de mejorar y optimizar los resultados de la inversión de recursos (económicos, humanos y sociales). En consecuencia, es esencial la búsqueda

de los factores que inciden en la mejora de los mismos y, por tanto, se puede y se debe realizar otros trabajos para la consecución de variables idóneas.

El interés de mejorar la calidad educativa, busca la mejora del rendimiento académico. Este indicador es de interés para varios sectores de la sociedad (instituciones educativas, instituciones gubernativas, sector empresarial, ciudadanos, etc.). Para el presente estudio se consideraron variables controlables por la institución de educación superior, porque el aporte es dirigido principalmente a las universidades, sin embargo, se pudiera incluir determinantes socioeconómicas, socioculturales, institucionales, intelectuales, afectivas, académicas, etc.

Sería interesante analizar la asociación entre dos variables de respuesta a través del modelo de regresión logístico bivariante para medir la tasa de deserción o de graduación. Para ello es muy conveniente ampliar la modelización bivariante ordinal y profundizar en sus posibilidades de interpretación en contextos de investigación educativa y procesos de aprendizaje.



## **REFERENCIAS BIBLIOGRÁFICAS**

---

- Abela, J. (2011). El análisis multinivel: una revisión actualizada en el ámbito sociológico. *Metodología de Encuestas*, 161-176.
- Acevedo, C., & Rocha, F. (2011). Estilos de aprendizaje, género y rendimiento académico. *Journal of Learning Styles*, 71-84.
- Agudo, A., Hernandez, A., & Iglesias, S. (2012). Predicting academic performance with learning analytics in virtual learning environments: a comparative study of three interaction classifications. *IEEE Xplore, digital library*.
- Agudo, A., Iglesias, S., Conde, M., & Hernández, Á. (2014). Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning. *Computers in Human Behavior*, 542–550.
- Aitkin, M., & Longford, N. (1986). Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society*, 1-43.
- Alhajraf, N., & Alasfour, A. (2014). The impact of demographic and academic characteristics on academic performance. *International Business Research*, 92-100.
- Ayesha, S., Mustafa, T., Sattar, A., & Khan, M. (2010). Data mining model for higher education system. *European Journal of Scientific Research*, 24-29.
- Baepler, P., & James, C. (2010). *Academic Analytics and Data Mining in Higher Education*. Recuperado en marzo de 2013, de [http://academics.georgiasouthern.edu/ijstotl/v4n2/essays\\_about\\_sotl/PDFs/\\_BaeplerMurdoch.pdf](http://academics.georgiasouthern.edu/ijstotl/v4n2/essays_about_sotl/PDFs/_BaeplerMurdoch.pdf)
- Bienkowski, M., Feng, M., & Means, B. (2012). *Enhancing teaching and learning through educational data mining and learning analytics*. Recuperado en enero de 2013, de U.S. Department of Education Office of Educational Technology: <http://www.ed.gov/edblogs/technology/files/2012/03/edm-la-brief.pdf>

- Bin Mat, U., Buniyamin, N., Arsad, P., & Kassim, R. (2013). An overview of using academic analytics to predict and improve students' achievement: a proposed proactive intelligent intervention. *Engineering Education*, 2013 IEEE 5th Conference, 126-130.
- Biniaminov, I., & Glasman, N. (1983). School determinants of student achievement in secondary education. *American Educational Research Journal*, 251-268.
- Boado, H. (2013). *Introducción al análisis multinivel*. Madrid: CIS-Centro de Investigaciones Sociológicas.
- Borde, S. (1998). Predictors of student academic performance in the introductory marketing. *Journal of Education for Business*, 302-306.
- Bourner, T., & Hamed, M. (1987). Entry qualifications and degree performance: summary findings report. Council for National Academic Awards. London.
- Bridgeman, B., & Wendler, C. (1991). Gender differences in predictors of college mathematics performance and in college mathematics course grades. *Journal of Educational Psychology*, 275-284.
- Brown, M. (2011). *Learning Analytics: the coming third wave*. Recuperado en julio de 2014, de <https://net.educause.edu/ir/library/pdf/ELIB1101.pdf>
- Bryk, A., & Raudenbush, S. (1992). *Hierarchical linear models: applications and data analysis methods*. Thousand Oaks, CA, US: Sage Publications.
- Caldas, S., & Bankston, C. (1999). Multilevel examination of student, school, and district-level effects on academic achievement. *The Journal of Educational Research*, 91-100.
- Calero, J., Choi, Á., & Waisgrais, S. (2010). *Determinantes del riesgo de fracaso escolar en España: una aproximación a través de un análisis logístico multinivel aplicado a PISA-2006*. Recuperado en julio de 2014, de [http://www.revistaeducacion.educacion.es/re2010/re2010\\_09.pdf](http://www.revistaeducacion.educacion.es/re2010/re2010_09.pdf)
- Campbell, J., & Oblinger, D. (2007). *Academic Analytics*. Recuperado en diciembre de 2012, de <http://net.educause.edu/ir/library/pdf/PUB6101.pdf>
- Campbell, T., & Campbell, D. (1997). Faculty/student mentor program: effects on academic performance and retention. *Research in Higher Education*, 727-742.

- Carrión, E. (2002). *Validación de características al ingreso como predictores del rendimiento académico en la carrera de medicina*. Recuperado en febrero de 2015, de [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S0864-21412002000100001&lng=es&nrm=iso](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21412002000100001&lng=es&nrm=iso)
- Cascón, I. (2000). Análisis de las calificaciones escolares como criterio de rendimiento académico. *Colegio Público Juan García Pérez*, <https://campus.usal.es/~inico/investigacion/jornadas/jornada2/comun/c17.html>.
- Celis, M., Martínez, J., & Lozano, J. (2005). Los programas de becas de la UNAM: características de operación y análisis inicial de su impacto académico. *Ponencia presentada en el 5º Congreso Retos y Expectativas de la Universidad: Experiencias y Dilemas de la Reforma*. Tampico: <http://www.congresoretosyexpectativas.udg.mx>.
- Coleman, J. (1968). Equality of educational opportunity. *Integrated Education*, 19-28.
- Cresswell, J., & Underwood, C. (2004). *Cresswell, J., & Underwood, C. (2004). Location, location, location: Implications of geographic situation on Australian student performance in PISA 2000*. Australia: Australian Council for Educational Research.
- Dale, J. (1986). Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics*, 909-917.
- Darda, A. (2009). *Modelling of African Farm Dynamics Using Bivariate Binary Logistic Regression in WinBUGS*. Recuperado en abril de 2015, de [http://www.researchgate.net/profile/Md\\_Darda/publication/50319233\\_Modelling\\_of\\_African\\_Farm\\_Dynamics\\_Using\\_Bivariate\\_Binary\\_Logistic\\_Regression\\_in\\_WinBUGS/links/09e4150f9f400bb872000000.pdf](http://www.researchgate.net/profile/Md_Darda/publication/50319233_Modelling_of_African_Farm_Dynamics_Using_Bivariate_Binary_Logistic_Regression_in_WinBUGS/links/09e4150f9f400bb872000000.pdf)
- Dayioğlu, M., & Türüt-Aşık, S. (2007). Gender differences in academic performance in a large public university in Turkey. *Higher Education*, 255-277.
- de Miguel, M., Urquijo, P., Arias, J., Escudero, T., Rodríguez, S., & Vidal, J. (2002). Evaluación del rendimiento en la enseñanza superior. Comparación de resultados entre alumnos procedentes de la LOGSE y del COU. *Revista de Investigación Educativa*, 357-383.



- Di Gresia, L., & Porto, A. (2005). *Una nota sobre la asignación de horas de estudio*. Recuperado en junio de 2014, de <http://www.depeco.econo.unlp.edu.ar/espec/notamat01.pdf>
- Díaz de Rada, V. (2002). *Técnicas de análisis multivariante para investigación social y comercial. Ejemplos prácticos utilizando SPSS*. Madrid: RA-MA S.A.
- Díaz, M., Urquijo, P., Arias, J., Escudero, T., Rodríguez, S., & Vidal, J. (2002). Evaluación del rendimiento en la enseñanza superior. Comparación de resultados entre alumnos procedentes de la LOGSE y del COU. *Revista de Investigación Educativa*, 357-383.
- Dickerson, S., & Hazelton, N. (2012). Data Mining and Analysis Methodology for Higher Education Curriculum Development and Recruitment Practices: An Interim Report. *Surveying and Land Information Science*, 129-143.
- Draper, D. (1995). Inference and hierarchical modeling in the social sciences (with discussion). *Journal of Educational and Behavioral Statistics*, 115-147.
- Duff, A., Boyle, E., Dunleavy, K., & Ferguson, J. (2004). The relationship between personality, approach to learning and academic performance. *Personality and individual differences*, 1907-1920.
- Duval, E. (2011). *Attention Please ! Learning Analytics for Visualization and Recommendation*. Obtenido de <https://lirias.kuleuven.be/bitstream/123456789/315113/1/>
- Dyckhoff, A., Zielke, D., Bültmann, M., Chatti, M., & Schroeder, U. (2012). Design and Implementation of a Learning Analytics Toolkit for Teachers. *Journal of Educational Technology & Society*, 58-76.
- Edel Navarro, R. (2003). *El rendimiento académico: concepto, investigación y desarrollo*. Recuperado en octubre de 2012, de <http://www.ice.deusto.es/RINACE/reice/vol1n2/Edel.pdf>
- Elías, T. (2011). *Learning Analytics: Definitions, Processes and Potential*. Recuperado en diciembre de 2012, de <http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf>

- Elias, T. (2011). *Learning Analytics: the Definitions, the Processes, and the Potential*. Recuperado en enero de 2013, de <http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf>
- Enea, M., & Attanasio, M. (2015). An association model for bivariate data with application to the analysis of university students' success. *Journal of Applied Statistics*, 1-12.
- Escudero, T. (1999). Indicadores del rendimiento académico: una experiencia en Zaragoza. En *Indicadores en la Universidad: información y definiciones* (págs. 251-262). Madrid: Ministerio de Educación y Cultura. Centro de Publicaciones.
- Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 304-317.
- Friedman, L. (1989). Friedman, L. (1989). Mathematics and the gender gap: A meta-analysis of recent studies on sex differences in mathematical tasks. *Review of Educational research*, 185-213.
- Fullana, J. (1992). *Revisió de la recerca educativa sobre les variables explicatives del rendiment acadèmic: Apunts per a l'ús del criteri de "modificabilitat pedagògica" de les variables*. Recuperado en abril de 2012, de [http://dugi-doc.udg.edu/bitstream/handle/10256/1669/revisio\\_recerca\\_educativa.pdf?sequence=1](http://dugi-doc.udg.edu/bitstream/handle/10256/1669/revisio_recerca_educativa.pdf?sequence=1)
- Gammie, P., & Gammie, D. (2003). Gender differences in accounting education: an undergraduate exploration. *Accounting Education: An International Journal*, 177-196.
- Garzón, R., Rojas, M., del Riesgo, L., & Pinzón, M. (2010). Factores que pueden influir en el rendimiento académico de estudiantes de Bioquímica que ingresan en el programa de Medicina de la Universidad del Rosario-Colombia. *Educación médica*, 85-96.
- Glass, G., & Smith, M. (1978). *Meta-Analysis of research on the relationship of class-size and achievement. The class size and instruction project*. Recuperado en diciembre de 2014, de ERIC: <http://files.eric.ed.gov/fulltext/ED168129.pdf>
- Goldstein, H. (2011). *Multilevel statistical models (Vol. 922)*. John Wiley & Sons.

- Goldstein, H., & Spiegelhalter, D. (1996). League tables and their limitations: statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society*, 385-443.
- Goldstein, H., Rasbash, J., Yang, M., Woodhouse, G., Pan, H., Nuttall, D., y otros. (1993). A multilevel analysis of school examination results. *Oxford review of education*, 425-433.
- Goldstein, P., & Katz, R. (2005). *Academic analytics: The uses of management information and technology in higher education*. Recuperado en abril de 2014, de Educase: <http://net.educause.edu/ir/library/pdf/ERS0508/ekf0508.pdf>
- González, A. (1988). Indicadores del rendimiento escolar: relación entre pruebas objetivas y calificaciones. *Revista de educación*, 31-54.
- Guruler, H., & Istanbulu, A. (2014). Modeling Student Performance in Higher Education Using Data Mining. *Educational Data Mining*, 105-124.
- Haist, S., Wilson, J., Elam, C., Blue, A., & Fosson, S. (2000). The effect of gender and age on medical school performance: an important interaction. *Advances in health sciences Education*, 197-205.
- Hox, J. (1995). *Applied Multilevel Analysis*. Amsterdam: TT-Publikaties.
- Hox, J. (2002). *Multilevel analysis: Techniques and applications*. Psychology Press.
- Hoxby, C. (2000). The Effects of Class Size on Student Achievement: New Evidence from Population Variation. *The Quarterly Journal of Economics*, 1239-1285 .
- Jiménez, Á., & Álvarez, H. (2010). *Minería de datos en la educación*. Recuperado en enero de 2013, de <http://www.it.uc3m.es/jvillena/irc/practicas/10-11/08mem.pdf>
- Johnson, L., Adams Becker, S., Cummins, M., Estrada, V., Freeman, A., & Ludgate, H. (2013). *Horizon Report: 2013 Higher Education Edition*. Austin, Texas: The New Media Consortium.
- Johnson, L., Smith, R., Willis, H., Levine, A., & Haywood, K. (2011). *The 2011 Horizon Report*. Recuperado en diciembre de 2012, de <http://net.educause.edu/ir/library/pdf/hr2011.pdf>

- Khan, I., & Choi, J. (2014). An Application of Educational Data Mining (EDM) Technique for Scholarship Prediction. *International Journal of Software Engineering and its Applications*, 31-42.
- Kosuke, I., King, G., & Lau, O. (2008). Toward a common framework for statistical analysis and development . *Journal of Computational and Graphical Statistics*, 892-913.
- Kotzé, M., & Niemann, R. (2013). Psychological resources as predictors of academic performance of first-year students in higher education. *Acta Academica*, 85-121.
- Kreft, I. (1995). *The Effects of Centering in Multilevel Analysis: Is the Public School the Loser or the Winner?* Recuperado en diciembre de 2014, de <http://files.eric.ed.gov/fulltext/ED392837.pdf>
- Latiesa, M. (1992). *La deserción universitaria*. Madrid: Centro de investigaciones sociológicas.
- Lauría, E., Moody, E., Jayaprakash, S., Jonnalagadda, N., & Baron, J. (2013). Open academic analytics initiative: initial research findings. *Proceedings of the Third International Conference on Learning Analytics and Knowledge* , 150-154.
- Le Cessie, S., & Van Houwelingen, J. (1994). Logistic Regression for Correlated Binary Data. *Applied Statistics*, 95-108.
- Lonn, S., Aguilar, S., & Teasley, S. (2014). Investigating student motivation in the context of a learning analytics intervention during a summer bridge program. *Computers in Human Behavior*.
- Luan, J. (2002). Data Mining and Its Applications in Higher Education. En *New Directions for Institutional Research* (págs. 17–36).
- Maccoby, E., & Jacklin, C. (1974). *The psychology of sex differences*. California: Stanford University Press.
- Marjoribanks, K. (1976). Academic achievement, intelligence, and creativity: a regression surface analysis. *Multivariate behavioral research*, 105-118.
- Marrero, H., & Espino, O. (1988). Evaluación comparativa del poder predictor de las aptitudes sobre notas escolares y pruebas objetivas. *Revista de educación*, 97-112.

- Martín, E., García, L., Torbay, Á., & Rodríguez, T. (2008). Estrategias de aprendizaje y rendimiento académico en estudiantes universitarios. *International Journal of Psychology and Psychological Therapy*, 401-412.
- Martín, J. (2005). Los factores definitorios de los grandes grupos de edad de la población: tipos, subgrupos y umbrales. *Revista electrónica de geografía y ciencias sociales*, 181-204.
- McArdle, J., Hamagami, F., Chang, J., & Hishinuma, E. (2014). Longitudinal dynamic analyses of depression and academic achievement in the Hawaiian High Schools Health Survey using contemporary latent variable change models. *Structural equation modeling: a multidisciplinary journal*, 608-629.
- McArdle, J., Paskus, T., & Boker, S. (2013). A Multilevel Multivariate Analysis of Academic Performances in College Based on NCAA Student-Athletes. *Multivariate Behavioral Research*, 57-95.
- Miguel-Vázquez, V., Bartolomei-Suárez, S., & Grey, W. (2013). The socio-economic background effect on UPRM engineering students' academic performance. *IIE Annual Conference and Expo 2013*, 587-596.
- Molnar, A., Smith, P., Zahorik, J., Palmer, A., Halbach, A., & Ehrle, K. (1999). Evaluating the SAGE program: A pilot program in targeted pupil-teacher reduction in Wisconsin. *Educational Evaluation and Policy Analysis*, 165-177.
- Moral, J. (2006). *Predicción del rendimiento académico universitario*. Recuperado en junio de 2014, de Perfiles educativos: [http://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S0185-26982006000300003&lng=es&tlng=es](http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0185-26982006000300003&lng=es&tlng=es)
- Morris, L., Wu, S., & Finnegan, C. (2005). Predicting retention in online general education courses. *The American Journal of Distance Education*, 23-36.
- Murillo Torrecilla, F. J. (2008). Los modelos multinivel como herramienta para la investigación educativa. *MagisRevista internacional de investigación en educación*, 45-62.
- Nasir, M. (2012). Demographic characteristics as correlates of academic achievement of university students. *Academic Research International*, 400-405.

- Nuttall, D., Goldstein, H., Prosser, R., & Rasbash, J. (1989). Differential school effectiveness. *International Journal of Educational Research*, 769-776.
- Oladejo, M., Ige, N., Fagunwa, A., & Arewa, O. (2010). Socio-demographic variables and distance learners' academic performance at the University of Ibadan, Nigeria. *European Journal of Scientific Research*, 540-553.
- Olmos, M., & Corrin, L. (2012). Academic analytics in a medical curriculum: enabling educational excellence. *Australasian Journal of Educational Technology*, 1-15.
- Ortlieb, E., & Cheek, E. (2008). How geographic location plays a role within instruction: Venturing into both rural and urban elementary schools. *Educational Research Quarterly*, 51-68.
- Othman, M., & Musa, A. (2014). The Improvement of Students' Academic Performance by Using Social Media through Collaborative Learning in Malaysian Higher Education. *Asian Social Science*, 210-221.
- Page, M. (1990). *Hacia un modelo causal del rendimiento académico*. Madrid: Centro de publicaciones - Secretaria General Técnica, Ministerio de Educación y Ciencia.
- Palmer, S. (2013). Modelling engineering student academic performance using academic analytics. *International journal of engineering education*, 132-138.
- Pantoja, A., & Alcaide, M. (2013). La variable Género y su relación con el autoconcepto y el rendimiento académico de alumnado universitario. *Revista científica electrónica de Educación y Comunicación en la Sociedad del Conocimiento*, 124-140.
- Penny, M., & White, W. (1998). Developmental Mathematics Students' Performance: Impact of Faculty and Student Characteristics. *Journal of Developmental Education*, 2-12.
- Pong, S.-I., & Pallas, A. (2001). Class size and eighth-grade math achievement in the United States and abroad. *Educational evaluation and policy analysis*, 251-273.
- Porto, A., & Di Gresia, L. (2004). Rendimiento de estudiantes universitarios y sus determinantes. *Revista de Economía y Estadística*, 93-113.
- Prensky, M. (2001). *Digital Natives, Digital Immigrants*. Recuperado el enero de 2015, de <http://www.marcprensky.com/writing/Prensky%20-%20Digital%20Natives,%20Digital%20Immigrants%20-%20Part1.pdf>

- Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods*. California: Sage publications.
- Richardson, J. (1995). Mature students in higher education: II. An investigation of approaches to studying and academic performance. *Studies in Higher Education*, 5-17.
- Richardson, J., Morgan, A., & Woodley, A. (1999). Approaches to studying in distance education. *Higher Education*, 23-55.
- Rodríguez, M., & Ruiz, M. (2011). Indicadores de rendimiento de estudiantes universitarios versus créditos acumulados. *Revista de educación*, 467-492.
- Rodriguez, S. (1982). *Factores de rendimiento escolar*. Barcelona: Oikos-tau.
- Rodríguez, S., Fita, E., & Torrado, M. (2004). El rendimiento académico en la transición secundaria-universidad. *Revista de educación*, 391-414.
- Roig, J. (1983). Más en torno a la variable "agrupamiento" en relación con el rendimiento. *Revista de pedagogía*, 279-314.
- Rojas, L. (2013). *Validez predictiva de los componentes del promedio de admisión a la Universidad de Costa Rica utilizando el género y el tipo de colegio como variables de control*. Recuperado en marzo de 2013, de [http://revista.inie.ucr.ac.cr/uploads/tx\\_magazine/validez-predictiva-componentes-promedio-admision-universidad-costa-rica-rojas.pdf](http://revista.inie.ucr.ac.cr/uploads/tx_magazine/validez-predictiva-componentes-promedio-admision-universidad-costa-rica-rojas.pdf)
- Sadler-Smith, E. (1996). Approaches to studying: age, gender and academic performance. *Educational Studies*, 367-379.
- Schram, C. (1996). A meta-analysis of gender differences in applied statistics achievement. *Journal of Educational and Behavioral Statistics*, 55-70.
- Sellström, E., & Bremberg, S. (2006). Is there a "school effect" on pupil outcomes? A review of multilevel studies. *Journal of Epidemiology and community health*, 149-155.
- Sepúlveda, M., López, M., & otros. (2011). Diferencias de género en el rendimiento académico y en el perfil de estilos y de estrategias de aprendizaje en estudiantes de Química y Farmacia de la Universidad de Concepción. *Revista de estilos de aprendizaje*, 135-150.

- Sharkey, M. (2011). Academic analytics landscape at the University of Phoenix. *In Proceedings of the 1st International Conference on Learning Analytics and Knowledge*, 122-126.
- Shin, Y., & Raudenbush, S. (2011). The causal effect of class size on academic achievement multivariate instrumental variable estimators with data missing at random. *Journal of Educational and Behavioral Statistics*, 154-185.
- Shyamala, K., & Rajagopalan, S. (2006). Data mining model for a better higher educational system. *Information Technology Journal*, 560-564.
- Siemens, G., Gasevic, D., Haythornthwaite, C., Dawson, S., Buckingham, S., Ferguson, R., y otros. (2011). *Open Learning Analytics : an integrated & modularized platform Proposal to design , implement and evaluate an open platform to integrate heterogeneous learning analytics techniques*. Obtenido de <http://solaresearch.org/OpenLearningAnalytics.pdf>
- Tejedor, J. (2003). Poder explicativo de algunos determinantes del rendimiento en los estudios universitarios. *Revista española de pedagogía*, 5-32.
- Tempelaar, D., Rienties, B., & Giesbers, B. (2014). Computer Assisted, Formative Assessment and Dispositional Learning Analytics in Learning Mathematics and Statistics. *Communications in Computer and Information Science*, 67-78.
- Tirados, R. (1990). Principales dificultades en el rendimiento académico en primer año de carrera de ingeniería. *La investigación educativa sobre la Universidad* (págs. 261-279). Madrid: Centro de Investigación y Documentación Educativa.
- Tomás, J., Expósito, M., & Sempere, S. (2014). Factors in students' academic performance in bachelor's degrees. A study in management and business administration. *Revista de Investigacion Educativa*, 379-392.
- Tournon, J. (1984). *Factores del rendimiento académico en la universidad*. España: Ediciones Universidad de Navarra, S.A. .
- Tourón, J. (1985). La predicción del rendimiento académico: procedimientos, resultados e implicaciones. *revista española de pedagogía*, 473-495.
- Turner, R., Camilli, G., Kroc, R., & Hoover, J. (1986). Policy strategies, teacher salary incentive, and student achievement: An explanatory model. *Educational Researcher*, 5-11.



- Vallejo, G., Arnau, J., & Bono, R. (2008). Construcción de modelos jerárquicos en contextos aplicados. *Psicothema*, 830.
- Xu, R., & Wunsch, D. (2009). Clustering. *Computational Intelligence Magazine*, 92-95.
- Yu, T., & Jo, I. (2014). Educational Technology Approach toward Learning Analytics : Relationship between Student Online Behavior and Learning Performance in Higher Education. *ACM International Conference Proceeding Series*, 269-270.
- Zax, J., & Rees, D. (2002). IQ, academic performance, environment, and earnings. *Review of Economics and Statistics*, 600-616.
- Zhang, Y., Oussena, S., Clark, T., & Hyensook, K. (2010). Using data mining to improve student retention in HE: a case study. *Paper presented at the ICEIS 2010 - Proceedings of the 12th International Conference on Enterprise Information Systems*, 190-197.

***ANEXOS***

---

## ANEXO 1.

### Estimación conjunta predictores del nivel 2

Mixed-effects ML regression	Number of obs =	23583
Group variable: AULA	Number of groups =	468
	Obs per group: min =	1
	avg =	50.4
	max =	470
	Wald chi2(11) =	2559.22
Log likelihood = -83489.128	Prob > chi2 =	0.0000

REND_ACADEMICO	Coef.	Std. Err.	z	P>z	[95% Conf. Interval]	
N_matriculados	.0000351	.0002016	0.17	0.862	-.00036	.0004302
Tasa repetidores	.7006228	1.034866	0.68	0.498	-1.327678	2.728923
Tasa aprobación	19.47311	.5709319	34.11	0.000	18.3541	20.59211
Créditos	-.3188622	.1319581	-2.42	0.016	-.5774954	-.060229
Ciclo	.0184366	.0446542	0.41	0.680	-.069084	.1059572
Experiencia	-.0330937	.0156945	-2.11	0.035	-.0638543	-.0023331
Evaluación	.0373765	.0232904	1.60	0.109	-.0082718	.0830248
Tipo_docente						
Tiempo completo	0	(base)				
Administrativo	.4133271	.3471715	1.19	0.234	-.2671165	1.093771
Invitado	.3686677	.2200823	1.68	0.094	-.0626857	.800021
Formación						
Grado	0	(base)				
Maestría	.3332799	.1927417	1.73	0.084	-.0444869	.7110467
Doctorado	.4120786	.4861203	0.85	0.397	-.5406998	1.364857
_cons	14.45224	13.01438	11.10	0.000	11.90147	17.00301

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
AULA: Identity				
var(_cons)	.9973223	.1564203	.7333864	1.356245
var(Residual)	68.94602	.6389946	67.70492	70.20987

LR test vs. linear regression: chibar2(01) = 141.15 Prob >= chibar2 = 0.0000

## ANEXO 2.

### Regresiones multinivel por grupo de variables del nivel 1

Variables	Modelo1	Modelo2	Modelo3	Modelo4
REND_ACADEMICO	Sociodemográficas	Antecedentes	Learning analytics	Grupal
Edad	0.148 <sup>***</sup>			0.0876 <sup>***</sup>
Genero(mujer)	0.572 <sup>***</sup>			0.147
Region (pequeña9	0.152			0.198
Rinde_supletorio		-1.766 <sup>***</sup>		-1.470 <sup>***</sup>
Repite_materia		1.293 <sup>***</sup>		2.035 <sup>***</sup>
Tiene Beca		6.014 <sup>***</sup>		3.487 <sup>*</sup>
N_comentarios			0.0955 <sup>***</sup>	0.0805 <sup>***</sup>
N_mensajes			0.317 <sup>*</sup>	0.306 <sup>*</sup>
N_accesos_curso			-0.168	-0.123
N_accesos_LMS			0.0439 <sup>***</sup>	0.0432 <sup>***</sup>
Tiempo_LMS			0.0104 <sup>***</sup>	0.0111 <sup>***</sup>
Participa_chat			1.317 <sup>***</sup>	1.340 <sup>***</sup>
Participa_foro			2.046 <sup>***</sup>	2.079 <sup>***</sup>
Participa_video			1.401 <sup>***</sup>	1.369 <sup>***</sup>
Constant	20.84 <sup>***</sup>	26.41 <sup>***</sup>	21.49 <sup>***</sup>	18.95 <sup>***</sup>
$\sigma^2$	67.84723	68.13585	56.35935	54.79256
$\tau_\gamma$	15.09004	15.62792	13.7757	12.88085
Varianza explicada	0.018	0.014	0.184	0.207
Deviance	167353.13	167465.02	163014.53	162334.59

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

En el modelo 1, la edad y género son estadísticamente significativos, en el modelo 2 todas las variables son significativas, donde, se entiende que si el estudiante se queda

suspenso el rendimiento académico tenderá a disminuir aproximadamente en 2 puntos. En el modelo 3 que considera los predictores del “learning analytics” solo una de las ocho variables no es significativa, aquí la participación de los estudiantes en el chat, video colaboración, en promedio, aumentarán entre 1 y 2 puntos la nota del rendimiento académico y además explica la varianza un 20% más que los dos primeros modelos.

### ANEXO 3

#### Determinando la significancia de los coeficientes de las variables del nivel 1 incluidas como efectos aleatorios

a) AIC Nonrandomly varying slopes models

Model	Obs	ll (null)	ll (model)	df	AIC	BIC
.	23583	.	-80895.55	19	161829.1	161982.4

b) Incluyendo Edad

Model	Obs	ll(model)	ll (Edad)	df	AIC	BIC
.	23583	-80895.55	-80895.55	20	161831.1	161992.5

No es significativa AIC del modelo incluyendo EDAD > que AIC del modelo a)

c) Incluyendo Rinde supletorio

Model	Obs	ll (modelo)	ll (R_supletorio)	df	AIC	BIC
.	23583	-80895.55	-80815.33	20	161670.7	161832

Es significativa AIC del modelo incluyendo Rinde supletorio < que AIC del modelo a)

d) Incluyendo Repite materia

Model	Obs	ll (R_supletorio)	ll (Re_materia)	df	AIC	BIC
.	23583	-80815.33	-80812.11	21	161666.2	161835.6

Es significativo AIC del modelo incluyendo Repite materia < que AIC del modelo c)

e) Incluyendo Participa chat

Model	Obs	ll (Re_materia)	ll (P_chat)	df	AIC	BIC
.	23583	-80812.11	-80808.64	22	161661.3	161838.8

Es significativo AIC del modelo incluyendo Participa en chat < que AIC del modelo d)

f) Incluyendo Participa en foro

Model	Obs	ll (P_chat)	ll (P_foro)	df	AIC	BIC
.	23583	-80808.64	-80807.15	23	161660.3	161845.9

Es significativo AIC del modelo incluyendo Participa en foro < que AIC de modelo e)

g) Incluyendo Participa en video-colaboración

Model	Obs	ll (P_foro)	ll (P_video)	df	AIC	BIC
.	23583	-80807.15	-80805.91	24	161659.8	161853.5

Es significativo AIC del modelo incluyendo Participa en video < que AIC de modelo f)

h) Incluyendo N° comentarios

Model	Obs	ll (P_video)	ll (N°comentarios)	df	AIC	BIC
.	23583	-80805.91	-80804.99	25	161660	161861.7

No es significativo AIC del modelo incluyendo N°comentarios > que AIC de modelo g)

i) Incluyendo N°accesos al LMS

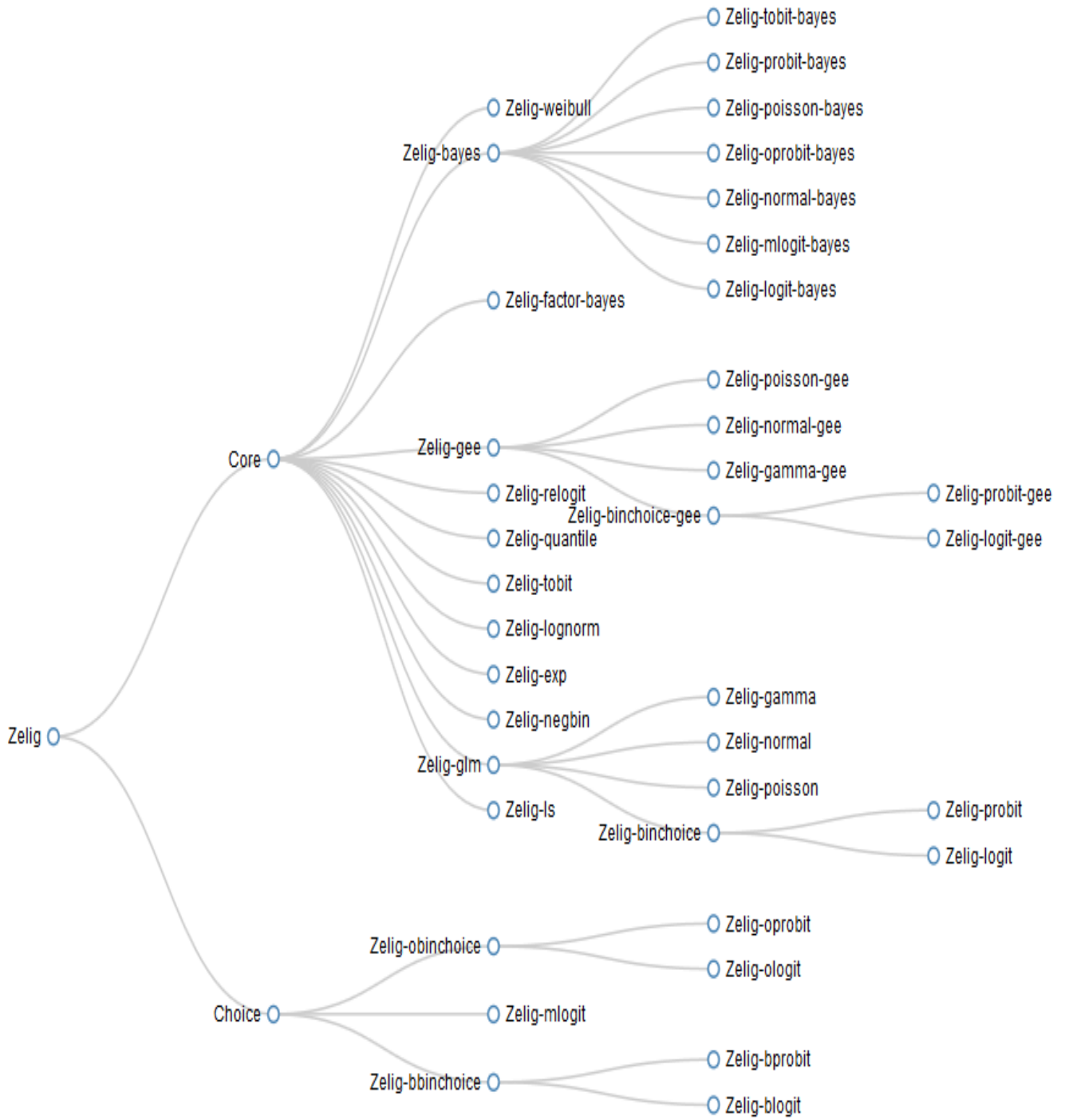
Model	Obs	ll (P_video)	ll (N°accesos)	df	AIC	BIC
.	23583	-80805.91	-80798.39	25	161646.8	161848.5

No es significativo AIC del modelo incluyendo N°accesos LMS > que AIC de modelo g)

Por lo tanto las variables que se deben incluir en el modelo final son Rinde supletorio, repite materia, participa chat, participa en foro y participa en video.

## ANEXO 4

### Modelos estadísticos de Zelig



Fuente: <http://zeligproject.org/> (2015).



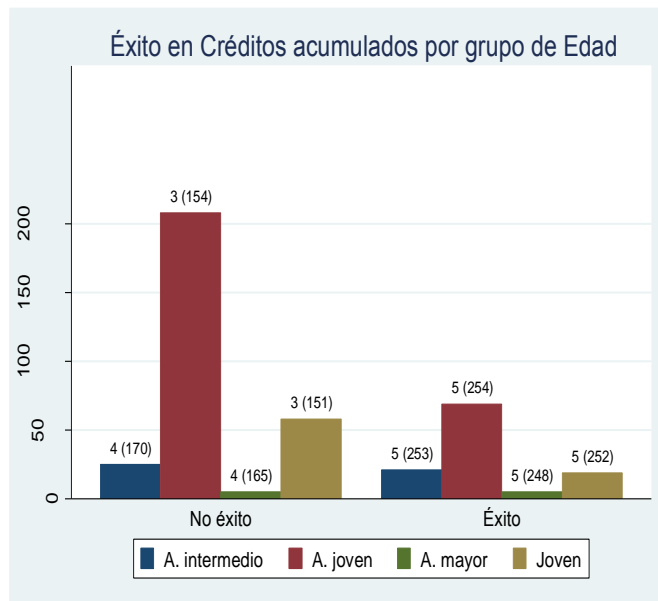
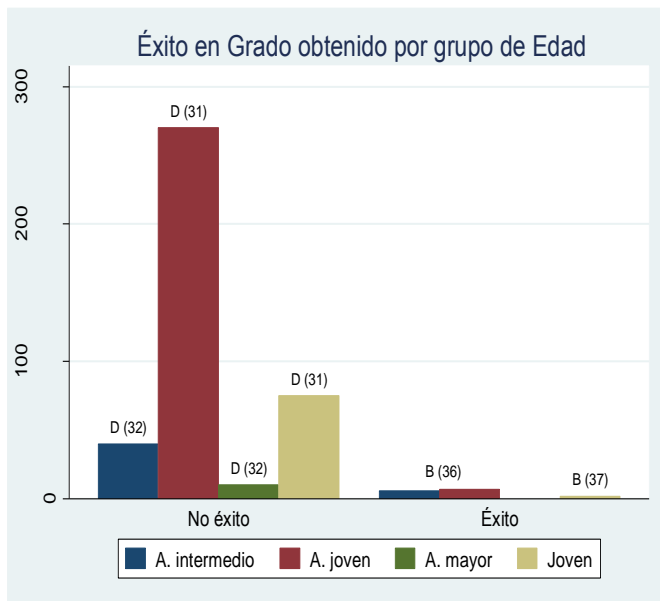
## ANEXO 5. Gráficas de asociación de covariables con variables de respuesta

a) Asociación de Grados y créditos con covariables, casos de éxito

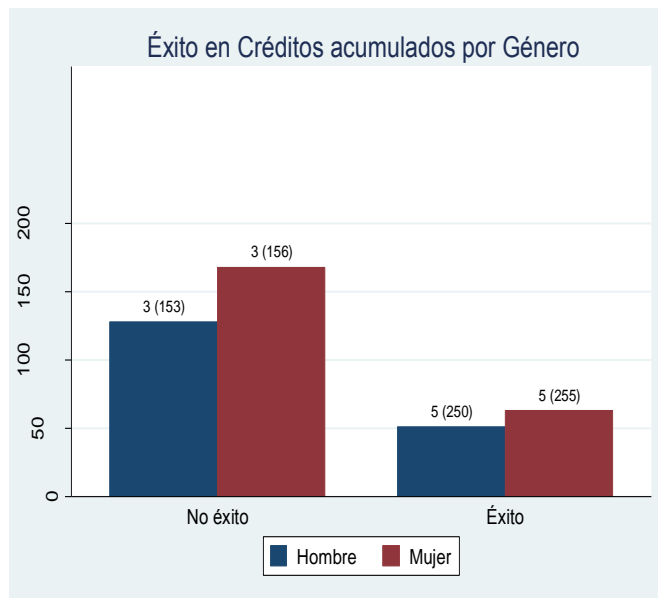
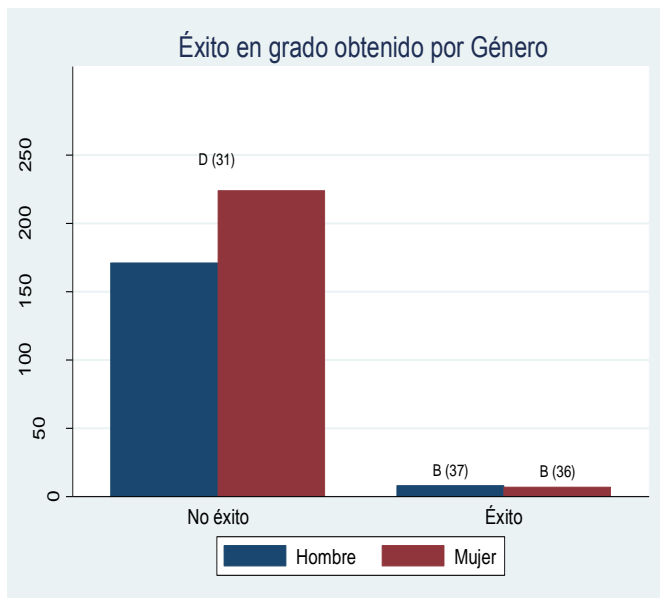
**GRADOS**

**CRÉDITOS**

Edad



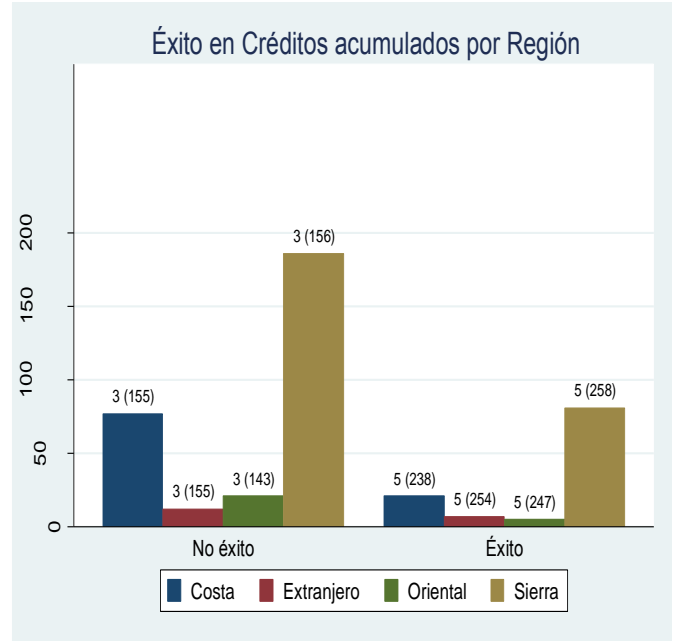
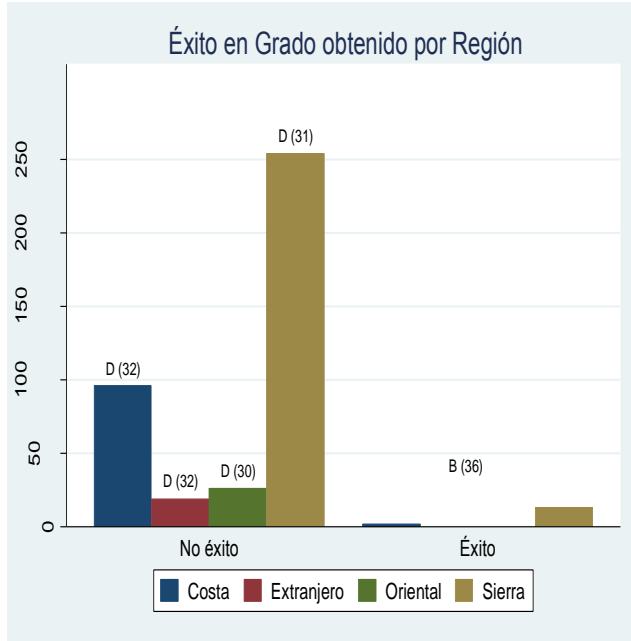
Género



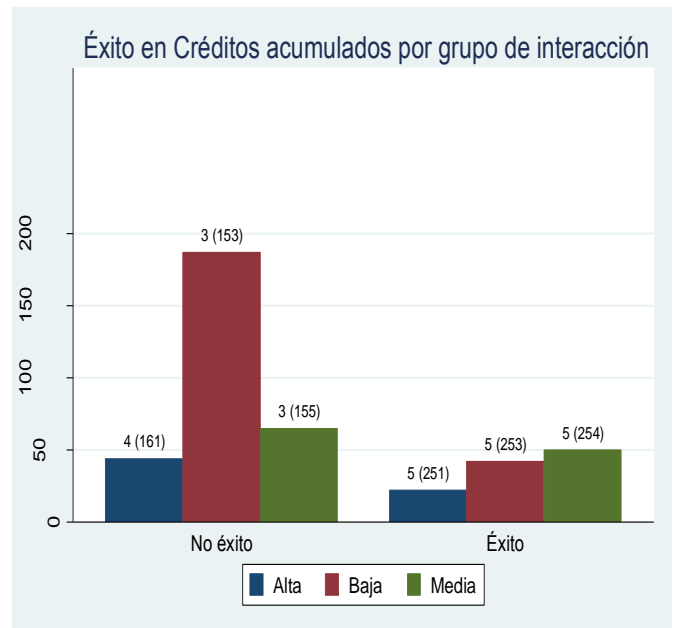
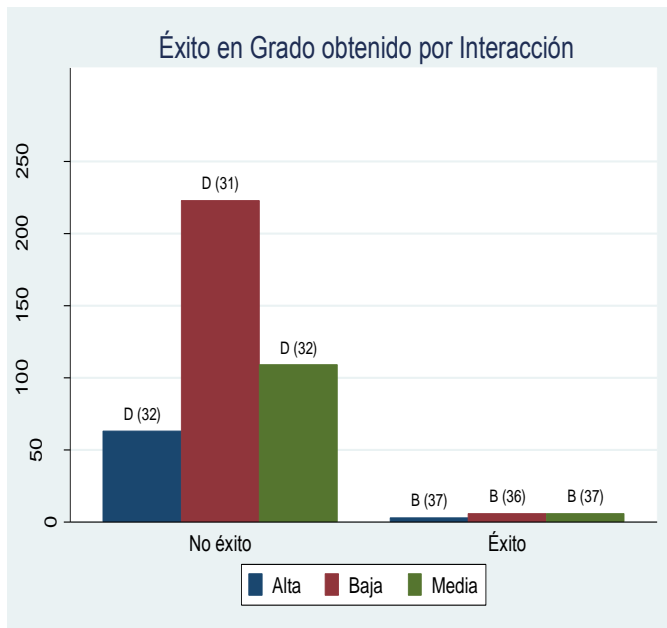
**GRADOS**

**CRÉDITOS**

Región



Interacción

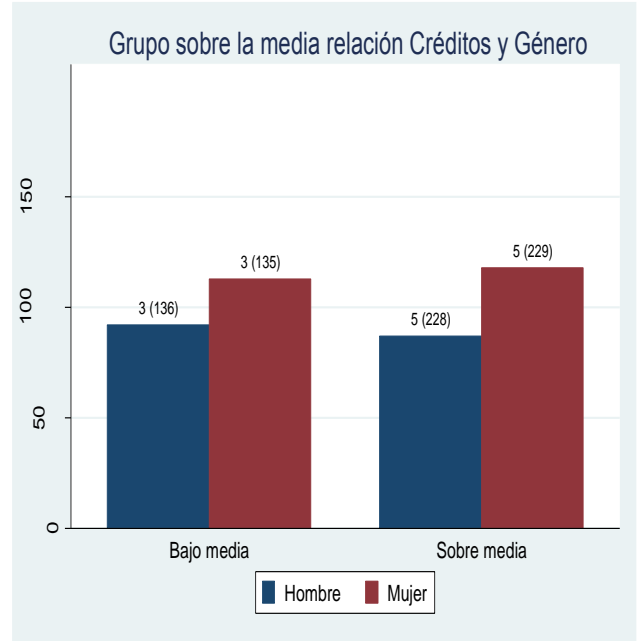
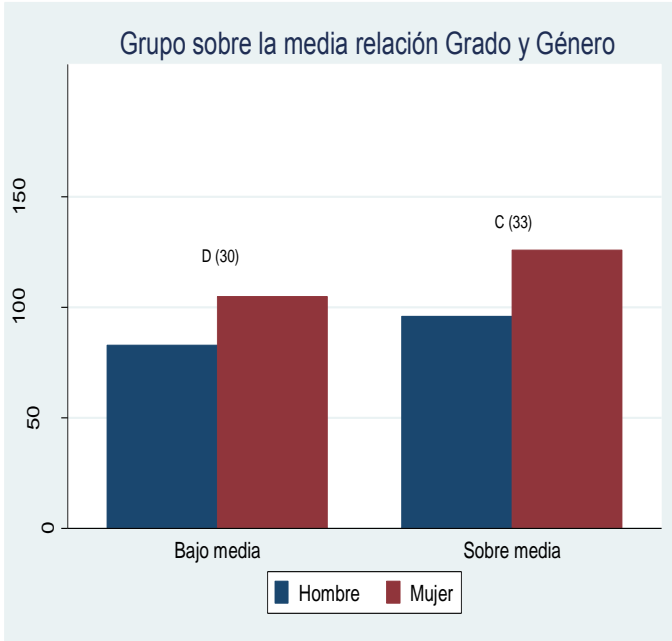


b) Asociación de Grados y créditos con covariables, casos sobre la media

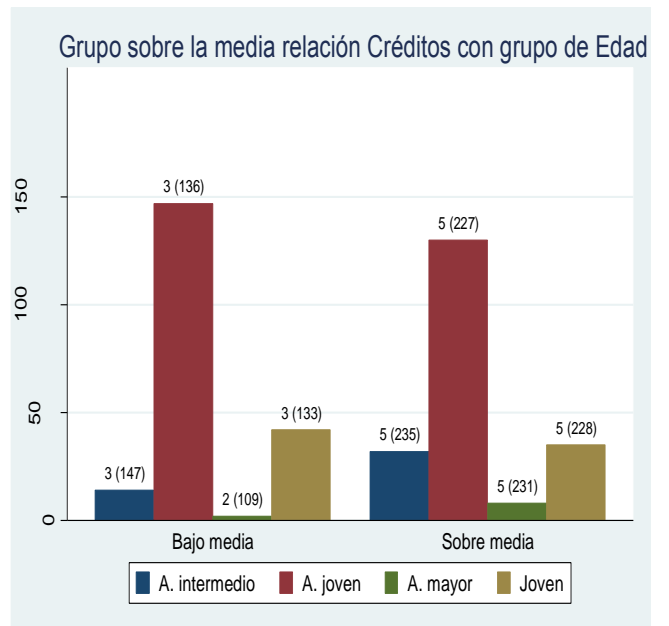
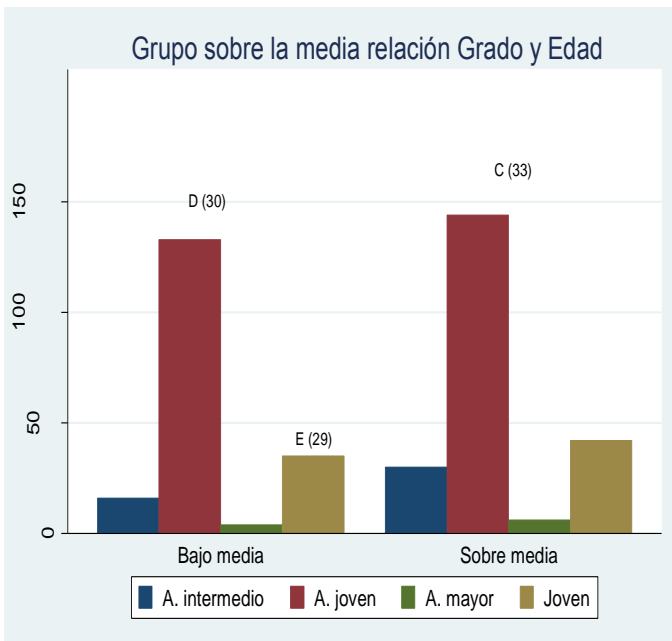
**GRADOS**

**CRÉDITOS**

**Género**



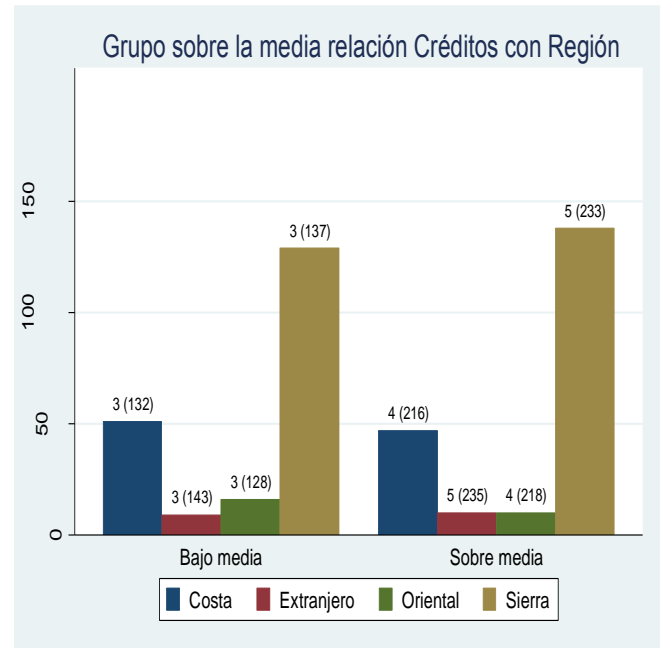
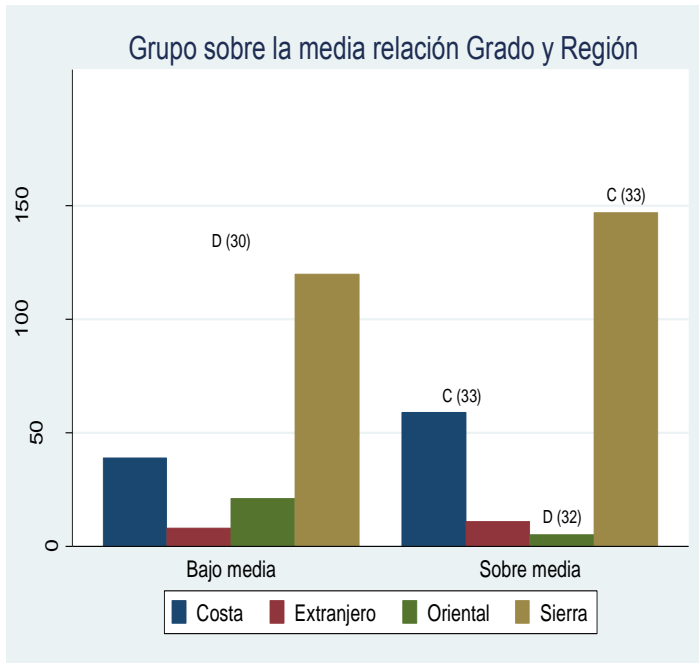
**Edad**



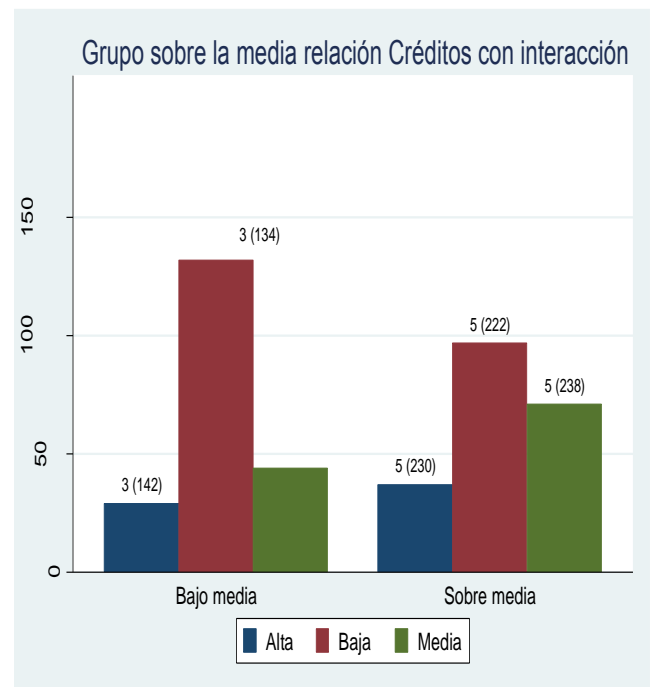
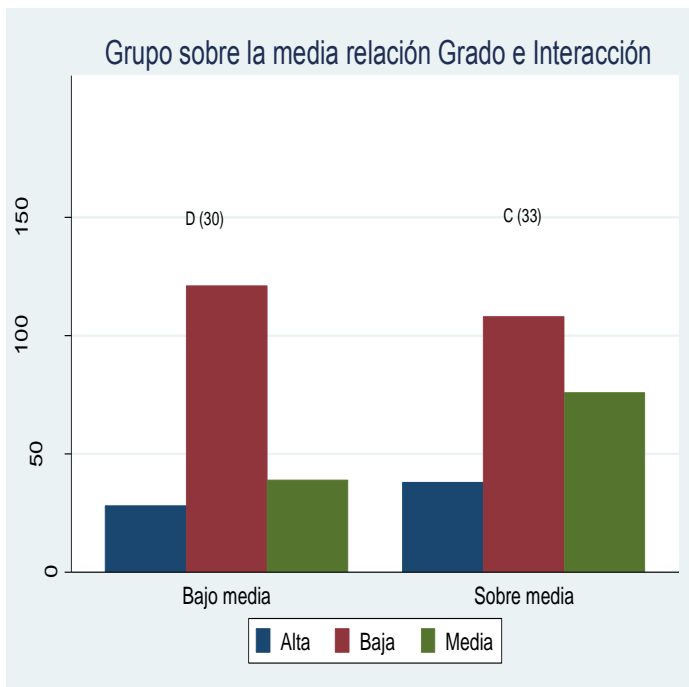
**GRADOS**

**CRÉDITOS**

Región



Interacción



**Anexo 6: Estimación bivariada logística con dos variables de respuesta binarias y varias covariables - Rendimiento académico óptimo**

Deviance: 5541.961

Coefficients	Estimate	Std. Error	z value	Pr(> z )
<b>GRADOS</b>				
<i>Edad</i>				
Adulto		0	base	
Joven	-1.461	0.5555	-2.629	0.008554 **
<i>Género</i>				
Hombre		0	base	
Mujer	-0.3483	0.5430	-0.642	0.521195
<i>Región</i>				
Pequeña		0	base	
Grande	2.6290	1.1050	0.000	0.999810
<i>Participación</i>				
Alta		0	base	
Media	0.3533	0.7426	0.476	0.634287
Baja	-0.4597	0.7512	-0.612	0.540603
(Intercept)	-2.80600	1.1050	0.000	0.999797
<b>CRÉDITOS</b>				
<i>Edad</i>				
Adulto		0	base	
Joven	-0.9587	0.3222	-2.975	0.002929 **
<i>Género</i>				
Hombre		0	base	
Mujer	-0.04604	0.2500	-0.184	0.853889
<i>Región</i>				
Pequeña		0	base	
Grande	-3.83300	1.10500	-34.680	< 2e-16 ***
<i>Participación</i>				
Alta		0	base	
Media	0.4.761	0.3463	1.375	0.169261
Baja	-0.7683	0.3358	-2.288	0.022160 *
(Intercept)	3.83300	1.1050	34.680	< 2e-16 ***
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

## Anexo 7

### Estimación bivariada logística con dos variables de respuesta binarias y varias covariables- Rendimiento académico positivo

Log-likelihood: -497.5817

Deviance: 995.1633

Coefficients	Estimate	Std.Error	z value	Pr(> z )
<b>GRADOS</b>				
<i>Edad</i>				
Adulto	0 base			
Joven	-0.43483	0.30679	-1.417	0.156379
<i>Género</i>				
Hombre	0 base			
Mujer	0.03373	0.20674	0.163	0.870408
<i>Región</i>				
Pequeña	0 base			
Grande	0.84371	0.33426	2.524	0.011600 *
<i>Participación</i>				
Alta	0 base			
Media	0.37473	0.32139	1.166	0.243631
Baja	-0.40467	0.28573	-1.416	0.156702
Intercept	-0.09994	0.48161	-0.208	0.835611
<b>CRÉDITOS</b>				
<i>Edad</i>				
Adulto	0 base			
Joven	-1.07574	0.32224	-3.338	0.000843 ***
<i>Género</i>				
Hombre	0 base			
Mujer	0.16333	0.20764	0.787	0.431526
<i>Región</i>				
Pequeña	0 base			
Grande	0.17499	0.32606	0.537	0.591479
<i>Participación</i>				
Alta	0 base			
Media	0.23615	0.31870	0.741	0.458711
Baja	-0.53088	0.28691	-1.850	0.064266 .
Intercept	0.92308	0.48928	1.887	0.059212 .

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 '' 1