

UNIVERSIDAD POLITÉCNICA DE VALENCIA

**DEPARTAMENTO DE ESTADÍSTICA,
INVESTIGACIÓN OPERATIVA APLICADAS Y
CALIDAD**



**MÁSTER EN INGENIERÍA DE ANÁLISIS DE
DATOS, MEJORA DE PROCESOS Y TOMA DE
DECISIONES**

**ANÁLISIS DEL RENDIMIENTO ACADÉMICO DE
LOS ESTUDIANTES DE LA UNIVERSIDAD DEL
MAGDALENA SEGÚN VARIABLES
SOCIOECONÓMICAS Y FAMILIARES**

TRABAJO FINAL DE MÁSTER

ALUMNO:

DELIMIRO ALBERTO VISBAL CADAVID

DIRECTORA ACADÉMICA:

MÓNICA MARTÍNEZ GÓMEZ

VALENCIA, January de 2019

Contenido

1.	INTRODUCCIÓN.....	6
2.	OBJETIVOS.....	10
2.1	Objetivo General.....	10
2.2	Objetivos Específicos.....	10
3	ESTADO DEL ARTE Y METODOLOGÍA.....	11
3.1	Estado del arte.....	11
3.2	Metodología.....	15
3.2.1	Modelo de Regresión Logística Simple.....	15
3.2.2	Modelo de Regresión Logística Múltiple.....	24
3.2.3	Regresión Logística Ordinal.....	34
3.2.4	Método de selección de variables por paso (Stepwise).....	37
4	MATERIALES Y MÉTODOS.....	40
4.1	Matricula estudiantil.....	40
4.1.1	Distribución por facultad.....	40
4.1.2	Distribución por género.....	41
4.1.3	Procedencia de los estudiantes por departamento.....	44
4.1.4	Información socioeconómica.....	46
4.2	Instrumento.....	47
5	RESULTADOS.....	53
5.1	Análisis de la muestra.....	53
5.2	Modelo de Regresión Logística Ordinal.....	59
6	DISCUSIÓN.....	69
7	CONCLUSIÓN.....	75
8	LINEAS FUTURAS DE INVESTIGACIÓN.....	78
9	BIBLIOGRAFÍA.....	79
10	ANEXOS.....	86

Índice de Tablas

Tabla 1. Codificación de variables dummy.....	25
Tabla 2. Conformación Académica de la Universidad del Magdalena.....	40
Tabla 3. Distribución por facultad.....	41
Tabla 4. Distribución género por facultad.....	42
Tabla 5. Distribución de género por programa.....	44
Tabla 6. Procedencia de los estudiantes según departamento.....	45
Tabla 7. Distribución según estrato socio-económico.....	47
Tabla 8. Variables consideradas en el modelo con su descripción y/o categorías.....	50
Tabla 9. Categorización del Rendimiento Académico.....	53
Tabla 10. Variables significativas mediante AIC y BIC.....	60
Tabla 11. Resultados modelo ajustado mediante AIC.....	61
Tabla 12. Resultados modelo AIC sin variable Edad.....	62
Tabla 13. Información de ajuste de los modelos.....	63
Tabla 14. Prueba de Brant (Lineas Paralelas).....	63
Tabla 15. Pruebas de Bondad de Ajuste del Modelo.....	64
Tabla 16. Pseudo R ²	64
Tabla 17. Tabla de Clasificación o Matriz de Confusión.....	68
Tabla 18. Estudiantes entrevistados según facultad.....	86
Tabla 19. Estudiantes entrevistados según programa académico.....	86
Tabla 20. Estudiantes entrevistados según estrato socio-económico.....	86
Tabla 21. Estudiantes entrevistados según tipo de colegio (Público/Privado).....	87
Tabla 22. Estudiantes entrevistados según provengan de un municipio certificado en educación	87
Tabla 23. Estudiantes entrevistados según género.....	87
Tabla 24. Estudiantes entrevistados según la clasificación ordinal del rendimiento académico.	87
Tabla 25. Estudiantes entrevistados según jornada de su programa académico.....	87
Tabla 26. Estudiantes entrevistados según su estado civil.....	87
Tabla 27. Estudiantes entrevistados según estado civil (Soltero/Otros).....	87
Tabla 28. Estudiantes entrevistados según edad.....	88
Tabla 29. Estudiantes entrevistados según su número de hijos.....	88
Tabla 30. Estudiantes entrevistados según número de personas que dependen económicamente de él.....	88
Tabla 31. Estudiantes entrevistados según con quien viven mientras estudian.....	89
Tabla 32. Estudiantes entrevistados según ocupación de su padre.....	89
Tabla 33. Estudiantes entrevistados según ocupación de su madre.....	89
Tabla 34. Estudiantes según nivel educativo logrado por su padre.....	89
Tabla 35. Estudiantes según nivel educativo logrado por su madre.....	90
Tabla 36. Estudiantes según el origen del dinero para costear sus estudios.....	90
Tabla 37. Estudiantes según si desempeñan o no alguna actividad que les represente ingresos económicos.....	90
Tabla 38. Estudiantes entrevistados según el tiempo transcurrido entre la finalización del bachillerato y el ingreso a la universidad.....	90

Tabla 39. Estudiantes entrevistados según el tipo de bachillerato del cual se graduó.....	91
Tabla 40. Estudiantes entrevistados según la jornada del colegio donde recibió su título de bachiller.....	91
Tabla 41. Estudiantes entrevistados según si han o no interrumpido sus estudios universitarios	91
Tabla 42. Estudiantes entrevistados según si han realizado o no otros estudios.....	91
Tabla 43. Estudiantes entrevistados según si realizan o no otros estudios.....	91
Tabla 44. Estudiantes entrevistados según el grado en que se sientes respetados y orientados por sus profesores.....	91
Tabla 45. Estudiantes entrevistados según sus intereses de continuar estudiando luego de adquirir su título universitario.....	92
Tabla 46. Estudiantes entrevistados según a quien acuden en caso de conflictos personales.....	92
Tabla 47. Estudiantes entrevistados según su relación con su familia.....	92
Tabla 48. Estudiantes entrevistados según su relación con sus amigos.....	92
Tabla 49. Estudiantes entrevistados según su relación con la Universidad del Magdalena.....	92
Tabla 50. Estudiantes entrevistados según su relación con la ciudad de Santa Marta.....	93
Tabla 51. Estudiantes entrevistados según su relación respecto a Colombia.....	93
Tabla 52. . Estudiantes entrevistados según las actividades desarrolladas en su tiempo libre....	93
Tabla 53. Estudiantes entrevistados según su pertenencia a grupos.....	93
Tabla 54. Apreciación hacia la calidad de los servicios de bienestar estudiantil.....	93

Índice de Figuras

Figura 1. Función logística univariada.....	19
Figura 2. Relación lineal entre Logit(X) y X.....	19
Figura 3. Dendograma Estado Civil.....	49
Figura 4. Dendograma Ocupación Padre.....	49

1. INTRODUCCIÓN

Actualmente una de las prioridades a nivel mundial es poder garantizar la calidad en la educación superior, conforme a lo establecido en la Declaración Mundial sobre Educación Superior para el Siglo XXI: Visión y Acción, en su artículo once establece la importancia de la calidad en la educación superior y el desarrollo de mecanismos para que se garantice (UNESCO, 1998).

En este contexto, el rendimiento académico de los estudiantes universitarios constituye un factor imprescindible y fundamental para la valoración de la calidad educativa en la enseñanza superior (de Miguel et al., 2002). Se ha convertido en un indicador estratégico que proporciona una sólida información para la toma de decisiones en educación superior, especialmente a nivel institucional, ya que, en la actualidad, la calidad de la educación superior se relaciona con criterios de financiamiento, capacitación del personal académico, evaluación, innovación, alternativas educativas, cooperación interinstitucional y regional (Verdugo Saldivar y Urias Montes, 2007).

La Declaración de la Conferencia Regional de la Educación Superior en América Latina y el Caribe (CRES), celebrada en Cartagena de Indias (Colombia) en el año 2008, definió a la educación superior como “un bien público y social, un derecho humano y universal, y un deber del Estado”. Los Estados tienen el deber fundamental de garantizar este derecho. “Los Estados, las sociedades nacionales y las comunidades académicas deben ser quienes definan los principios básicos en los cuales se fundamenta la formación de los ciudadanos y ciudadanas, velando para que ella sea pertinente y de calidad. (IESALC, 2008).

Los estudios del Rendimiento Académico en la educación superior se han convertido en una necesidad en la actual coyuntura mundial, debido al dinamismo que experimenta el sector universitario.

Centrándonos en la educación colombiana, las instituciones públicas de educación superior son conscientes del reto al que se enfrentan de mejorar su calidad académica, lo que conlleva una fuerte preocupación por mejorar los procesos y productos.

El debido aprovechamiento y desarrollo del conocimiento de los estudiantes universitarios en Colombia es uno de los temas que despierta el interés en las instituciones universitarias del país, con el propósito de orientar políticas y desarrollar las acciones institucionales pertinentes en las diversas áreas de la comunidad estudiantil.

Entre estos intereses es de particular importancia el diseño de estrategias para mejorar el desempeño académico de los estudiantes, apoyados en estudios cuantitativos basados en técnicas estadísticas. Sin embargo, delimitar el concepto y ámbito de aplicación del término “rendimiento académico” no ha sido tarea fácil, dado el carácter complejo y multidimensional que da cuerpo a esta variable en el área educativa (Montero, Villalobos & Valverde, 2007). Existen multitud de estudios sobre los factores que pueden afectar al rendimiento académico, pero casi todos ellos se centran en dimensiones institucionales o pedagógicas. Muy pocos tienen en cuenta la situación propia del estudiante respecto sus condiciones demográficas, socioeconómicas, afectivos o psicosociales entre otras.

En este sentido, de Miguel et al., 2002, consideran que el rendimiento académico de los alumnos universitarios constituye un factor ineludible en el abordaje del tema de la calidad de la educación superior, pues se considera que es un indicador que permite una aproximación a la realidad educativa.

El bajo rendimiento académico es uno de los determinantes del abandono en las enseñanzas universitarias, e incluso cuando no se considera el abandono escolar, el bajo rendimiento y la repetición de cursos provocan dificultades para la institución y para la población estudiantil (Montero et al., 2007), conformándose por consiguiente en un factor determinante de la eficacia del sistema educativo.

Analizar los elementos relacionados con el desempeño de los estudiantes universitarios se convierte en un factor importante a la hora de diseñar políticas institucionales en pro de la mejora de los procesos educativos universitarios (Rodríguez, 2014). Por ello, convendría plantearse ¿qué sabemos de los estudiantes? ¿Cuáles son los factores que influyen en la consecución de sus metas académicas?

Los requerimientos de la sociedad actual demandan que el quehacer de la universidad se caracterice por un amplio contenido social y humanístico del estudiante, donde no prime una formación de alta capacidad técnica, sino una formación basada fundamentalmente

en principios de libertad y justicia. Para tal propósito, es indispensable determinar las características y situaciones en las que vive y se encuentra actualmente el estudiante universitario, teniendo presente los aspectos demográficos, sociales, afectivos, económicos, académicos, recreativos y ambientales.

La Asociación Colombiana de Universidades (ASCUN) emprendió a finales de la década de los años 80, un programa tendiente a definir las características del estudiante universitario colombiano y a impulsar la organización de un sistema de información sobre su perfil.

Para el caso específico que nos ocupa, la Universidad del Magdalena en la que se desarrolla el presente trabajo, cabe destacar la diversificación de los orígenes sociales de donde provienen sus estudiantes: aunque son en su mayoría de la ciudad de Santa Marta, el número de estudiantes provenientes de otros municipios del departamento, de otros departamentos de la costa atlántica y del interior del país ha ido en aumento a consecuencia de la consolidación que en los últimos años ha vivido nuestra alma mater. Esto, trae consigo que muchos estudiantes de provincia no se adapten, o su proceso de acomodamiento se torne difícil, reflejándose obviamente en su desempeño académico y en el peor de los casos en la deserción.

El presente trabajo tiene como finalidad fundamental esclarecer cuáles son los factores que, de una u otra manera, están relacionados con el rendimiento académico de los estudiantes de la Universidad del Magdalena mediante un modelo de Regresión Logística Ordinal (RLO), no ya con fines predictivos, sino más bien, con fines explicativos.

De hecho, la principal diferencia entre un modelo explicativo y uno predictivo está en los objetivos del análisis. Si lo que se desea es realizar una intervención con miras a la mejora de la situación estudiada, la explicación es mucho más importante que la predicción. Si deseamos tomar medidas para alterar un resultado, lo mejor es tratar de determinar que variables, dentro de un grupo amplio de posibilidades, están incidiendo y de qué manera, en el fenómeno bajo estudio. Los resultados de un modelo explicativo, le dirá al investigador cómo intervenir, es decir, en qué variables explicativas debemos actuar. Por otro lado, si el interés radica en “ver” cómo será el comportamiento futuro

de una variable, sin ninguna intención (o capacidad) de intervenir, es más probable que sea apropiado un modelo predictivo.

Existen distintas publicaciones que abordan el estudio del rendimiento académico desde distintas perspectivas. En este sentido Gómez, Oviedo, y Martínez, (2011) afirma:

“Los factores que inciden en el rendimiento académico se clasifican de diversas maneras, a saber: variables demográficas o de identificación (sexo, edad, estado civil, experiencia laboral), variables académicas (tipos de estudios cursados, curso, opción en que se estudia una carrera, rendimiento previo), variables sociofamiliares (estudios de los padres, situación laboral de los mismos, lugar de residencia familiar, lugar de estudio).” (p.91)

Los indicadores de Rendimiento Académico que se considerarán en nuestro estudio se centrarán en analizar la influencia de factores demográficos, sociales, económicos, emocionales, cognitivos, uso del tiempo libre, en el rendimiento académico del estudiante de la Universidad del Magdalena.

La determinación de estas características en los estudiantes de la Universidad del Magdalena, y el grado de influencia de éstos en el desempeño académico pueden ser utilizados para el diseño e implementación de políticas en el programa de Bienestar Universitario y/o de las instancias institucionales internas que las requieran, orientando directrices claras y objetivas para mejorar y fortalecer los programas académicos, reforzando, apoyando y guiando al estudiante de acuerdo a sus motivaciones, expectativas y situaciones personales, para así alcanzar la integralidad en cuanto a lo académico, personal y profesional.

2. OBJETIVOS

2.1 Objetivo General

De acuerdo a lo expuesto anteriormente, el objetivo fundamental del presente trabajo es determinar los factores demográficos, familiares, afectivos, académicos, personales, socio-económicos y ambientales que se relacionan con el rendimiento académico de los estudiantes de la Universidad del Magdalena.

2.2 Objetivos Específicos

Para la consecución del objetivo general del trabajo, se puede descomponer en los siguientes objetivos específicos:

1. Identificar si el género, estado civil, número de hijos, número de personas a cargo, jornada de estudio y procedencia del estudiante de la Universidad del Magdalena se relacionan con su Rendimiento Académico.
2. Determinar cómo se relaciona la formación educativa de los padres, ocupación de los padres, el desempeño académico de los estudiantes de la Universidad del Magdalena.
3. Identificar cómo se relaciona, en los estudiantes de la Universidad del Magdalena, el estrato socioeconómico, dependencia laboral, financiación de matrícula y gastos educativos con el Rendimiento Académico.
4. Analizar si el tipo de colegio de procedencia, tiempo transcurrido entre la finalización de sus estudios de bachillerato y el ingreso a la universidad se relacionan con el Rendimiento Académico.
5. Y finalmente, se determinará si otras variables como los sentimientos respecto a su familia, a su universidad, a la ciudad, al país, el respeto y orientación de sus profesores, el género y la utilización del tiempo libre de los estudiantes de la Universidad del Magdalena, pueden estar relacionados con su desempeño académico.

3 ESTADO DEL ARTE Y METODOLOGÍA

3.1 Estado del arte

Hasta el momento es muy escasa la atención que se le presta en la Universidad del Magdalena a las características del estudiante universitario, y al efecto que dichas características tienen sobre su desempeño académico. De ahí la importancia de realizar la presente investigación, cuyos resultados permitirán a la institución una orientación más certera en el desarrollo académico, personal y profesional de cada estudiante.

El estudio que se pretende efectuar con los estudiantes de los diferentes programas de la Universidad del Magdalena ayudará a planear una educación que concuerde con los rasgos y necesidades específicas de la población universitaria. Existen diversos estudios similares en distintas universidades del mundo (de Miguel et al., 2002; Garbanzo, 2007; Shahzadi y Ahmad, 2011; Rodríguez, 2014; Mora, 2015; Redondo, Corrás, Novo, y Fariña, 2017; Fajardo, Maestre, Felipe, León y Polo, 2017; Tomás-Miquel, Expósito-Langa, y Sempere-Castelló, 2014; García, Alvarado, y Jiménez, 2000; Montero, et al., 2007; Gómez et al. 2011; Heredia, Rodríguez, y Vilalta, (2012), 2012; Tourón, 1985), entre otros.

Centrándonos en las áreas de las ciencias sociales, ciencias de la educación y estadística aplicada, es habitual hallar investigaciones que tienen como propósito analizar las variables o factores que contribuyen al rendimiento académico de los estudiantes de todos los niveles educativos. Algunas de estas investigaciones buscan establecer la importancia de dichos factores (de Miguel et al., 2002; Garbanzo, 2007; Shahzadi y Ahmad, 2011; Gómez et al., 2011; Rodríguez, 2014; Mora, 2015; Redondo et al., 2017), y otras intentan predecir el desempeño futuro de los estudiantes con base a estos factores (Fajardo et al., 2017; Tomás-Miquel et al., 2014; García et al., 2000; Montero et al., 2007; Heredia et al., 2012; Mora, 2015; Tourón, 1985). De esta manera, podemos volver a reafirmar que existe un verdadero interés por parte de los investigadores, por establecer los factores que influyen de una u otra manera, en el desempeño estudiantil.

Según Fajardo et al., (2017), “hasta la década de los años sesenta, las causas de los bajos rendimientos se buscaban exclusivamente en el alumno; además, no existía una preocupación sobre si el alumno rendía y sobre el porqué de su fracaso” (p.212). De esta manera, siguiendo a Delgado (1994), citado en Fajardo et al., (2017), “el bajo rendimiento académico solía explicarse de una manera muy práctica y tranquilizante; el alumno era vago o tonto; o no tiene capacidad suficiente” (p.212). Sin embargo, con el avance tecnológico y en especial con el de los sistemas computacionales y las herramientas estadísticas fácilmente asequibles a los investigadores, se ha venido estudiando la influencia de factores tanto internos como externos del estudiante y su posible influencia en su desempeño académico.

Existe mucha controversia en cómo medir el rendimiento académico. Para Tomás-Miquel et al., (2014) el concepto del rendimiento académico es un término complejo e interpretado de distintas maneras según el significado que tiene para las diferentes audiencias implicadas. Y según González-López, (2004), citado por Tomás-Miquel et al., (2014), se considera un concepto multidimensional, relativo y contextual, del que es difícil aportar un criterio que lo defina y sea aceptado por todos.

Garbanzo (2007), considera el rendimiento académico como 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. Según Pérez, Ramón y Sánchez (2000) y Vélez y Roa (2000), citado en Garbanzo (2007), mediante las calificaciones obtenidas, con una valoración cuantitativa, se muestran los resultados de las materias ganadas o perdidas, la deserción y el grado de éxito académico.

En este mismo orden de ideas, Pita y Corengia (2005), citado en Gómez et al., (2011) y en Tomás-Miquel et al., (2014), razona que “el rendimiento académico no es el producto de una única capacidad, sino que se percibe como el resultado sintético de una serie de factores que actúan en, y desde, la persona que aprende. Puede afirmarse, en términos educativos, que el rendimiento académico es un resultado del aprendizaje suscitado por la actividad educativa del profesor y producido en el alumno, aunque es claro que no todo aprendizaje

es producto de la acción docente. Se le expresa en una calificación cuantitativa y cualitativa, una nota que si es consistente y válida será el reflejo de un determinado aprendizaje y del logro de unos objetivos preestablecidos” (p.90 y p.381).

Por su parte Artunduaga (2008), citado en Tomás-Miquel et al., (2014), indica que:

“Los factores relevantes para el rendimiento académico de los estudiantes universitarios pueden estar relacionados con variables personales del individuo, como las demográficas (edad, sexo, experiencia laboral, etc.), las cognoscitivas (aptitudes, rendimiento académico previo, capacidades y habilidades, etc.) y las actitudinales (interés, autoconcepto, habilidades sociales, etc.), y con variables contextuales como las pedagógicas (entre otras, actitud, formación, experiencia y personalidad del profesor, tamaño del grupo, clima de la clase, etc.), institucionales (tipo y tamaño del centro y políticas educativas), de entorno social, familiar y cultural (por ejemplo, nivel educativo del padre y de la madre, clima educativo familiar, integración social del estudiante, etc.), entre las más destacables”. (p.281)

En la misma línea Redondo et al., (2017) va un poco más allá y evalúa cómo distintas variables psicoemocionales (entre ellas las expectativas, el optimismo y la autoeficacia), se relacionan o afectaban el rendimiento académico del alumnado en la enseñanza superior.

De Miguel et al., (2002) ponen de manifiesto la posibilidad de utilizar distintos indicadores para definir el rendimiento académico, como por ejemplo las calificaciones académicas, el número de créditos aprobados o las tasas de rendimiento entre otras, y que adicionalmente cada uno de ellos puede a su vez ser matizado por índices correctores, como número de convocatorias utilizadas, total de créditos matriculados, tipología de los créditos, número de cursos de la carrera, etc.

Garbanzo (2007) considera que “por sus características, hay consenso en la comunidad de especialistas de lo difícil que resulta identificar el rendimiento

académico en la educación superior, debido a que es problemático y confuso identificar el rendimiento académico con las notas” (p46).

García (1989) manifiesta que existe un inconformismo más o menos acusado en cuanto al uso de las calificaciones otorgadas por los profesores al estudiante como proxy del rendimiento académico debido a sus numerosas limitaciones, pero, aun así, en la práctica no se encuentra un mejor criterio. De este modo, según Tejedor (1998), citado por Tomás-Miquel et al., (2014), “desde una perspectiva operativa del término, se define como la nota o calificación media obtenida durante el periodo universitario que cada alumno haya cursado”. Observamos, pues, que aún hoy en día, en todos los niveles educativos, se consideran las calificaciones como el ente que guía todo el proceso evaluativo del estudiante, convirtiéndose en un indicador directo de la calidad académica y del proceso educativo.

Para el desarrollo de este trabajo seguiremos el concepto de Rendimiento Académico en el que se enmarca el trabajo de Montero et al. (2007), el cual está basado en el propuesto por Tournon (1984), citado por Montero et al (2007), en este se establece que:

“El rendimiento académico es un resultado del aprendizaje, suscitado por la intervención pedagógica del profesor o la profesora, y producido en el alumno. No es el producto analítico de una única aptitud, sino más bien el resultado sintético de una suma (nunca bien conocida) de elementos que actúan en, y desde la persona que aprende, tales como factores institucionales, pedagógicos, psicosociales y sociodemográficos”. (p.24)

Teniendo en cuenta lo anterior, este trabajo forma parte de un estudio más amplio sobre la caracterización del estudiante de la Universidad del Magdalena y las implicaciones que estas características tienen sobre el rendimiento académico. Para ello, se analizará si existe relación entre el rendimiento, es decir, el promedio académico ponderado de los estudiantes, y distintos factores relacionados con aspectos demográficos, sociales, afectivos, económicos, académicos, recreativos y ambientales.

Entre los distintos estudios encontrados en la revisión bibliográfica realizada, se han identificado distintas técnicas estadísticas para el análisis del Rendimiento Académico entre las que podemos destacar la Regresión Lineal (García et al. 2000; Rodríguez y Ruíz, 2011; Mora, 2015), los Modelos Jerárquicos Lineales (Montero et al., 2007; Rodríguez, 2014), Modelos de Ecuaciones Estructurales (Shahzadi y Ahmad, 2011; Kim y Nembhard, 2017), Análisis de Varianza (García, 1989; Gómez et al., 2011, Fajardo et al., 2017), Regresión Logística (García et al. 2000; Ibarra y Michalus, 2010; Heredia et al., 2012; Cortez, Tutiven y Villavicencio, 2017), Diseños Factoriales de Efectos Fijos (de Miguel et al., 2002), Análisis Discriminante (Tomás-Miquel, et al., 2014), Análisis Factorial Confirmatorio (Fernández, 2015) y Análisis Correlacional (Redondo et al., 2017). De esta manera, se pretende poner de manifiesto la diversidad de modelos y métodos estadísticos que se han utilizado en el estudio del Rendimiento Académico.

En nuestro caso, se ha decidido optar por la Regresión Logística Ordinal, debido a que el interés recae en determinar las variables asociadas a la cualidad del estudiante por obtener determinado nivel de Rendimiento Académico, más no en la calificación numérica exacta del estudiantado.

3.2 Metodología

Como se ha comentado anteriormente, el objetivo del presente trabajo vislumbrar cuáles son los factores que de una u otra manera están relacionados con el Rendimiento Académico de los estudiantes de la Universidad del Magdalena mediante un modelo de Regresión Logística Ordinal (RLO) con fines explicativos y no predictivos.

Para la aplicación de esta técnica y debido a que algunas variables explicativas tienen varias categorías y muy pocas observaciones en algunas de estas categorías, se procedió a disminuir el número de categorías agrupando individuos por características homogéneas mediante análisis clúster.

3.2.1 Modelo de Regresión Logística Simple

Hair, Andersen, Tathan y Black (2010), consideran que la regresión lineal múltiple es la técnica de dependencia multivariante más extensamente utilizada.

Siendo el principal factor de la popularidad de la regresión lineal su capacidad para explicar y predecir las variables métricas. Sin embargo, en contextos donde la variable dependiente es no métrica, es decir las variables son categóricas nominal u ordinal, la regresión lineal multivariante ya no es adecuada. En este contexto la regresión logística es una de las técnicas más utilizadas.

En este mismo sentido, Hosmer, Lemeshow y Sturdivant (2013), consideran importante entender que el objetivo de un análisis que utiliza el modelo de regresión logística es similar al de cualquier otro modelo de regresión utilizado en estadística, es decir, encontrar el modelo más apropiado y el más parsimonioso, que sea fácilmente interpretable para describir la relación entre una variable dependiente o respuesta y las variables independientes (explicativas). Las variables independientes a menudo se llaman covariables.

En la regresión lineal, el método más utilizado para la estimación de los parámetros desconocidos es el de mínimos cuadrados. Sin embargo, cuando el método de mínimos cuadrados se aplica a un modelo con un resultado dicotómico, los estimadores no poseen las propiedades estadísticas deseables, esto es, que sea insesgado, es decir, que su esperanza matemática sea igual al parámetro que trata de estimar, y que sea eficiente, es decir, que su varianza sea lo más pequeña posible. Por ello, cuando la variable respuesta es dicotómica el método empleado para estimar los parámetros desconocidos es el de máxima verosimilitud.

La regresión logística cuenta con amplias aplicaciones en situaciones donde el objetivo es identificar la categoría al cual un objeto (por ejemplo, una persona, una empresa, o un producto) pertenece, incluyendo, por ejemplo, la predicción de éxitos o fracasos de un nuevo producto, decidir si un estudiante debe ser admitido en una universidad, clasificar a los estudiantes por sus intereses vocacionales, determinar en qué categoría de riesgo de crédito se encuentra una persona o predecir si una empresa tendrá éxito o no. En cada caso los objetos están incluidos en grupos o categorías y se desea poder explicar o predecir la pertenencia a cada grupo mediante un conjunto de variables independientes seleccionadas por el investigador.

El modelo de regresión logística simple permite estimar la probabilidad de un evento, indicado por una variable Y , siendo $Y=1$ cuando el evento ocurre y $Y=0$ cuando no ocurre. El modelo es del tipo simple pues incluye sólo una variable explicativa X , siendo ésta de tipo cuantitativa, nominal u ordinal. Para representar la media condicional de Y dado x_i , $E(Y|x_i)$, el modelo de regresión logística es dado por:

$$E(Y|x_i) = P(Y=1|X=x_i) = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} = P(x_i) \quad (1)$$

Dónde:

β_0 : Parámetro de intercepto

β_1 : Parámetro de regresión

e : Número de Euler, es aproximadamente igual a 2,71828 y es la base de los Logaritmos Naturales

$Y=1$: Indica la ocurrencia del evento (respuesta)

X : Variable explicativa

Una transformación de $P(Y=1|X=x_i)$ que es fundamental para nuestro estudio de regresión logística es la *transformación logit*, la cual es la más utilizada por su facilidad de manejo, al transformarse en un modelo lineal sobre la variable logit. Esta transformación se define, en términos de $P(Y=1|X=x_i)$, como:

$$\text{Logit} \left(\frac{P(Y=1|X=x_i)}{P(Y=0|X=x_i)} \right) = \ln \left(\frac{P(Y=1|X=x_i)}{P(Y=0|X=x_i)} \right) = \beta_0 + \beta_1 x_i = \frac{P(x_i)}{1 - P(x_i)} \quad (2)$$

En donde

$$P(Y=0|X=x_i) = \frac{1}{1 + e^{\beta_0 + \beta_1 x_i}} \quad (3)$$

La variable “*logit*” representa, en escala logarítmica, la diferencia entre las probabilidades de pertenecer a cada categoría. Al cociente $\frac{P(Y=1|X=x_i)}{P(Y=0|X=x_i)}$ se le conoce como “*odd*” en el lenguaje de apuestas, que suele traducirse por

“riesgo”, y no es más que el cociente entre la probabilidad de “éxito” sobre la probabilidad de “fracaso”.

La importancia de esta transformación es que tiene muchas de las propiedades deseables de un modelo de regresión lineal, como por ejemplo la facilidad en la interpretación de los parámetros del modelo, por un lado, y la linealidad de sus parámetros, que puede ser continuo y puede variar de $-\alpha$ a $+\alpha$, dependiendo del rango de X , por otro.

Otra diferencia importante entre los modelos de regresión lineal y logística se refiere a la distribución condicional de la variable respuesta. En el modelo de regresión lineal suponemos que una observación de la variable respuesta puede expresarse como $y = E(Y|x) + \varepsilon$. La cantidad ε se llama *residuo* y expresa la desviación de una observación de la media condicional. La suposición más común es que ε sigue una distribución normal con media cero y varianza constante. Se deduce que la distribución condicional de la variable de resultado dada x es normal con la media $E(Y|x)$, y una varianza que es constante. Este no es el caso con una variable de resultado dicotómica. En esta situación, podemos

expresar el valor de la variable respuesta, Y , dado x como $y = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} + \varepsilon$. Aquí

la cantidad ε puede asumir uno de dos valores posibles. Si $y=1$ entonces

$\varepsilon = 1 - \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$ con probabilidad $\frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$, y si $y=0$ entonces $\varepsilon = \frac{-e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$ con

probabilidad $1 - \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$. Por lo tanto, ε tiene una distribución con media cero y

varianza igual a $\left(\frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \right) \left(1 - \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \right)$. Es decir, la distribución condicional

de la variable respuesta sigue una distribución binomial con probabilidad dada

por la media condicional, $\frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$.

Cuando la variable respuesta es politómica con L categorías de respuestas, el modelo anterior se ajusta en dos categorías a la vez y se formarán $L-1$ logits.

Si se considera el caso en que la variable X es cuantitativa y se representan los

valores de $P(Y=1|X=x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$ en función de los valores de X en un sistema

de ejes cartesianos, se podrá comprobar que la expresión gráfica del modelo, por ejemplo, cuando $\beta_0=1.6$, $\beta_1=1.3$ y $-6 \leq X \leq 6$ es como se muestra en la figura 1. Por su parte, la figura 2 muestra la relación lineal entre $\text{logit}(X)$ y X para los mismos valores de β_0 y β_1 , en el mismo intervalo de valores para X .

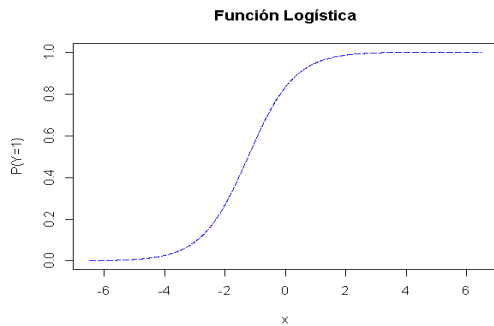


Figura 1. Función logística univariada
Fuente: Elaboración Propia

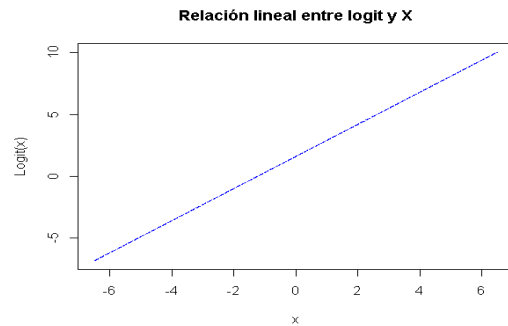


Figura 2. Relación lineal entre Logit(X) y X
Fuente: Elaboración Propia

3.2.1.1 Estimación del modelo logístico binario

Supongamos que tenemos una muestra aleatoria de n observaciones independientes del par $(x_i, y_i), i=1, 2, \dots, n$, donde y_i denota el valor de una variable respuesta dicotómica y x_i es el valor de la variable independiente para la i -ésima observación. Además, supongamos que la variable respuesta se ha codificado como 0 o 1, representando la ausencia o la presencia de la característica, respectivamente. Ajustar el modelo de regresión logística de la ecuación (1) a un conjunto de datos requiere la estimación de los parámetros desconocidos β_0 y β_1 .

Como se ha comentado previamente, cuando la variable dependiente es dicotómica el método de estimación más apropiado es el método de máxima verosimilitud que proporciona valores para los parámetros desconocidos que maximizan la probabilidad de obtener el conjunto de datos observado. Para aplicar este método, primero debemos

construir una función, llamada función de verosimilitud. Esta función expresa la probabilidad de los datos observados como una función de los parámetros desconocidos. Los estimadores de máxima verosimilitud de los parámetros son los valores que maximizan esta función. Por lo tanto, los estimadores resultantes son los que están más de acuerdo con los datos observados. A continuación, se describe cómo encontrar estos valores para el modelo de regresión logística:

Si Y se codifica como 0 o 1, la expresión para $P(x)$ dada en la ecuación (1) proporciona (para un valor arbitrario de $\beta = (\beta_0, \beta_1)$, el vector de parámetros) la probabilidad condicional de que Y sea igual a 1 dado x . Se deduce que la cantidad $1 - P(x)$ da la probabilidad condicional de que Y es igual a cero, habiéndose dado x , $P(Y = 0|x)$. Por lo tanto, para esos pares (x_i, y_i) , donde $y_i = 1$, la contribución a la función de verosimilitud es $P(x_i)$, y para aquellos pares donde $y_i = 0$, la contribución a la función de verosimilitud es $1 - P(x_i)$, donde la cantidad $P(x_i)$ denota el valor de $P(x)$ calculado en x_i . Una forma conveniente de expresar la contribución a la función de verosimilitud para el par (x_i, y_i) , es a través de la expresión

$$P(x_i)^{y_i} [1 - P(x_i)]^{1 - y_i} \quad (4)$$

Como se supone que las observaciones son independientes, la función de verosimilitud se obtiene como el producto de los términos dados en la ecuación (4), de la siguiente manera:

$$l(\beta) = \prod_{i=1}^n P(x_i)^{y_i} [1 - P(x_i)]^{1 - y_i} \quad (5)$$

El principio de máxima verosimilitud establece que se utiliza como nuestra estimación de β , el valor que maximiza la expresión en la ecuación (5). Sin embargo, matemáticamente es más fácil trabajar con el logaritmo natural de la ecuación (5). Esta expresión, $\ln[l(\beta)]$, se define como

$$L(\beta) = \ln[l(\beta)] = \sum_{i=1}^n \ln \left[P(x_i)^{y_i} [1 - P(x_i)]^{1 - y_i} \right]$$

Para encontrar el valor de β que maximiza $L(\beta)$, diferenciamos $L(\beta)$ con respecto a β_0 y β_1 , y establecemos las expresiones resultantes igual a cero. Estas ecuaciones, conocidas como ecuaciones de verosimilitud, quedarían

$$\sum_{i=1}^n [y_i - P(x_i)] = 0 \quad (7)$$

y

$$\sum_{i=1}^n x_i [y_i - P(x_i)] = 0 \quad (8)$$

Las expresiones en las ecuaciones (7) y (8) son no lineales en β_0 y β_1 , requiriendo métodos iterativos no lineales para su solución, como *Newton-Raphson*. (Silva y Barroso, 2004, P.57).

El valor de β dado por la solución a las ecuaciones (7) y (8) se denomina *estimación de máxima verosimilitud* y se denota como $\hat{\beta}$. En general, el uso del símbolo " $\hat{\cdot}$ ", denota la estimación de máxima verosimilitud de la cantidad respectiva.

Por ejemplo, $\hat{P}(x_i)$ es la estimación de máxima verosimilitud de $P(x_i)$. Esta cantidad proporciona la estimación de la probabilidad condicional de que Y es igual a 1, dado que x es igual a x_i . Como tal, representa el valor ajustado o predicho para el modelo de regresión logística. Una consecuencia interesante de la ecuación (7) es que

$$\sum_{i=1}^n y_i = \sum_{i=1}^n \hat{P}(x_i) \quad (9)$$

La ecuación (9) indica que la suma de los valores observados de y , es igual a la suma de los valores predichos (esperados).

3.2.1.2 Evaluación de la significación de los coeficientes β_0 y β_1 .

Después de estimar los coeficientes, el primer criterio de validación del modelo ajustado, se refiere comúnmente a evaluar la importancia de las variables en el modelo. Es decir, analizar si las variables independientes en el modelo son significativas.

Un enfoque para probar la importancia del coeficiente de una variable en cualquier modelo se relaciona con la siguiente pregunta. ¿El modelo que incluye la variable en cuestión nos dice más sobre la variable respuesta que un modelo que no incluye esa variable? Esta pregunta se responde comparando los valores observados de la variable respuesta con los predichos por cada uno de los dos modelos; el primero con, y el segundo sin, la variable en cuestión. La función matemática utilizada para comparar los valores observados y predichos depende del problema particular. Si los valores predichos con la

variable en el modelo son mejores, o más precisos en cierto sentido, que cuando la variable no está en el modelo, entonces se puede considerar que la variable en cuestión es "significativa".

En regresión logística la comparación de los valores observados con los predichos se realiza a través de la función de verosimilitud y se basa en la siguiente expresión, llamada *devianza*.

$$D = -2 \ln \left[\frac{\text{Verosimilitud del modelo ajustado}}{\text{Verosimilitud del modelo saturado}} \right] \quad (10)$$

La cantidad dentro de los corchetes en la expresión anterior se llama *razón de verosimilitud*.

$$D = -2 \sum_{i=1}^n \left[y_i \ln \left(\frac{\hat{P}(x_i)}{y_i} \right) + (1 - y_i) \ln \left(\frac{1 - \hat{P}(x_i)}{1 - y_i} \right) \right] \quad (11)$$

El estadístico D , la *devianza*, para la regresión logística, juega el mismo papel que la suma de cuadrados residual (SSR) en la regresión lineal. De hecho, la devianza como se muestra en la ecuación (11), cuando se calcula para la regresión lineal, es idéntica a la SSR (Hosmer et al., 2013).

En situaciones donde los valores de la variable respuesta es 0 o 1, la verosimilitud del modelo saturado es igual a 1.0.

Por lo tanto, se deduce de la ecuación (10) que la *devianza* es

$$D = -2 \ln [\text{Verosimilitud del modelo ajustado}] \quad (12)$$

La devianza juega un papel fundamental para verificar la eficacia del ajuste. En la hipótesis nula, se prueba que en grandes muestras (prueba asintótica) la estadística D sigue una distribución de chi-cuadrado con p grados de libertad (Hosmer et al., 2013).

En particular, para evaluar la significancia de una variable independiente, comparamos el valor de D con y sin la variable en la ecuación. El cambio en D debido a la inclusión de la variable independiente en el modelo es

$$G = D \dot{}$$

Dado que la verosimilitud del modelo saturado es común para ambas situaciones, G puede expresarse como:

$$G = -2 \ln \left[\frac{\text{Verosimilitud del modelo sin la variable}}{\text{Verosimilitud del modelo con la variable}} \right] \quad (14)$$

Para el caso específico de una sola variable independiente, es fácil mostrar que cuando

la variable no está en el modelo, la estimación de máxima verosimilitud de β_0 es $\ln \left(\frac{n_1}{n_2} \right)$

donde $n_1 = \sum y_i$ y $n_0 = \sum (1 - y_i)$ y la probabilidad pronosticada para todas las

observaciones es constante e igual a $\frac{n_1}{n_1 + n_0}$. En esta configuración, el valor de G es:

$$G = -2 \ln \left[\frac{\left(\frac{n_1}{n_1} \right)^{n_1} \left(\frac{n_0}{n_1} \right)^{n_0}}{\prod_{i=1}^n \hat{P}(x_i)^{y_i} (1 - \hat{P}(x_i))^{(1-y_i)}} \right] \quad (15)$$

o

$$G = 2 \hat{\chi}^2$$

Bajo la hipótesis nula de que $\beta_1 = 0$, el estadístico G sigue una distribución chi-cuadrado con 1 grado de libertad, requiriéndose básicamente tener un tamaño de muestra suficientemente grande, n , y suficientes sujetos con $y=0$ e $y=1$.

Hay otra prueba estadísticamente equivalente para analizar la significatividad de los parámetros es la prueba de Wald.

La prueba de Wald bajo la hipótesis nula y las suposiciones del tamaño de la muestra, sigue una distribución normal estándar.

Aunque no se ha discutido formalmente (más adelante se hará) cómo se obtienen las estimaciones de los errores estándar (\widehat{SE}) de los parámetros estimados, el software los estima habitualmente.

El estadístico de Wald se determina de la siguiente manera:

$$W = \frac{\hat{\beta}_1}{\widehat{SE}(\hat{\beta}_1)} \quad (17)$$

Algunos paquetes de software muestran la estadística $W^2 = Z^2$, que se distribuye como chi-cuadrado con 1 grado de libertad.

Finalmente, un complemento importante para probar la importancia del modelo, es el cálculo e interpretación de los intervalos de confianza para los parámetros del modelo.

En particular, los estimadores de intervalo de confianza para la pendiente (β_1) y el intercepto (β_0) se basan, en sus respectivas pruebas de Wald y, a veces, se los denomina intervalos de confianza basados en Wald. El intervalo de confianza de $100(1-\alpha)\%$ para el coeficiente de la pendiente es

$$\hat{\beta}_1 \pm Z_{1-\frac{\alpha}{2}} \widehat{SE}(\hat{\beta}_1) \quad (18)$$

y para el intercepto es

$$\hat{\beta}_0 \pm Z_{1-\frac{\alpha}{2}} \widehat{SE}(\hat{\beta}_0) \quad (19)$$

donde $Z_{1-\frac{\alpha}{2}}$ es el $100\left(1-\frac{\alpha}{2}\right)\%$ punto superior de la distribución normal estándar y $\widehat{SE}(\cdot)$ denota un estimador del error estándar basado en el modelo del respectivo parámetro.

3.2.2 Modelo de Regresión Logística Múltiple

Una generalización de la regresión logística simple es la realizada por Hosmer et al., (2013), para p variables independientes. Supongamos que tenemos una muestra de n observaciones independientes $(x_i, y_i) i=1, 2, \dots, n$. Como en el caso univariado, ajustar el modelo requiere que obtengamos estimaciones del vector $\beta = (\beta_0, \beta_1, \dots, \beta_p)$, el vector de parámetros desconocidos asociados a las p variables explicativas.

En el modelo de regresión múltiple dicotómica, la probabilidad de éxito es dada por la siguiente expresión:

$$P(Y=1|X=X_i) = \frac{e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}}} = \frac{e^{\beta_0 + X_i \beta}}{1 + e^{\beta_0 + X_i \beta}} = P(X_i) \quad (20)$$

Y el complemento de esta probabilidad es:

$$P(Y=0|X=X_i) = \frac{1}{1 + e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}}} = \frac{1}{1 + e^{\beta_0 + X_i \beta}} = 1 - P(X_i) \quad (21)$$

Note que

$$\frac{P(X_i)}{1 - P(X_i)} = e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}} \quad (22)$$

Al aplicar el logaritmo natural en (22) se obtiene la función de unión $g(x)$ que, en el contexto de modelos lineales generalizados, es una función monótona, diferenciable y relaciona el promedio de Y , dada por $E(Y) = P(Y=1)$, con sus predictores lineales (Hosmer et al., 2013) dada por:

$$g(x) = \ln\left(\frac{P(X_i)}{1 - P(X_i)}\right) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} \quad (23)$$

donde, para el modelo de regresión logística múltiple,

$$P(Y=1|X=X_i) = \frac{e^{g(x)}}{1 + e^{g(x)}} = \frac{1}{1 + e^{-g(x)}} \quad (24)$$

Si algunas de las variables independientes son discretas, variables de escala nominal como raza, sexo, tipo de escuela, etc., no es apropiado incluirlas en el modelo como si fueran variables de escala de intervalo. Los números utilizados para representar los distintos niveles de estas variables de escala nominal son meramente identificadores y no tienen significado numérico. La asignación de un número a cada categoría no resuelve el problema. La solución a este problema es crear tantas variables dicotómicas como número de posibles niveles menos uno. Estas nuevas variables, artificialmente creadas, reciben en la literatura anglosajona el nombre de *dummy*, traducándose con diferentes denominaciones como pueden ser variables internas, indicadoras, o variables de diseño.

Supongamos, por ejemplo, que una de las variables independientes es la raza, que se ha codificado como "blanco", "negro" y "mestizo". En este caso, se necesitan dos variables *dummy*. Una posible estrategia de codificación es que cuando el estudiante es "blanco",

las dos variables de diseño, D_1 y D_2 , se igualan a cero; cuando el estudiante es "negro", D_1 se establecerá igual a 1, mientras que D_2 aún igual a 0; cuando la raza del estudiante es "mestizo", usaríamos $D_1 = 0$ y $D_2 = 1$. La Tabla 1 ilustra esta codificación de las variables de diseño.

Tabla 1. Codificación de variables dummy

Raza	D_1	D_2
Blanco	0	0
Negro	1	0
Mestizo	0	1

Fuente: Hosmer et al., (2013)

En general, si una variable nominal tiene k valores posibles, entonces se necesitan $k - 1$ variables de diseño. La razón para utilizar una variable menos de diseño es que, a menos que se indique lo contrario, nuestros modelos tienen un término constante. Para ilustrar la notación utilizada para las variables de diseño, supongamos que la j -ésima variable independiente x_j tiene k_j niveles. Las variables de diseño $k_j - 1$ se denotarán como D_{jl} y los coeficientes para estas variables de diseño se denotarán como β_{jl} , $l = 1, 2, \dots, k_j - 1$.

Por lo tanto, el logit para un modelo con p variables, con la j -ésima variable cualitativa quedaría:

$$g(x) = \ln \left(\frac{P(X_i)}{1 - P(X_i)} \right) = \beta_0 + \beta_1 x_1 + \dots + \sum_{l=1}^{k_j-1} \beta_{jl} D_{jl} + \beta_p x_p \quad (25)$$

3.2.2.1 Estimación del modelo de regresión logística múltiple

Supongamos que tenemos una muestra aleatoria de n observaciones independientes del par (x_i, y_i) , $i = 1, 2, \dots, n$, como en el caso univariable, ajustar el modelo requiere la estimación del vector de parámetros $\beta = (\beta_0, \beta_1, \dots, \beta_p)$. El método de estimación utilizado en el caso multivariable es el mismo que en la situación univariable: máxima verosimilitud. La función de verosimilitud es casi idéntica a la dada en la ecuación (5) con el único cambio de que $P(x)$ ahora se define como en la ecuación (23).

En este caso tendríamos $p+1$ ecuaciones de verosimilitud obtenidas al diferenciar la función de verosimilitud logarítmica con respecto a los $p+1$ parámetros (los coeficientes del modelo). Las ecuaciones de verosimilitud que resultan se pueden expresar de la siguiente manera:

$$\sum_{i=1}^n [y_i - P(x_i)] = 0 \quad (26)$$

y

$$\sum_{i=1}^n x_{ij} [y_i - P(x_i)] = 0 \quad (27)$$

para $j = 1, 2, \dots, p$

Como en el caso de regresión logística simple, la solución de las ecuaciones de verosimilitud (26) y (27) requieren de software y está disponible en prácticamente todos los paquetes estadísticos. De este modo, digamos que β es la solución a estas ecuaciones. Por lo tanto, los valores ajustados para el modelo de regresión logística múltiple son $P(x_i)$, según la expresión en la ecuación (24) calculada utilizando β y x_i .

El método de estimación de las varianzas y covarianzas de los coeficientes estimados se deriva de la teoría de estimación de máxima verosimilitud. Esta teoría establece que los estimadores se obtienen a partir de la matriz de segundas derivadas parciales de la función del logaritmo de la verosimilitud, $\ln[l(\beta)]$. Estas derivadas parciales tienen la siguiente forma general

$$\frac{\partial^2 L(\beta)}{\partial \beta_j^2} = - \sum_{i=1}^n x_{ij}^2 P(x_i) [1 - P(x_i)] \quad (28)$$

$$\frac{\partial^2 L(\beta)}{\partial \beta_j \partial \beta_l} = - \sum_{i=1}^n x_{ij} x_{il} P(x_i) [1 - P(x_i)] \quad (29)$$

para $j, l = 0, 1, 2, \dots, p$

Y se construye una matriz cuadrada de dimensiones $p+1$ que contiene el negativo de los términos dados en las ecuaciones (28) y (29), la cual denotaremos como $I(\beta)$. Esta matriz se conoce como *matriz de información observada*. Las varianzas y covarianzas de los coeficientes estimados se obtienen de la inversa de esta matriz, que denotamos como $\text{Var}(\beta) = I^{-1}(\beta)$. Excepto en casos muy especiales, no es posible escribir una expresión explícita para los elementos en esta matriz. Por lo tanto, usaremos la notación $\text{Var}(\beta_j)$ para denotar el j -ésimo elemento diagonal de esta matriz, que es la varianza de $\hat{\beta}_j$, y $\text{Cov}(\beta_j, \beta_l)$ para denotar el jl -ésimo elemento fuera de la diagonal, que es la covarianza entre $\hat{\beta}_j$ y $\hat{\beta}_l$. Los estimadores de las varianzas y covarianzas, que se

denotarán con $\widehat{Var}(\widehat{\beta}_j)$, se obtienen al evaluar $Var(\beta)$ en $\widehat{\beta}$. Se utiliza $\widehat{Var}(\widehat{\beta}_j)$ y $\widehat{Cov}(\widehat{\beta}_j, \widehat{\beta}_l), j, l=0, 1, 2, \dots, p$ para denotar los valores en esta matriz. Casi siempre se utilizará únicamente las estimaciones de los errores estándar de los coeficientes estimados, que designamos como

$$\widehat{SE}(\widehat{\beta}_j) = [\widehat{Var}(\widehat{\beta}_j)]^{\frac{1}{2}} \quad (30)$$

para $j, l = 0, 1, 2, \dots, p$. Dicha notación es utilizada para evaluar la significancia de los coeficientes y para estimar los intervalos de confianza.

3.2.2.2 Evaluación de la significación del modelo

Al igual que en el caso de una única variable independiente, luego de estimar los coeficientes, se procederá al proceso de evaluación del modelo. Por lo tanto, el primer paso en este proceso suele ser evaluar la significación de las variables en el mismo. La prueba de razón de verosimilitud para la significancia global de los p coeficientes de las variables independientes en el modelo se realiza exactamente de la misma manera que en el caso univariable. La prueba se basa en la estadística G dada en la ecuación (13). La única diferencia es que los valores de $P(x_j)$ son ajustados por el modelo que contiene $p + 1$ parámetros, β . Bajo la hipótesis nula de que los p coeficientes para las variables explicativas en el modelo son iguales a cero, la distribución del estadístico G sigue una distribución chi-cuadrado con p grados de libertad.

Antes de concluir que alguno o todos los coeficientes son distintos de cero, debemos evaluar la significancia estadística de cada uno de los coeficientes mediante la prueba de Wald, utilizando el siguiente estadístico

$$W_j = \frac{\widehat{\beta}_j}{\widehat{SE}(\widehat{\beta}_j)} \quad (31)$$

Bajo la hipótesis nula de que un parámetro individual es cero, el estadístico, W_j , sigue una distribución normal. Debiéndose calcular el valor p para determinar la significación estadística del coeficiente en cuestión dado un nivel de confianza determinado.

Finalmente, como el objetivo es ajustar el mejor modelo minimizando el número de parámetros, el próximo paso es determinar un modelo reducido que contenga solo las

variables que se consideran significativas, y comparar este modelo con el modelo que contiene todas las variables.

Hay que tener en cuenta que siempre que se incluya (o excluya) una variable independiente categórica de un modelo, todas sus variables de diseño deberían incluirse (o excluirse); hacer lo contrario implica que hemos recodificado la variable. Si k es el número de niveles de una variable categórica, entonces la contribución a los grados de libertad para la prueba de razón de verosimilitud para la exclusión de esta variable es $k - 1$.

Los intervalos de confianzas al $100(1 - \alpha)\%$, basado en Wald, para los coeficientes de las variables independientes, β_j , se obtienen de la ecuación (18) y para el término constante β_0 , mediante la ecuación (19).

3.2.2.3 Evaluación del ajuste del modelo

Según Hosmer et al. (2013), la evaluación de la bondad del ajuste no es una comparación relativa, es una comparación absoluta. Cuando evaluamos la bondad del ajuste, comparamos los valores ajustados con los valores observados, donde podemos pensar que los valores observados provienen del mejor modelo posible, el modelo saturado. Otra forma de describir esto es considerar una secuencia de modelos progresivamente más grandes (más covariables). El modelo más pequeño contiene solo el término constante, β_0 . El proceso de creación de modelos agrega variables al modelo y se evalúa la importancia de las covariables agregadas haciendo referencia hacia atrás, a un modelo más pequeño, uno sin las covariables agregadas. Al evaluar la bondad del ajuste, comparamos el modelo ajustado con el modelo más grande posible, el modelo saturado, no con un modelo más pequeño.

Hosmer et al. (2013), utiliza el término *patrón de covariables* para describir una configuración particular de valores para las covariables en un modelo. Por ejemplo, en un conjunto de datos que contiene valores de edad, raza, sexo y peso para cada sujeto, la combinación de estos factores puede dar lugar a tantos patrones de covariables diferentes como sujetos haya. Por otro lado, si el modelo contiene únicamente raza y sexo, cada uno codificado en dos niveles, sólo hay cuatro posibles patrones de covariables.

Si denotamos por J el número de distintos patrones de covariables y por n el número de observaciones, si algunas observaciones tienen el mismo patrón de covariables entonces $J < n$. Denotamos como m_j el número de observaciones con el j -ésimo patrón de covariables, con $j=1,2,3,\dots,J$, por lo tanto $\sum m_j = n$. Si y_j es el número de observaciones con $y=1$ entre los m_j observaciones, entonces $\sum y_j = n_1$ es el número total de observaciones con $y=1$.

En la práctica, cuando hay al menos una covariable continua, entonces $J \approx n$.

3.2.2.3.1 Medidas de bondad de ajuste: El estadístico Chi-cuadrado de Pearson, la Devianza, la prueba de Hosmer-Lemeshow, y los índices "Pseudo R^2 "

En regresión lineal tanto las medidas de ajustes del modelo como los efectos de algunas observaciones sobre el ajuste son funciones de los residuos, definidos como la diferencia entre al valor observado y el valor ajustado por el modelo $(y - \hat{y})$.

Cuando la variable respuesta es ordinal, los residuos son más difíciles de definir, existiendo pocas definiciones en la literatura que tenga una distribuciones de referencia simples e interpretables. Algunas propuestas producen múltiples residuos, cuando la variable respuesta es ordinal, siendo difícil de interpretar. Otras interpretaciones de residuos se basan en una respuesta discreta y son discretos en sí mismos, lo que los hace menos útiles para el diagnóstico mediante gráficos (Greenwell, McCarthy, Boehmke, and Liu, 2018).

Esto se debe al hecho de que una variable ordinal tiene valores discretos que están etiquetados con valores numéricos, pero que en realidad no lo son. Los valores simplemente representan categorías ordenadas.

En regresión logística los valores ajustados por un determinado modelo se calculan para cada patrón de covariable y dependen de la probabilidad estimada para dicho patrón.

Si denotamos el valor ajustado para el j -ésimo patrón de covariable como \hat{y}_j

$$\hat{y}_j = m_j \hat{\pi}_j = m_j \left\{ \frac{e^{\hat{g}(x_j)}}{1 + e^{\hat{g}(x_j)}} \right\} \quad (32)$$

donde $\hat{g}(x_j) = \hat{\beta}_0 + \hat{\beta}_1 x_{j1} + \hat{\beta}_2 x_{j2} + \dots + \hat{\beta}_p x_{jp}$ es el logit estimado, $\hat{\pi}_j$ es la probabilidad estimada del j -ésimo patrón de covariable.

Los residuos usados ampliamente en la validación de los modelos de Regresión Logística, son los de Pearson y los de la Devianza.

A continuación, consideramos tres medidas para evaluar la diferencia entre los valores observados y ajustado: residual de Pearson, la devianza residual y la prueba de Hosmer-Lemeshow.

Para el j -ésimo patrón de covariables el *residual de Pearson* es

$$r(y_j, \hat{\pi}_j) = \frac{(y_j - m_j \hat{\pi}_j)}{\sqrt{m_j \hat{\pi}_j (1 - \hat{\pi}_j)}} \quad (33)$$

La medida resumen basado en el residual de Pearson es el estadístico *chi-cuadrado de Pearson*

$$X^2 = \sum_{j=1}^J [r(y_j, \hat{\pi}_j)]^2 \quad (34)$$

La devianza residual es

$$d(y_j, \hat{\pi}_j) = \pm \left\{ 2 \left[y_j \ln \left(\frac{y_j}{m_j \hat{\pi}_j} \right) + (m_j - y_j) \ln \left(\frac{(m_j - y_j)}{m_j (1 - \hat{\pi}_j)} \right) \right] \right\}^{\frac{1}{2}} \quad (35)$$

Donde el signo + o - es el mismo signo de $(y_j - m_j \hat{\pi}_j)$. Para los patrones de covariables con $y_j = 0$ la devianza residual es

$$d(y_j, \hat{\pi}_j) = -\sqrt{2m_j |\ln(1 - \hat{\pi}_j)|} \quad (36)$$

y cuando $y_j = m_j$ la devianza residual es

$$d(y_j, \hat{\pi}_j) = \sqrt{2m_j |\ln(\hat{\pi}_j)|} \quad (37)$$

La medida resumen basado en la devianza residual es la *devianza*

$$D = \sum_{j=1}^J [d(y_j, \hat{\pi}_j)]^2 \quad (38)$$

Cuando $J = n$ la devianza es igual al mostrado en la ecuación (11)

La distribución de los estadísticos X^2 y D bajo la hipótesis nula de que el modelo ajustado es correcto en todos los aspectos sigue una distribución chi-cuadrado con $J - (p + 1)$ grados de libertad. De tal modo, que si el modelo ajustado es correcto se debe

aceptar la hipótesis nula, por ende un p -valor superior a 0,05 implica que lo que observamos se ajusta suficientemente a lo esperado bajo el modelo.

Básicamente, la prueba de Hosmer y Lemeshow consiste en dividir el total de la muestra (observaciones) en g grupos de igual tamaño (n/g) en función de las probabilidades estimadas, el primer par de conteos (observados y ajustados) corresponde a las n/g observaciones que tienen las probabilidades estimadas más altas, el siguiente par se refiere a las n/g observaciones que tienen las segundas probabilidades más altas, y así sucesivamente. Cada grupo tiene un recuento observado de observaciones con cada resultado y un valor ajustado para cada resultado. El valor ajustado para un resultado es la suma de las probabilidades estimadas para ese resultado para todas las observaciones en ese grupo. Luego el estadístico de bondad de ajuste de Hosmer-Lemeshow, \hat{C} , se obtiene calculando el estadístico chi-cuadrado de Pearson a la tabla $g \times 2$ de las frecuencias observadas y estimadas, mediante la siguiente formula

$$\hat{C} = \sum_{k=1}^g \left[\frac{(O_{1k} - \hat{e}_{1k})^2}{\hat{e}_{1k}} + \frac{(O_{0k} - \hat{e}_{0k})^2}{\hat{e}_{0k}} \right] \quad (39)$$

donde

$$O_{1k} = \sum_{j=1}^{J_k} y_j \quad (40)$$

$$O_{0k} = \sum_{j=1}^{J_k} (m_j - y_j) \quad (41)$$

$$\hat{e}_{1k} = \sum_{j=1}^{J_k} m_j \hat{\pi}_j \quad (42)$$

$$\hat{e}_{0k} = \sum_{j=1}^{J_k} m_j (1 - \hat{\pi}_j) \quad (43)$$

Donde J_k es el número de patrones de covariables en el k -ésimo grupo. El estadístico también puede escribirse de la siguiente manera

$$\hat{C} = \sum_{k=1}^g \left[\frac{(O_{1k} - n_k \hat{\pi}_k)^2}{n_k \hat{\pi}_k} + \frac{(O_{0k} - n_k (1 - \hat{\pi}_k))^2}{n_k (1 - \hat{\pi}_k)} \right] \quad (44)$$

donde $\hat{\pi}_k$ es la probabilidad promedio estimada para el k -ésimo grupo.

$$\hat{\pi}_k = \frac{1}{n_k} \sum_{j=1}^{J_k} m_j \hat{\pi}_j \quad (45)$$

Bajo la hipótesis nula de que el modelo ajustado es correcto, el estadístico \hat{C} se distribuye chi-cuadrado con $g-2$ grados de libertad $X^2(g-2)$. De igual modo que con los estadísticos *chi-cuadrado de Pearson* (X^2 , ecuación 34) y *devianza* (D , ecuación 38), si el modelo ajustado es correcto, se debe aceptar la hipótesis nula, por ende un *p-valor* superior a 0,05 implica que lo que observamos se ajusta suficientemente a lo esperado bajo el modelo.

En resumen, para las pruebas *chi-cuadrado de Pearson*, *devianza* y *Hosmer-Lemeshow* lo que se desea es que no haya significación, es decir aceptar dichas hipótesis nulas de que el modelo propuesto se ajusta a lo observado.

En regresión lineal el coeficiente de determinación R^2 , resume la proporción de varianza en la variable respuesta que puede explicarse por las variables independientes. Sin embargo para los modelos de regresión logística binaria, multinomial y ordinal no es posible calcular la misma R^2 que en regresión lineal, por lo que en la mayoría de las aplicaciones se calculan tres aproximaciones, los llamados índices “pseudo R^2 ”. Sin embargo, es bastante común encontrar valores bajos para estos “pseudo R^2 ” incluso cuando el modelo es bueno. Por esta razón, puede ser mejor evitar estas estadística, excepto quizás con el propósito de comparar modelos compatibles, Faraway (2016).

Camarero, Llorente y Ramírez (2015) consideran estos índices como medidas de probabilidad que nos orientan sobre la ganancia que nos produce el modelo sobre situaciones en las que no tenemos modelos. Estos indicadores nos muestran qué ganamos utilizando un modelo, pero no nos indican qué capacidad tenemos de predecir los datos, algo que si ofrece el coeficiente de determinación (R^2).

Según Smith y McKenna (2013), el R^2 de McFadden es quizás el más sencillo de tales índices “pseudo R^2 ”, en el sentido de que refleja tanto el criterio que se minimiza en la estimación de un modelo de regresión logística como la varianza explicada por el modelo de regresión logística ajustado. Se determina mediante la siguiente expresión:

$$R_{MF}^2 = 1 - \frac{\ln[\text{Verosimilitud del modelo ajustado}]}{\ln[\text{Verosimilitud del modelo nulo}]} \quad (46)$$

El índice R^2 de Cox y Snell se representa como

$$R_{CS}^2 = 1 - \left(\frac{L(\text{Modelo nulo})}{L(\text{Modelo ajustado})} \right)^{\frac{2}{n}} \quad (47)$$

Donde $L(\text{Modelo nulo})$ y $L(\text{Modelo ajustado})$ son los valores de la función de verosimilitud del modelo con solo el intercepto y del modelo ajustado, respectivamente. n es el número de observaciones. Debido a que el R_{CS}^2 puede eventualmente exceder de 1.0 se propuso una medida re-escalada para este estadístico, el llamado R^2 de Nagelkerke (R_{NK}^2).

$$R_{NK}^2 = \frac{1 - \left(\frac{L(\text{Modelo nulo})}{L(\text{Modelo ajustado})} \right)^{\frac{2}{n}}}{1 - L(\text{Modelo nulo})^{\frac{2}{n}}} \quad (48)$$

donde el re-escalamiento se realiza dividiendo R_{CS}^2 por su valor máximo posible. De esta manera el R_{NK}^2 está comprendido entre 0 y 1, un rango que es idéntico al del R^2 de los modelos de regresión lineal.

3.2.3 Regresión Logística Ordinal

Hasta este punto se ha considerado la regresión logística binaria, es decir la variable respuesta tiene dos posibles valores, como por ejemplo éxito y fracaso. Ahora bien, si la variable respuesta es nominal con más de dos valores, por ejemplo, considere un estudio de elección de un plan de seguro entre tres planes ofrecidos. La variable de resultado tiene tres niveles que indican qué plan, A, B o C se elige. Las posibles covariables pueden incluir el género, la edad, el ingreso, el tamaño de la familia, el nivel educativo y otros. El objetivo es estimar la probabilidad de elegir cada uno de los tres planes, así como, estimar las probabilidades de elección del plan en función de las covariables. Este problema se puede modelar mediante la regresión logística multinomial.

Sin embargo, hay situaciones donde la variable respuesta puede tomar uno de varios niveles ordenados. Ejemplos comunes de resultados ordinales incluyen variables como la extensión de la enfermedad (ninguna, alguna, grave), el desempeño laboral (inadecuado, satisfactorio, sobresaliente) y la opinión sobre la posición de un grupo de interés sobre algún tema (muy en desacuerdo, en desacuerdo, de acuerdo, muy de acuerdo), que implican un determinado orden. En tal contexto, se podría usar el modelo logístico multinomial. Este análisis, sin embargo, no tomaría en cuenta la naturaleza

ordinal del resultado y , por lo tanto, los cocientes de probabilidades estimados podrían no abordar las preguntas del análisis. En esta sección consideramos un modelo de regresión logística que tiene en cuenta el orden de clasificación de los resultados.

Supongamos que la variable respuesta ordinal, Y , puede tomar $K+1$ valores codificados $0, 1, 2, \dots, K$. Denotamos una expresión general para la probabilidad de que el resultado sea igual a k condicional en un vector, x , de p covariables como

$$Pr[Y = k | x] = \phi_k(x) \quad (49)$$

El modelo a considerar es el de Odds Proporcionales o logístico acumulativo, en este se compara la probabilidad de una respuesta menor o igual a k a la probabilidad de una respuesta mayor que k , esto es

$$g_k(x) = \ln \left[\frac{Pr(Y \leq k | x)}{Pr(Y > k | x)} \right] = \ln \left[\frac{\phi_0(x) + \phi_1(x) + \dots + \phi_k(x)}{\phi_{k+1}(x) + \phi_{k+2}(x) + \dots + \phi_K(x)} \right] = \alpha_k - x' \beta \quad (50)$$

para $k = 0, 1, \dots, K-1$.

Este modelo asume que el cociente de odds, el cual establece el efecto de las variables explicativas sobre la variable respuesta, es el mismo para cualquiera de las posibles comparaciones que puedan realizarse, al margen del punto de corte realizado. Es decir, este modelo considera que el efecto de las covariables $x_1, x_2, x_3, \dots, x_p$ es el mismo en todas la categorías, en escala logarítmica, de este modo sólo el intercepto depende de la categoría.

El signo negativo de $x' \beta$ se incluye para permitir la interpretación habitual de que un valor positivo de β_p significa que a medida que x_p aumenta, también aumenta la probabilidad de valores más altos de Y .

Debido a ello cada una de las $K-1$ ecuaciones $\text{logit} g_k(x)$, tendrán un valor para el intercepto (α_k) y los mismos coeficientes para cada una de las p variables explicativas (β).

Por lo tanto, las inferencias del modelo de odds proporcionales se prestan a una discusión general de la dirección de la respuesta y no tienen que centrarse en categorías de resultados específicas. Los resultados son mucho más simples de describir que los de

cualquiera de los modelos no restringidos. De ahí que sea el modelo de regresión logística ordinal más utilizado en la práctica.

El odds (riesgo) de que un evento ocurra es el cociente de la probabilidad de que un evento ocurra sobre la probabilidad de que no ocurra.

El método utilizado para ajustar este modelo es una adaptación del algoritmo de máxima verosimilitud utilizado para la regresión logística binaria.

Como cualquier modelo de regresión, al ajustar un modelo de regresión logística ordinal se debe evaluar si cumple con los supuestos del modelo, y si las probabilidades producidas por el modelo reflejan con exactitud los resultados reales de los datos. Esto último se conoce como la bondad de ajuste del modelo.

Se han propuestos muchas pruebas estadísticas para evaluar si se cumple la proporcionalidad de los odds, las cuales de alguna manera comparan el modelo en la ecuación (41) con un modelo aumentado en el cual se permite que los coeficientes para las covariables del modelo puedan ser diferentes. Esto es

$$g_k(x) = \ln \left[\frac{Pr(Y \leq k | x)}{Pr(Y > k | x)} \right] = \alpha_k - x' \beta_k \quad (51)$$

Donde $\alpha_k < \alpha_{k+1}$ para $k = 1, \dots, K$. En este trabajo utilizaremos el test de Brant, el cual compara los valores estimados para los coeficientes ajustados según el modelo de la ecuación (51), con los obtenidos mediante el ajuste de la ecuación (50). La prueba tiene $[(K+1)-2] \times p$ grados de libertad. La idea es aceptar la hipótesis nula de que se cumple la proporcionalidad de los odds.

Para evaluar la bondad de ajuste del modelo utilizamos la versión ordinal de la prueba de Hosmer-Lemeshow,

$$\hat{C}_O = \sum_{l=1}^g \sum_{k=0}^K \frac{(O_{lk} - \hat{E}_{lk})^2}{\hat{E}_{lk}} \quad (52)$$

Bajo la hipótesis nula de que el modelo ajustado es correcto, el estadístico \hat{C}_O se distribuye chi-cuadrado con $K \times (g-2) + (K-1)$ grados de libertad. Una vez más, la idea es aceptar la hipótesis nula de que el modelo se ajusta a lo observado.

Resumiendo, en un modelo de RLO se deben tener en cuenta las siguientes hipótesis.

- 1) Proporcionalidad de los Odds, el cual se verifica mediante la Prueba de líneas paralelas, con el test de Brant. La idea es aceptar la hipótesis nula de que se cumple la proporcionalidad de los odds, por lo cual un *p-valor* mayor a 0,05 de dicha hipótesis es lo deseado.

- 2) La significación del modelo, lo cual puede verificarse mediante el cambio en la *devianza* debido a la inclusión conjunta de las variables explicativas en el modelo en contraste con el modelo nulo (modelo con solo el intercepto) mediante el estadístico *G* de la ecuación (14) y/o mediante el estadístico de Wald de la ecuación (17), el cual evalúa la significancia de cada una de las variables de manera individual. En el primer caso se debe aceptar la hipótesis nula de que el modelo que incluye de manera conjunta las variables explicativas es mejor que el modelo nulo, por lo cual un *p-valor* mayor a 0,05 en dicha hipótesis es lo deseado, mientras que en el estadístico de Wald se requiere rechazar la hipótesis nula de que el coeficiente en cuestión es igual a cero ($\beta_i=0$), requiriéndose de un *p-valor* menor a 0,05 en dicha hipótesis.

- 3) El ajuste del modelo, el cual se realiza mediante las *pruebas chi-cuadrado de Pearson* (X^2), la *devianza* (D) y la *prueba de Hosmer-Lemeshow* (\hat{C}_o). En éstas se prueba la hipótesis nula de que el modelo ajustado es correcto, por lo tanto, la idea es aceptar la hipótesis nula, por ende un *p-valor* superior a 0,05 implica que lo que observamos se ajusta suficientemente a lo esperado bajo el modelo.

3.2.4 Método de selección de variables por paso (Stepwise)

Según Shalizi (2015), la selección de variables por pasos o por etapas es una familia de métodos para agregar o eliminar variables de un modelo de forma secuencial.

Shalizi (2015), manifiesta que la regresión por pasos hacia adelante (forward stepwise) comienza con un modelo pequeño (tal vez sólo una intersección), y agrega la variable que es mejor de acuerdo con alguna medida o criterio. Este criterio puede ser "valor p más bajo", " R^2 ajustado más alto", "Cp de Mallows más bajo", "AIC más bajo". Este proceso se repite, agregando siempre una variable a la vez, hasta que el criterio deja de

mejorar. En la regresión por pasos hacia atrás (backward stepwise), comenzamos por el contrario, con el modelo más grande que estamos dispuestos a contemplar, y seguimos eliminando las variables hasta que ya no mejoramos. Por su parte, el procedimiento de selección de variables mediante una estrategia mixta contemplará tanto la adición como la eliminación de una variable en cada paso.

Según Shalizi (2015), podríamos ver todos los modelos lineales posibles basados en un conjunto dado de variables, y calcular nuestro criterio para cada uno de ellos; esto se denomina selección de variables de todos los subconjuntos, porque cada modelo corresponde a un subconjunto de las variables. Con p variables hay 2^p modelos posibles, por lo que la regresión de todos los subconjuntos se vuelve, literalmente, exponencialmente más lenta, siendo esto la única justificación real para los procedimientos paso a paso.

Hosmer et al. (2013), consideran que los métodos paso a paso pueden ser útiles como herramientas efectivas de análisis de datos. En particular, hay ocasiones en que el resultado que se estudia es relativamente nuevo y las covariables importantes pueden no conocerse y las asociaciones con el resultado no se comprenden bien. En estos casos, la mayoría de los estudios recopilan muchas covariables posibles y las analizan en busca de asociaciones significativas. El empleo de un procedimiento de selección por pasos puede proporcionar un medio rápido y eficaz para detectar una gran cantidad de variables y ajustar una serie de ecuaciones de regresión logística simultáneamente.

Upton (2017) considera que una medida útil para comparar varios modelos, es una que equilibre la complejidad del modelo (número de covariables) y la bondad de ajuste. La medida más común es el Criterio de Información de Akaike (AIC , por su sigla en inglés). A partir de un conjunto de modelos candidatos, AIC intenta seleccionar el modelo que proporciona la descripción más adecuada de una realidad que probablemente haya sido influenciada por una gran cantidad de variables no medidas. El criterio tiene una aplicación de modelado general, pero, para datos categóricos, basta con calcular

$$AIC = -2 \ln(L) + 2p \quad (53)$$

donde p es el número de parámetros en el modelo y L es la máxima verosimilitud del modelo considerado. Se prefiere un modelo con un valor AIC más pequeño que uno con un valor mayor.

Otro criterio también utilizado con frecuencia es el Criterio de Información Bayesiano (*BIC*, por su sigla en inglés), el cual puede ser determinado con la siguiente expresión,

$$BIC = -2 \ln(L) + 2 \ln(n) \quad (54)$$

donde n es el número de observaciones en la muestra utilizada para determinar el modelo de regresión. Al igual que *AIC*, modelos con valores menores de *BIC* son preferidos ante los modelos con valores mayores. Debido a que $\ln(n)$ será mayor que 2, el modelo seleccionado por *BIC* nunca será más complicado que el elegido mediante *AIC*.

Upton (2017) considera que *AIC* y *BIC* tienen motivaciones diferentes: *AIC* busca seleccionar el modelo, de entre los disponibles, que más se asemeja al verdadero modelo (que estará gobernado por una gran cantidad de consideraciones no medidas y no estará entre los modelos considerados), mientras que *BIC* asume que el modelo correcto se encuentra entre los que se ofrecen y busca identificar ese modelo óptimo. En realidad, dado que ninguno de los modelos que consideramos proporcionará una descripción perfecta (excepto, quizás, si estamos estudiando alguna relación científica), el *AIC* debería ser la medida utilizada.

4 MATERIALES Y MÉTODOS

4.1 Matricula estudiantil

4.1.1 Distribución por facultad.

La Universidad del Magdalena académicamente está compuesta por seis Facultades, las cuales a su vez están conformadas por 23 Programas Académicos, incluyendo el recién creado Programa de Profesional en Deporte, según se puede observar en la tabla 2.

Tabla 2. Conformación Académica de la Universidad del Magdalena

Facultad	Programa Académico
Ciencias Empresariales y Económicas	Administración de Empresas
	Administración de Empresas Turísticas y Hoteleras
	Negocios Internacionales
	Contaduría Pública
	Tecnología en Administración Hotelera y Turísticas
Humanidades	Economía
	Cine y Audiovisuales
	Antropología
	Derecho
Ciencias de la Educación	Profesional en Deporte
	Lic. en Educación Básica con Énfasis en Informática
Ciencias de la Salud	Licenciatura en Preescolar
	Medicina
	Enfermería
	Odontología
Ingenierías	Psicología
	Ingeniería Agronómica
	Ingeniería Ambiental y Sanitaria
	Ingeniería Civil
	Ingeniería de Sistemas
	Ingeniería Electrónica
Ciencias Básicas	Ingeniería Industrial
	Ingeniería Pesquera
	Biología

Fuente: Elaboración Propia

Los datos utilizados en el estudio corresponden a la totalidad de los registros de matriculados en el periodo 2014-2 (segundo semestre del año 2014) y fueron suministradas por el Grupo de Admisiones, Registro y Control de la Universidad del Magdalena y según se muestra en la tabla 3 puede descomponerse de la siguiente manera: las Facultades de Ingeniería y Ciencias Empresariales concentra el 61,49% y el restante 38,51% se distribuye en las Facultades de Ciencias de la Salud, con un 15,38%, Humanidades con un 13,01% a pesar de ser las más recientes en su conformación. La Facultad de Ciencias de la Educación alcanza un 7,84% del total, que corresponde a los

programas creados en el año 2000. No obstante, la antigüedad de la Facultad es de 36 años de existencia. El bajo número de estudiantes se debe a que para esa época (año 2000) la administración central, decidió cerrar los programas presenciales existentes y sólo ofertarlos a distancia, quedando reducida la Facultad a los programas de Licenciatura en Preescolar y en Educación Básica con énfasis en Informática, creados en el año 2000.

La tabla 3 muestra que Facultad de Ciencias Básicas es la más pequeña. Cuenta con un programa de pregrado y sus estudiantes sólo representan el 2,3% de los estudiantes de la Universidad del Magdalena.

Tabla 3. Distribución por facultad

Facultades	Frecuencia	Porcentaje
Facultad de Ciencias Básicas	306	2,30
Facultad de Ciencias de la Educación	1044	7,84
Facultad de Ciencias de la Salud	2049	15,38
Facultad de Ciencias Empresariales y Económicas	3861	28,98
Facultad de Humanidades	1733	13,01
Facultad de Ingeniería	4331	32,51
Total general	13.324	100

Fuente: Elaboración Propia

4.1.2 Distribución por género

La tabla 4 muestra la distribución según género por Facultad, mientras que en la tabla 5 se presenta la distribución según género por Programa Académico, en estas se puede observar que se presenta una relativa paridad en la matrícula estudiantil. Ambos géneros, prácticamente están representados en proporciones iguales. Sin embargo, existen particularidades en varias facultades y programas con predominancia de uno u otro género. La Facultad de Ciencias Empresariales, por ejemplo, es una que se caracteriza porque la matrícula femenina prácticamente dobla a la masculina.

En la Facultad de Ciencias de la educación el género femenino alcanza los mayores porcentajes, pues triplica a la masculina (es 3,2 veces más). Esto se debe fundamentalmente al programa de Preescolar, donde ingresan mujeres llegando a tener prácticamente el 100% de la matrícula total del programa (sólo tiene un estudiante de género masculino) y en el de Licenciatura en Informática están prácticamente equilibrados los dos géneros. Así mismo, es importante la representación del género

femenino en los programas de la facultad de salud; en particular Enfermería y Psicología, en donde casi cuadruplican a la matrícula masculina. Se considera que esto se debe a la tradición de los programas, y al colectivo entre los recién graduados del bachillerato, que piensa que estas profesiones son fundamentalmente para mujeres. Entre estos dos programas se concentra el 15% de toda la matrícula de la universidad.

Sin embargo, se observa una situación inversa en los programas de ingeniería, pues en ellos, el género masculino es más del doble que el femenino. Sin embargo, es la facultad que ocupa un tercer lugar respecto del género femenino, con un 19%, seguida por Humanidades y Educación, ambas con un 12% y cierra la Facultad de Ciencias Básicas con un 2%.

Analizando la distribución de varones matriculados en las distintas Facultades, se observa que se encuentra concentrada mayoritariamente en la Facultad de Ingeniería, con un 46% del total del género masculino matriculado en la Universidad para el período 2014-2, y en la Facultad de Ciencias Empresariales con un 23%, es decir casi el 70% del total. Estas facultades tienen siete y seis programas académicos, respectivamente.

Estas facultades son las más grandes pues tienen un considerable número de programas de alta demanda, lo cual justifica la cantidad de estudiantes matriculados. Así mismo, las políticas institucionales relacionadas con la ampliación de cobertura han generado la ampliación de cupos en varios de sus programas, logrando que ingrese mayor cantidad de estudiantes que en años anteriores a los de la implementación de la política.

Tabla 4. Distribución género por facultad

Facultad	Género				Frecuencia	Porcentaje
	Femenino		Masculino			
	Frecuencia	Porcentaje	Frecuencia	Porcentaje		
Facultad de Ciencias Básicas	135	2,01	171	2,59	306	2,30
Facultad de Ciencias de la Educación	797	11,86	247	3,74	1044	7,84
Facultad de Ciencias de la Salud	1323	19,68	726	11,00	2049	15,38
Facultad de Ciencias Empresariales y Económicas	2344	34,88	1517	22,97	3861	28,98
Facultad de Humanidades	830	12,35	903	13,68	1733	13,01
Facultad de Ingeniería	1292	19,22	3039	46,02	4331	32,51
Total	6.721	100	6.603	100	13.324	100

Fuente: Elaboración Propia

A continuación, se realiza un análisis descriptivo de los datos por programa. Puede apreciarse en la tabla 5 que existe cierta diferencia en la distribución de la matrícula. Por ejemplo, se observa que aunque la Facultad de Ciencias Empresariales posee el mayor porcentaje de matrícula femenina, internamente sus programas tienen diferencias que oscilan entre 0 y 2 puntos porcentuales, teniendo en cuenta los porcentajes globales institucionales. Es decir, para algunos casos no hay diferencias o es baja, alcanzando un punto porcentual de diferencia, mientras que en otros como en Contaduría y Negocios Internacionales, es alta llegando casi a duplicar la matrícula masculina. Estos programas también consiguieron un considerable incremento en el número de matriculados a partir de la ampliación de cobertura.

La tabla 5 también muestra que en la Facultad de Ingeniería se presentan grandes diferencias porcentuales entre la matrícula femenina y la masculina. Mientras en programas como Ingeniería Ambiental y Sanitaria la variación es mínima, en otros como en Ingeniería Electrónica, el género masculino alcanza a superar en 10 veces al género femenino, la cual se encuentra representado por 48 mujeres. En Ingeniería de Sistemas, aunque esta diferencia es menor, es igualmente elevado, ascendiendo en esta ocasión a 5,5 veces. Los demás programas presentan porcentajes menores, sin que esto signifique que no sean importantes: en Ingeniería Civil el género masculino supera en 3,1 veces al femenino, en Agronomía en 2,4, Pesquera 1,8 e Ingeniería Industrial 1,5 veces. Recordemos que en esta Facultad se presenta la mayor cantidad de estudiantes masculinos de toda la universidad, con un 46% del total.

En la tabla 5 se puede observar que en los programas de la Facultad de Humanidades la variación es baja y ambos géneros son más o menos similares, siendo Antropología el programa de esta Facultad que presenta un ligero incremento de mujeres matriculadas. En Cine y Audiovisuales y en Derecho se presenta lo contrario, pero también muy ligeramente.

En la Facultad de Ciencias de la Salud los estudiantes de género femenino es mayor casi dos veces a la masculina, no obstante, hay diferencias marcadas: en el programa de Enfermería el género femenino alcanza a ser 4,7 veces más que el masculino, y en Psicología 2,5 veces, mientras que en Medicina y Odontología la variación es mínima,

aunque la femenina es algo mayor a la masculina en ambos programas. También en la facultad de informática la variación es mínima

Finalmente, la Facultad de Ciencias de la Educación tiene una característica particular en su programa de Licenciatura en Preescolar, ya que tiene sólo un estudiante de género masculino en su matrícula, por lo que prácticamente el 100% de la matrícula son de género femenino.

Tabla 5. Distribución de género por programa

Programa académico	Género				Frecuencia	Porcentaje
	Femenino		Masculino			
	Frecuencia	Porcentaje	Frecuencia	Porcentaje		
Administración de Empresas	521	7,75	446	6,75	967	7,26
Administración de Empresas Turísticas y Hoteleras	34	0,51	14	0,21	48	0,36
Antropología	218	3,24	173	2,62	391	2,93
Biología	135	2,01	171	2,59	306	2,30
Cine y Audiovisuales	134	1,99	250	3,79	384	2,88
Contaduría Pública	644	9,58	391	5,92	1035	7,77
Derecho	478	7,11	480	7,27	958	7,19
Economía	205	3,05	173	2,62	378	2,84
Enfermería	269	4,00	57	0,86	326	2,45
Ingeniería Agronómica	103	1,53	248	3,76	351	2,63
Ingeniería Ambiental y Sanitaria	351	5,22	361	5,47	712	5,34
Ingeniería Civil	173	2,57	534	8,09	707	5,31
Ingeniería de Sistemas	109	1,62	601	9,10	710	5,33
Ingeniería Electrónica	48	0,71	497	7,53	545	4,09
Ingeniería Industrial	405	6,03	615	9,31	1020	7,66
Ingeniería Pesquera	103	1,53	183	2,77	286	2,15
Lic. en Educación Básica con Énfasis en Informática	225	3,35	246	3,73	471	3,53
Licenciatura en Preescolar	572	8,51	1	0,02	573	4,30
Medicina	293	4,36	278	4,21	571	4,29
Negocios Internacionales	613	9,12	355	5,38	968	7,27
Odontología	274	4,08	193	2,92	467	3,50
Psicología	487	7,25	198	3,00	685	5,14
Tecnología en Administración Hotelera y Turística	327	4,87	138	2,09	465	3,49
Total	6.721	100	6.603	100	13.324	100

Fuente: Elaboración Propia

4.1.3 Procedencia de los estudiantes por departamento

La tabla 6 muestra la procedencia de los estudiantes según departamento. En ella se puede observar que la Universidad del Magdalena ha dejado de ser una Universidad local que atendía las necesidades educativas de la ciudad y el departamento, para ser regional y ha empezado a ser atractiva a nivel nacional, aunque aún hoy el 69,59% de su matrícula corresponde a estudiantes del propio Departamento del Magdalena. Otros departamentos de la Región Caribe Colombiana tienen una presencia significativa en la

matricula: Cesar, Atlántico, Bolívar, Guajira y Sucre en conjunto suman un 21,53% de estudiantes atendidos por la Universidad del Magdalena, con una mayor presencia del Cesar con un 6,10% del total, seguido de Atlántico con un 4,74%, Bolívar con un 3,90%, La Guajira con un 3,52% y Sucre con 3,27% y cierra Córdoba con un 1,01%. Entre los matriculados de otras regiones colombianas, cabe destacar que del Departamento de Cundinamarca hay un 2,69% de la matrícula de estudiantes y el 5,96% se distribuye en otros departamentos del país.

Tabla 6. Procedencia de los estudiantes según departamento

Departamento de origen	Frecuencia	Porcentaje	Porcentaje acumulado
Magdalena	9272	69,59	69,59
Cesar	813	6,10	75,69
Atlántico	631	4,74	80,43
Bolívar	519	3,90	84,32
La guajira	469	3,52	87,84
Sucre	436	3,27	91,11
Cundinamarca	359	2,69	93,81
Santander	219	1,64	95,45
Córdoba	135	1,01	96,47
Antioquia	122	0,92	97,38
Norte de Santander	66	0,50	97,88
Valle	52	0,39	98,27
Tolima	39	0,29	98,56
Extranjeros	30	0,23	98,78
Boyacá	28	0,21	98,99
Caldas	26	0,20	99,19
Huila	18	0,14	99,32
Meta	17	0,13	99,45
Quindío	12	0,09	99,54
Risaralda	9	0,07	99,61
Putumayo	7	0,05	99,66
San Andrés Islas	7	0,05	99,71
Arauca	6	0,05	99,76
Caquetá	6	0,05	99,80
Casanare	6	0,05	99,85
Cauca	6	0,05	99,89
Nariño	5	0,04	99,93
Choco	4	0,03	99,96
Amazonas	3	0,02	99,98
Guaviare	2	0,02	100
Total	13.324	100	100

Fuente: Elaboración Propia

Entre estos otros departamentos destacan Santander y Antioquia, y aunque el porcentaje de matriculados de estos departamentos asciende sólo a un 2,56% se pone en evidencia el interés en el país por los programas académicos de la Universidad del Magdalena. También es importante señalar que la Universidad cuenta entre sus estudiantes con 30 extranjeros, representando el 0,23%. De igual manera cabe subrayar, que la Universidad

del Magdalena tiene estudiantes desde San Andrés Islas hasta Caquetá, Guaviare, Arauca y Amazonas.

Respecto del municipio de origen de los estudiantes universitarios, se observa que la mayor parte de los estudiantes, son naturales de Santa Marta con 6.693 estudiantes, representando el 50,23%, seguido por 740 estudiantes de Ciénaga (5,6%), 535 estudiantes de la ciudad de Barranquilla (4%), en tanto que el 13,76% de los estudiantes son nativos de otros municipios del Departamento del Magdalena.

En términos regionales para el periodo académico 2014-2, los estudiantes procedían de los siguientes departamentos: 12.275 estudiantes (92,13%) provienen de los 7 departamentos de la Región Caribe colombiana, por ende 1.049 estudiantes (7,87%) provienen de otras regiones del país. Esto muestra cómo ha ido creciendo el interés en el país por nuestra Universidad. A tal punto que podemos afirmar que el 30,41% de los matriculados son oriundos de otros departamentos del país. Esto nos permite concluir que hay que desarrollar políticas que permitan una mayor inclusión de estudiantes provenientes del departamento. Es probable que estos estudiantes queden por fuera debido al bajo rendimiento académico que presentan los estudiantes del bachillerato en el Departamento del Magdalena, lo cual se evidencia en los resultados de las pruebas Saber 9 y Saber 11.

De acuerdo con lo anterior, la comunidad de la ciudad de Santa Marta, Capital del Departamento del Magdalena, sigue siendo la más beneficiada por la cantidad de estudiantes que logran acceder a los estudios superiores, aun cuando la institución no recibe aportes del Distrito de Santa Marta para financiar la educación superior.

4.1.4 Información socioeconómica

En Colombia, la estratificación socioeconómica es la clasificación de los inmuebles residenciales de un municipio, que se hace en atención al Régimen de los Servicios Públicos Domiciliario (Ley 142 de 1994) (DANE, 2018). El Artículo 102 de la Ley 142 de 1994 establece los siguientes seis estratos socioeconómicos: 1) bajo-bajo, 2) bajo, 3) medio-bajo, 4) medio, 5) medio-alto, y 6) alto.

El artículo 99.6 de la citada ley establece que a los usuarios de los servicios públicos residentes en los estratos 1, 2 y 3 se les aplicará un factor de subsidio máximo de 50%,

40% y 15%, respectivamente, sobre el costo medio del suministro. Mientras que a los usuarios de los estratos 5 y 6 y no residenciales se les aplicara un factor de contribución no superior al 20% del valor del servicio. Los usuarios del estrato 4 no reciben subsidio y tampoco contribuye a subsidiar a los usuarios de los estratos de menores ingresos.

En lo que respecta a la información socioeconómica de los estudiantes de la Universidad del Magdalena, la tabla 7 muestra que el 83,51% pertenecen a los estratos 1 y 2, mientras que el 14,25% son del estrato 3, el número de estudiantes de los estratos 4, 5 y 6 apenas supera el 2%. Esto demuestra que los estudiantes de la Universidad del Magdalena provienen esencialmente de los estratos socioeconómicos de más bajos niveles de ingresos. Esta distribución muestra la vocación de la Universidad de cumplir con su propósito misional en la población con menos posibilidades de acceso a la educación superior, teniendo en cuenta el carácter público de la Universidad del Magdalena.

Tabla 7. Distribución según estrato socio-económico

Estrato Socioeconómico	SEXO				Frecuencia	Porcentaje del total
	Femenino		Masculino			
	Frecuencia	Porcentaje	Frecuencia	Porcentaje		
Estrato 1	2.921	43,46	2632	39,86	5553	41,68
Estrato 2	2.802	41,69	2772	41,98	5574	41,83
Estrato 3	877	13,05	1022	15,48	1899	14,25
Estrato 4	96	1,43	148	2,24	244	1,83
Estrato 5	21	0,31	22	0,33	43	0,32
Estrato 6	4	0,06	7	0,11	11	0,08
Total	6.721	100	6.603	100	13.324	100

Fuente: Elaboración Propia

4.2 Instrumento

En el desarrollo del presente trabajo se decidió utilizar los datos de una encuesta realizada, vía web, por el Grupo Investigación en Currículo y Evaluación (GICE) de la Universidad del Magdalena en el marco del proyecto de investigación Caracterización de los Estudiantes de la Universidad del Magdalena, el cuál fue financiado por el Fondo Patrimonial de Investigación-FONCIENCIAS de la misma institución educativa. Dicha encuesta se complementó posteriormente con algunos datos obtenidos a partir del Grupo de Admisiones, Registro y Control de la Universidad del Magdalena.

La idea básica es aprovechar la información disponible de esta encuesta para intentar vislumbrar las posibles relaciones existentes entre los distintos factores relativos a los aspectos demográficos, sociales, afectivos, económicos, académicos, recreativos y ambientales, en que se desarrolla el estudiante de la Universidad del Magdalena, y su Rendimiento Académico.

En la encuesta se considera entre otras, el municipio de procedencia del estudiante. Sin embargo, en este estudio se categorizó esta variable para tener en cuenta si el estudiante proviene o no de un municipio certificado en materia educativa. De acuerdo con el Artículo 20 de la Ley 715 de 2001, esta certificación consiste en otorgar a municipios de más de 100.000 habitantes la administración de la educación primaria y secundaria en su jurisdicción. Aquellos municipios con menos de 100.000 habitantes que cumplan con los requisitos que señale el reglamento en materia de capacidad técnica, administrativa y financiera podrán certificarse.

En la tabla 8 se muestran las variables consideradas en este trabajo, así como la categorización o descripción de las variables tal y como se llevó a cabo la encuesta. La variable dependiente en el estudio será el Rendimiento Académico categorizada en 3 niveles, y como puede apreciarse en la tabla, 33 posibles variables explicativas.

Algunas de las variables explicativas tienen varias categorías y muy pocas observaciones en algunas de estas categorías. Por ello se procedió a disminuir el número de categorías agrupando individuos por características homogéneas mediante análisis clúster.

Por no extendernos demasiado se muestran aquí un par de ejemplos, para las variables Estado Civil y Ocupación Padre. El análisis clúster produce los dos dendogramas que se muestran en las figuras 3 y 4, indicándonos que Estado Civil puede resumirse en dos categorías (Solteros y Otros), mientras que Ocupación Padre en dos categorías (Empleado e Independiente, y Otros).

Procediendo de igual manera para el resto de variables, se obtuvo la categorización que se muestra en la tercera columna de la tabla 8, y que será utilizada para modelar las variables consideradas en el modelo tras haber llevado a cabo la reducción de categorías.

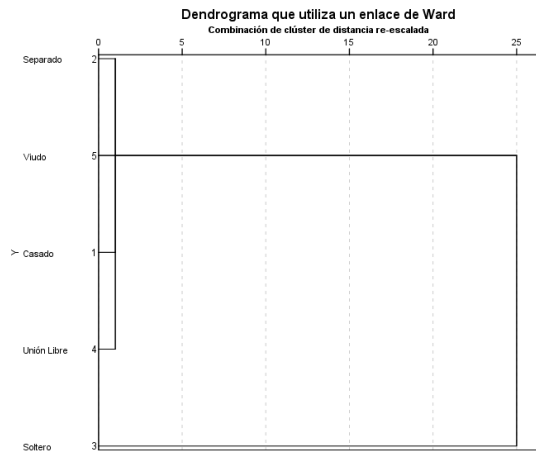


Figura 3. Dendrograma Estado Civil
Fuente: Elaboración Propia

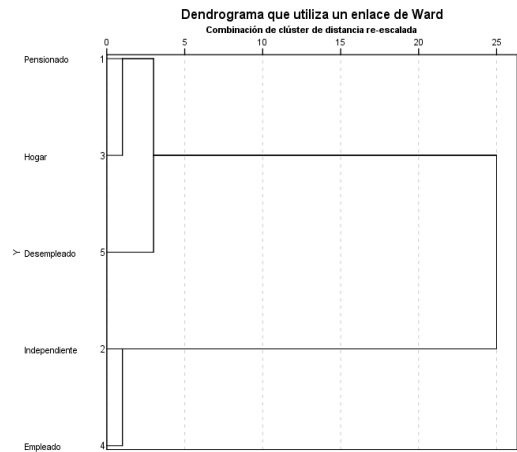


Figura 4. Dendrograma Ocupación Padre
Fuente: Elaboración Propia

Tabla 8. Variables consideradas en el modelo con su descripción y/o categorías

Variable	Categorías/Descripción	Categorías/Descripción Definitiva
Promedio Ponderado (Rendimiento Académico)	Variable que representa el promedio ponderado acumulado del estudiante. Está entre 320 - 500	Rendimiento Académico categorizado en 3 niveles. Básico, Alto y Superior.
Facultad	Facultad de Ciencias Básicas Facultad de Ciencias de la Educación Facultad de Ciencias de la Salud Facultad de Ciencias Empresariales y Económicas Facultad de Humanidades Facultad de Ingeniería	Tres Categorías. Categoría 1: Ciencias Básicas Categoría 2: Ciencias de la Educación, Ciencias de la Salud, Ciencias Empresariales y Económicas, y Humanidades. Categoría 3: Ingeniería
Estrato Socio-económico	Estrato 1 Estrato 2 Estrato 3 Estrato 4 Estrato 5 Estrato 6	Tres Categorías: Estrato 1 y 2 Estrato 3 Estrato 4, 5 y 6
Tipo Colegio	Público Privado	Público Privado
Certificación del Municipio de procedencia	Si No	Si No
Sexo	Masculino Femenino	Masculino Femenino
Jornada de su programa académico	Diurno Nocturno Mixto	Diurno Nocturno Mixto
Edad	Edad del estudiante (17 - 47 años)	Edad del estudiante (17 - 47 años)
Estado Civil	Casado Separado Soltero Unión Libre Viudo	Dos Categorías: Soltero Otros
Número de hijos	Número de hijos del estudiante (0 a 3 hijos)	Número de hijos del estudiante (0 a 3 hijos)
¿Con quién vives?	Alguien diferente a su familia Compañero o amigo Cónyuge Familia de su cónyuge Parientes Solo Familia de origen	Dos Categorías: Familia de origen No familia de origen
Ocupación de su padre	Desempleado Empleado Independiente Pensionado Hogar	Dos Categorías: Empleado e Independiente Desempleado, Pensionado y Hogar
Ocupación de su madre	Desempleado Empleado Independiente Pensionado Hogar	Dos Categorías: Empleado e Independiente Desempleado, Pensionado y Hogar
Número de personas a cargo	Número de personas que dependen económicamente del estudiante, está entre 0 - 5	Dos categorías: Con personas a cargo Sin personas a cargo

Nivel educativo padre	Ninguno Primaria Incompleta Primaria Completa Bachillerato incompleto Bachillerato completo Universitaria Incompleta Técnico Tecnólogo Universitaria Completa Diplomado Especialización Maestría Doctorado	Dos Categorías: Universitaria completa y más No universitaria
Nivel educativo madre	Ninguno Primaria Incompleta Primaria Completa Bachillerato incompleto Bachillerato completo Universitaria Incompleta Técnico Tecnólogo Universitaria Completa Diplomado Especialización Maestría Doctorado	Dos Categorías: Universitaria completa y más No universitaria
¿Cómo costea sus gastos educativos?	Auxilio empresarial Ayuda Familiar Beca Otra forma Préstamos Educativos Propio Sueldo	Dos Categorías: Propio Sueldo Otros
¿Desempeña alguna actividad económica?	Sí No	Sí No
Tiempo transcurrido entre la terminación del bachillerato y el ingreso a la universidad	Inmediatamente Un semestre después Un año después Más de un año después	Dos Categorías: Inmediatamente No Inmediatamente
Tipo Bachillerado	Académico Comercial Otro Industrial Pedagógico Técnico Validación	Dos Categorías: Académico No académico
¿Ha realizado otros estudios?	Sí No	Sí No
¿Está realizando otros estudios?	Sí No	Sí No
¿Por qué motivo ha interrumpido sus estudios universitarios?	Desastre natural Desplazamiento forzado Dificultad laboral Dificultad personal Dificultades académicas Dificultades económicas No ha interrumpido sus estudios	Dos Categorías: Quienes han interrumpido estudios Quienes no han interrumpido estudios

¿Se siente respetado y orientado por sus profesores?	Nunca Ocasionalmente Casi Siempre Siempre	Dos Categorías: Nunca y Ocasionalmente Casi Siempre y Siempre
¿Cómo complementaría su formación universitaria?	Autoformación Cursos de Actualización Investigaciones Otra carrera Postgrado	Dos Categorías: Autoformación, Cursos de Actualización Investigaciones ,Otra carrera, Postgrado
¿Cuándo tiene conflictos personales a quién acude?	Amigos Cónyuge Madre Novio(a) Otro Familiar Padre Psicólogo Sacerdote Profesor	Dos Categorías: Familia y Amigos Otros
¿Cómo se siente usted en su relación con respecto a su familia?	Mal Regular Bien Muy bien	Dos Categorías: Mal y Regular Bien y Muy bien
¿Cómo se siente usted en su relación con respecto a sus amigos?	Mal Regular Bien Muy bien	Dos Categorías: Mal y Regular Bien y Muy bien
¿Cómo se siente usted en su relación con respecto a la universidad?	Mal Regular Bien Muy bien	Dos Categorías: Mal y Regular Bien y Muy bien
¿Cómo se siente usted en su relación con respecto a Santa Marta?	Mal Regular Bien Muy bien	Dos Categorías: Mal y Regular Bien y Muy bien
¿Cómo se siente usted en su relación con respecto al país?	Mal Regular Bien Muy bien	Dos Categorías: Mal y Regular Bien y Muy bien
Actividades que realiza en sus ratos libres.	Actividades físicas Bailar Caminatas Dormir Hacer Deporte Cine Leer Escuchar Música Otra Pasear Actividades Artísticas Ver TV Videojuegos Visitar amigos/familiares	Dos categorías: Categoría 1: Actividades físicas, hacer deporte, leer, actividades artísticas y caminatas. Categoría 2: Bailar, Dormir, Cine, Escuchar Música, Otra, Pasear, Ver TV, Videojuegos, Visitar amigos.
¿Pertenece usted a grupos?	Académicos Culturales Deportivos Ninguno Religioso Sociales	Dos Categorías: Categoría 1: Académicos, Culturales y Deportivos Categoría 2: Ninguno, Religioso y Sociales.

¿Cómo califica la calidad de los programas y servicios que ofrece Bienestar Universitario?	Muy Mala Mala Regular Buena Muy Buena	Dos categorías. Categoría 1: Mala y Muy Mala Categoría 2: Regular, Buena y Muy Buena.
--	---	---

Fuente: Elaboración Propia

5 RESULTADOS

5.1 Análisis de la muestra

La muestra con la cual realizaremos la investigación consta de 623 estudiantes entrevistados de los cuales únicamente se pudo trabajar con 455, ya que 168 estudiantes están matriculados en su primer semestre académico, por lo tanto no tienen notas registradas.

En la Universidad del Magdalena las notas van de cero (0) a 500 puntos. El Reglamento Estudiantil de la Universidad del Magdalena (Acuerdo Superior No. 008 de 2003) establece una calificación promedio acumulada ponderada por los créditos de las materias cursadas de 320 puntos para que un estudiante pueda permanecer en la institución, es decir aquellos estudiantes cuyo promedio académico ponderado sea inferior a dicho valor quedarán excluidos de la universidad y no podrán seguir estudiando en el próximo semestre académico por bajo desempeño académico. A dichos estudiantes en el argot de la comunidad universitaria se les conoce como estudiantes **F.B.R.A (Fuera de la Universidad por Bajo Rendimiento Académico)**. Dichos estudiantes F.B.R.A no han sido considerados en la muestra, que como se ha comentado está constituida por 455 estudiantes matriculados en el periodo académico 2014-02 (segundo semestre académico del año 2014).

Recientemente hubo un cambio en el Reglamento Estudiantil, mediante el cual aquellos alumnos con Promedio Académico Ponderado Acumulado del estudiante entre 300 y 319 puntos, ambos inclusive, pueden permanecer en la Universidad en estado condicional durante dos semestres académicos, con el compromiso de subir su promedio como mínimo a 320 puntos. De no ser así será retirado de la institución educativa. Sin embargo, en el momento de realización de la encuesta este cambio no había sido realizado.

Para poder realizar el estudio, se procedió a categorizar la variable respuesta en tres grupo que se muestran en la tabla 9 siguiente:

Tabla 9. Categorización del Rendimiento Académico

Básico	Alto	Superior
320 - 359	360 - 379	380 - 500

Fuente: Elaboración Propia

Para dicha categorización del promedio acumulado en tres grupos ordenados (variable ordinal), se tuvo en cuenta lo establecido en el Título Octavo del Reglamento Estudiantil de la Universidad del Magdalena, en lo referente a los estímulos otorgados a los estudiantes según su desempeño académico. El Reglamento Estudiantil de la Universidad del Magdalena puede ser consultado en la siguiente dirección web https://www.unimagdalena.edu.co/Content/Public/Docs/Entrada_Direccion8/adjunto_1023-20180509151728_256.pdf En dicho reglamento se establecen los siguientes estímulos:

“ARTÍCULO 155. **MATRÍCULA DE HONOR.** A un (1) estudiante de cada semestre de cada programa de Pregrado, siempre y cuando cumpla con los siguientes requisitos:

- a. Haber cursado en el semestre anterior catorce (14) o más créditos académicos.
- b. No haber habilitado, repetido ni reprobado ninguna asignatura durante el semestre; incluyendo cursos especiales durante el período.
- c. Haber obtenido el mejor promedio ponderado del semestre, no inferior a cuatrocientos veinticinco (425) puntos, para su programa académico.
- d. No tener en la hoja de vida sanciones disciplinarias para el semestre académico analizado.

ARTÍCULO 157. Los estudiantes que habiendo cursado un mínimo de catorce (14) créditos académicos, no hayan repetido, habilitado o perdido cátedras o cursos incluyendo los cursos especiales y tengan un promedio semestral igual o superior a trescientos ochenta (380) puntos tendrán los siguientes estímulos:

- a. El primer puesto en su respectivo semestre y que no cumpla con los requisitos para obtener matrícula de honor, tendrá derecho a un descuento del cien por ciento (100%) del valor de su matrícula.

- b. El segundo y tercer mejor promedio semestral en su respectivo semestre, tendrán derechos a descuentos en el valor de su matrícula del setenta y cinco (75%) y del cincuenta (50%), respectivamente.

ARTÍCULO 158. EXONERACIÓN DE PAGOS DE DERECHOS DE MATRÍCULA. *La Universidad exonerará del pago semestral del derecho de matrícula a los estudiantes que integren los grupos artísticos, culturales, académicos, científicos y deportivos Ad Honorem debidamente organizados y reconocidos por la Universidad, siempre y cuando cumplan con los siguientes requisitos:*

- a. *Haber sido galardonado o haber ocupado primer, segundo o tercer lugar en una competencia o evento artístico, cultural, científico, y/o deportivo, a nivel regional, nacional o internacional, de reconocimiento en su respectiva disciplina durante el semestre anterior.*
- b. *Haber cursado en el semestre anterior a la aplicación de la beca del 60% o más créditos académicos.*
- c. *No haber reprobado ninguna asignatura durante el semestre anterior, incluyendo cursos especiales de ese periodo.*
- d. *Haber obtenido un promedio ponderado acumulado, igual o superior a trescientos sesenta (360) puntos.*
- e. *No tener en la hoja de vida sanciones disciplinarias para el semestre académico analizado.*

PARÁGRAFO. La exoneración del pago de los derechos de matrícula por la actuación sobresaliente en el deporte o en los grupos artísticos, la hará el Consejo Académico de la Universidad, previa recomendación de Bienestar Universitario. La recomendación deberá basarse en la participación, representación e imagen que a diferentes niveles brinden de la Universidad los estudiantes integrantes de estas actividades. La sola pertenencia a ellas no implica la concesión de este reconocimiento.

ARTÍCULO 159. Los estudiantes que han cursado por lo menos tres (3) semestres académicos y tienen promedios ponderados acumulados iguales o superiores a cuatrocientos (400) puntos, podrán participar en los concursos que en cada programa académico se autoricen para MONITORES ACADÉMICOS. La Universidad realizará un reconocimiento económico por los servicios como monitor en cuantía que será establecida por el Consejo Superior de la Institución.

ARTÍCULO 160. Los estudiantes que han cursado por lo menos tres (3) semestres académicos y tienen promedios ponderados semestrales iguales o superiores a trescientos ochenta (380) puntos podrán participar en los concursos que en la Universidad se autoricen para MONITORES ADMINISTRATIVOS. La Universidad realizará un reconocimiento económico hasta por trescientas (300) horas semestrales por los servicios como monitor administrativo por una cuantía que será establecida por el Consejo Superior de la institución.”

En la base de datos de los 13.324 estudiantes matriculados en el periodo 2014-02 sólo 291 estudiantes tienen un promedio académico acumulado ponderado mayor o igual a 425 puntos, correspondiendo esto al 2,18%. En la muestra de 455 estudiantes sólo un estudiante tiene un promedio académico acumulado ponderado igual o superior a 425 puntos, por ende sólo un estudiante en la muestra puede optar a la Matrícula de Honor.

Considerando que únicamente a partir de un puntaje de 360 puntos un estudiante podría optar a algún tipo de estímulo por su desempeño académico se consideró como primera categoría del Rendimiento Académico la comprendida entre 320 – 359 puntos, a la cual le hemos llamado Rendimiento a Nivel Básico. El segundo nivel, al cual hemos llamado Alto, está comprendido entre 360 – 379 puntos, esto considerando que el Reglamento Estudiantil ofrece la posibilidad de que un determinado estudiante pueda optar a ser monitor a partir de 380 puntos, por ende la siguiente categoría la cual hemos denominado Rendimiento a Nivel Superior la comprendida entre 380 – 500 puntos.

En Colombia a nivel de educación universitaria no existe un decreto similar al **Real Decreto 1125/2003**, de 5 de septiembre, por el que se establece el sistema europeo de créditos y el sistema de calificaciones en las titulaciones universitarias de carácter oficial y validez en todo el territorio nacional. En el citado Real Decreto, se instituye en su Artículo 5. Sistema de calificaciones, que “*Los resultados obtenidos por el alumno en cada una de las materias del plan de estudios se calificarán en función de la siguiente escala numérica de 0 a 10, con expresión de un decimal, a la que podrá añadirse su correspondiente calificación cualitativa:*

<i>Suspenso (SS)</i>	<i>Aprobado (AP)</i>	<i>Notable (NT)</i>	<i>Sobresaliente (SB)</i>
<i>0 - 4,9</i>	<i>5,0 - 6,9</i>	<i>7,0 - 8,9</i>	<i>9,0 - 10</i>

Sin embargo, el Ministerio de Educación Nacional de Colombia en el **Decreto No. 1290 de 2009** reglamenta la evaluación del aprendizaje y promoción de los estudiantes de los niveles de educación básica y media, establece en su Artículo 5: *“Escala de valoración nacional: Cada establecimiento educativo definirá y adoptará su escala de valoración de los desempeños de los estudiantes en su sistema de evaluación. Para facilitar la movilidad de los estudiantes entre establecimientos educativos, cada escala deberá expresar su equivalencia con la escala de valoración nacional:*

- *Desempeño Superior*
- *Desempeño Alto*
- *Desempeño Básico*
- *Desempeño Bajo*

La denominación desempeño básico se entiende como la superación de los desempeños necesarios en relación con las áreas obligatorias y fundamentales, teniendo como referente los estándares básicos, las orientaciones y lineamientos expedidos por el Ministerio de Educación Nacional y lo establecido en el proyecto educativo institucional. El desempeño bajo se entiende como la no superación de los mismos.”

Este es el motivo por el cual hemos denominado como Básico, Alto y Superior a los tres niveles en que se ha categorizado al Rendimiento Académico.

La variable Estrato socio-económico se ha categorizado en tres grupos, un grupo por los estudiantes de los estratos 1 y 2, otro por los estudiantes del estrato 3 y otro por los estudiantes de los estratos 4, 5 y 6.

La variable Facultad se categorizó en tres grupos (ver tabla 8).

Las variables Número de Hijos y Número de Personas a Cargo tienen un coeficiente de correlación de 0,75 por este motivo al momento de modelar la Regresión Logística se debe utilizar sólo una de ellas, habiéndose escogido en este caso, el Número de Personas a Cargo, la cual se ha categorizado en dos grupos (uno con los estudiantes que tienen personas a cargo y otro conformado por los estudiantes que no tienen personas a cargo).

De igual manera, dado que cada Programa Académico está vinculado a una Facultad se debe decidir cuál de estas dos variables utilizar en el estudio. Debido a que la variable Facultad posee menos categorías se decidió utilizar esta variable y no la variable Programa Académico.

En los datos correspondientes a la variable Tipo de Colegio se encontró que el 74,7% de los estudiantes terminaron su bachillerato en colegios públicos, evidenciándose una vez más la vocación de la Universidad del Magdalena de atender a los estudiantes con menores niveles de ingresos.

En cuanto a la variable Estrato Socio-económico se encontró que el 79,3% de los estudiantes proceden de los estratos 1 y 2, un 17,6% proviene del estrato 3, y solo un 3,1% proceden de los estratos 4, 5 y 6.

En la muestra el 76,3% de los estudiantes entrevistados proceden de municipios con certificación en educación, el 52,1% son de género masculino, el 40,44% tienen un Rendimiento Académico en el nivel Básico, un 29,67% en el nivel Alto y un 29,89% en el nivel Superior. Por su parte el 47% de los estudiantes entrevistados estudian en jornada Diurna, el 16,3% estudian en la jornada Nocturna, mientras que el 36,7% asisten a clases tanto en la jornada Diurna como Nocturna. En cuanto a la edad, los estudiantes entrevistados tienen 23 años en promedio, con una moda de 21 años. El 71,9% tienen una edad entre 19 y 24 años. El 81,8% de los estudiantes entrevistados son solteros. El 82% no tiene hijos, mientras que el 73,8% no tiene personas a cargo, 59,6% de los estudiantes entrevistados viven con su familia de origen.

El 73,2% de los padres de los estudiantes están empleados o trabajan como independientes (autónomos), mientras que para las madres este porcentaje se reduce a un 40,4%.

En cuanto al nivel educativo alcanzado por los padres, se encontró que el 18% de los padres tienen por lo menos un nivel de educación superior debidamente culminado y graduado, mientras que para las madres este valor es del 16,7%. Aproximadamente el 50% de los estudiantes entrevistados costea sus estudios por medio de recursos de su familia. El 38,9% de los estudiantes desempeñan alguna actividad que les representa ingresos económicos. El 36,7% de los estudiantes ingresaron a la Universidad inmediatamente después de terminar sus estudios de educación secundaria

(bachillerato), mientras que el 26,2% ingresaron a la Universidad más de un año después de alcanzar su título de bachiller. El 60% de los estudiantes se graduaron de bachiller académico, el 14,1 bachiller comercial y el 14,3% de bachiller técnico. El 98% de los estudiantes realizaron su bachillerato en jornada diurna.

El 66,6% de los estudiantes nunca ha interrumpido sus estudios universitarios, el 40,2% ha realizado algún tipo de estudio además de su educación secundaria, por su parte solo el 10,1% está realizando otros estudios además de sus estudios universitarios. Por otra parte, el 86,6% de los estudiantes entrevistados se sienten respetados y orientados por sus profesores, mientras que el 82,9% de los estudiantes piensan seguir su formación profesional mediante postgrados u otra carrera profesional.

Aproximadamente el 50% de los estudiantes acuden a sus padres cuando tienen algún tipo de conflicto personal, mientras que el 24% acude a sus amigos y el 26,8% acude a otras personas (cónyuge, novio(a), otro familiar, profesor, psicólogo o sacerdote).

En cuanto a los sentimientos en relación a su familia, el 90,3% se sienten Bien o Muy Bien, el 92,8% se sienten Bien o Muy Bien en cuanto a la relación con sus amigos, el 86,6% se sienten Bien o Muy Bien en cuanto a la relación con la Universidad del Magdalena, el 79,8% se sienten Bien o Muy Bien en cuanto a la relación con la ciudad de Santa Marta, el 71,5% se sienten Bien o Muy Bien en cuanto a la relación con Colombia.

Respecto de las actividades que realizan los estudiantes entrevistados en sus tiempos libre encontramos que el 48,5% llevan a cabo actividades relacionadas con el ejercicio físico, leer o artísticas (danza, coro, teatro), mientras que el 51,5% realizan actividades que hemos denominado de ocio (bailar, dormir, ir a cine, escuchar música, pasear, ver televisión, videojuegos, visitar amigos o familiares, y otras).

En cuanto a la vinculación de los estudiantes entrevistados a grupos, encontramos que el 24,4% está vinculado a grupos académicos, culturales o deportivos, mientras que los restantes están vinculados a grupos religiosos, sociales o a ningún tipo de grupo.

Finalmente, en lo referente a la apreciación de los estudiantes entrevistados sobre la calidad de los servicios prestados por Bienestar Estudiantil de la Universidad del Magdalena, se encontró que el 90,1% tiene una opinión Regular, Buena o Muy Buena

de los servicios ofrecidos, mientras que el 9,9% tienen una apreciación Mala o Muy Mala de los servicios ofrecidos.

5.2 Modelo de Regresión Logística Ordinal

Para ajustar el modelo de Odds Proporcionales de Regresión Logística Ordinal según la parametrización presentada en la ecuación (49) utilizamos la función *polr* (proportional odds logistic regression) del paquete MASS en R, pues esta función admite el uso de la función *step* o *stepAIC*, las cuales permiten implementar la selección de variables mediante *stepwise* permitiendo ajustar el parámetro que define si se utiliza como criterio AIC o BIC.

Debido a la gran cantidad de variables explicativas consideradas (33) en este trabajo se procedió a modelar la selección de variables mediante las estrategias por pasos hacia adelante, hacia atrás y mixta, y utilizando los criterios *AIC* y *BIC*. Los resultados obtenidos al ajustar los modelos mediante las tres estrategias de selección por pasos son los mismos tanto para el criterio *AIC* como para *BIC*.

Los resultados utilizando el criterio *AIC* muestran que 10 variables son significativas al 95% de confianza, mientras que para *BIC* este número se reduce a 6 variables. La tabla 10 siguiente muestra las variables significativas al 95%.

Tabla 10. Variables significativas mediante AIC y BIC

AIC	BIC
Facultad	Facultad
Edad	
¿Con quién vives?	
¿Tiene personas a cargo?	
¿Cómo costea sus gastos educativos?	¿Cómo costea sus gastos educativos?
¿Ha interrumpido estudio?	¿Ha interrumpido estudio?
¿Ha realizado otros estudios?	¿Ha realizado otros estudios?
¿Se siente respetado y orientado por sus profesores?	¿Se siente respetado y orientado por sus profesores?
¿Cómo complementaría su formación universitaria?	
¿Cómo se siente usted en su relación con respecto a su familia?	¿Cómo se siente usted en su relación con respecto a su familia?

Fuente: Elaboración Propia

Upton (2017) manifiesta que, “dado que ninguno de los modelos que consideremos proporcionará una descripción perfecta, como lo sugiere *BIC*, el *AIC* debería ser la medida utilizada”.

En este mismo orden de ideas Chakrabarti y Ghosh (2011), consideran que “AIC sobresale en aquellos problemas en los que todos los posibles modelos son incorrectos o hay como máximo un modelo correcto, mientras que BIC sobresale si el modelo correcto está entre los posibles modelos”.

Dziak, Coffman, Lanza y Li (2012), consideran que “desde una perspectiva AIC y para una muestra grande, un modelo con bajo ajuste se considera peor que un modelo sobreajustado. En contraste, con BIC se está dispuesto a correr un mayor riesgo de elegir un modelo demasiado pequeño, para mejorar la probabilidad de elegir el modelo verdadero (el modelo correcto más pequeño). BIC considera que el subajuste y el sobreajuste son igualmente indeseables, mientras que AIC considera que los modelos subajustados son más indeseables que los modelos sobreajustados, a menos que el tamaño de muestra sea muy pequeño. Ninguna de las dos perspectivas es siempre correcta o incorrecta. AIC no es un BIC defectuoso, ni viceversa”.

Dada las consideraciones anteriores y la posible importancia en el rendimiento académico del estudiante de los factores referidos a *con quien vive el estudiante, si tiene o no personas a cargo, y al interés del estudiante por complementar su formación profesional*, en este estudio decidimos quedarnos con el modelo ajustado mediante el criterio *AIC*.

En la tabla 11 mostramos los resultados obtenidos mediante *AIC*. En ella se observa que todas las variables son significativas al 95%.

Tabla 11. Resultados modelo ajustado mediante AIC

Variable	Coefficiente	Desviación Estándar	Valor Z	Valor p	OR	IC 95%	
Facultad-F2	1,247	0,198	6,297	3,038E-10	3,479	2,368	5,149
Edad	0,058	0,026	2,272	2,307E-02	1,060	1,009	1,116
ViveCon-Su Familia	-0,420	0,196	-2,145	3,195E-02	0,657	0,447	0,964
TienePersonasACargo-Si	-0,793	0,261	-3,036	2,394E-03	0,452	0,270	0,752
CosteaEducacion-Otros	0,541	0,195	2,775	5,518E-03	1,718	1,174	2,522
InterrumpidoEstudios-Si	-1,044	0,221	-4,718	2,382E-06	0,352	0,227	0,541
HaRealizadoOtrosEstudios-Si	0,616	0,199	3,089	2,006E-03	1,851	1,254	2,742
RespetoOrientacionProfes-Siempre	0,782	0,295	2,649	8,076E-03	2,186	1,237	3,953
ComplementaríaEstudios-Si	0,612	0,258	2,375	1,755E-	1,843	1,118	3,074

SentimientoFamilia-No Bien	-0,834	0,355	-2,351	1,872E-02	0,434	0,213	0,859
Básico Alto	2,400	0,691	3,474	5,127E-04			
Alto Superior	3,946	0,706	5,587	2,307E-08			
Residual Deviance: 862,8227							
AIC: 886,8227							

Fuente: Elaboración Propia

En este modelo, dado que la variable respuesta es ordinal con 3 categorías, habrá 2 ecuaciones.

En el modelo anterior, el coeficiente de Edad es $\beta_{Edad} = 0,058$ y $e^{\beta_{Edad}} = 1,060 = \hat{\lambda}_{Edad}$, el Odd Ratio de Edad (una medida de asociación) nos indica que por cada año que aumente la edad del estudiante la probabilidad de que este mejore su rendimiento, es decir, que pase a la categoría inmediatamente superior es un 6% mayor, manteniendo constantes todas las otras variables. Esta baja probabilidad junto con el hecho de que, como se ha mencionado antes, cuando hay al menos una covariable continua el número de patrones de covariables se aproxima al número de observaciones, pudiendo ser un problema al momento de evaluar el ajuste del modelo, se optó por no considerar la variable Edad y por lo tanto eliminarla del modelo final a utilizar.

Esto concuerda con los resultados del estudio llevado a cabo por Ebeñuwa-Okoh (2010), quien concluye que la edad, estatus financiero y género no tienen influencia en el rendimiento académico. Por su parte, Cruz, Medina, Vázquez, Espinosa y Antonio (2014) concluyen que existe una correlación muy débil entre el Rendimiento Académico con el nivel socioeconómico del estudiante.

Tras haber eliminado variable Edad, se procedió a volver a ajustar el modelo, obteniéndose los resultados que se muestran en la tabla 12, y que van a constituir los coeficientes del modelo a utilizar en este estudio.

Tabla 12. Resultados modelo AIC sin variable Edad

Variable	Coefficiente	Desviación Estándar	Valor Z	Valor p	OR	IC 95%	
Facultad-F2	1,298	0,197	6,603	4,03E-11	3,661	2,499	5,403
ViveCon-Su Familia	-0,440	0,195	-2,260	2,38E-02	0,644	0,439	0,942
TienePersonasACargo-Si	-0,588	0,242	-2,424	1,53E-02	0,556	0,344	0,892
CosteaEducacion-Otros	0,573	0,194	2,948	3,20E-03	1,774	1,213	2,601
InterrumpidoEstudios-Si	-0,923	0,213	-4,333	1,47E-05	0,397	0,261	0,601
HaRealizadoOtrosEstudiosSi	0,678	0,197	3,438	5,86E-04	1,969	1,340	2,904
RespetoOrientacionProfes-Siempre	0,789	0,293	2,689	7,17E-03	2,201	1,250	3,965
ComplementariaEstudios-Si	0,549	0,254	2,157	3,10E-02	1,731	1,056	2,868

SentimientoFamilia-No Bien	-0,935	0,350	-2,675	7,47E-03	0,392	0,194	0,767
Básico Alto	1,137	0,406	2,803	5,06E-03			
Alto Superior	2,669	0,421	6,338	2,32E-10			
Residual Deviance: 868,1726							
AIC: 890,1726							

Fuente: Elaboración Propia

Antes de proceder a la interpretación de los coeficientes de las variables explicativas, debe evaluarse si el modelo resultante proporciona un mejor ajuste que un modelo con solo intercepto (llamado modelo nulo). Es decir, se pretende responder a la pregunta de si las variables explicativas incluidas en el modelo contribuyen a explicar los resultados presentados en la variable respuesta. Para ello, se evaluó la significación de modelo mediante la prueba de razón de verosimilitud para la significancia global de los p coeficientes de las variables independientes comentada en la sección 3.2.2.2. Ésta prueba se basa en la estadística G dada en la ecuación (14), y es la diferencia entre -2 veces logaritmo natural de la verosimilitud para el modelo nulo y la del modelo final. Cuanto más pequeño sea el valor de este último mejor será el ajuste.

Bajo la hipótesis nula de que los p coeficientes para las variables explicativas en el modelo son iguales a cero, la distribución del estadístico G sigue una distribución chi-cuadrado con p grados de libertad. Por ende, se desea rechazar la hipótesis nula de que los p coeficientes para las variables explicativas son iguales a cero, por lo que un p -valor inferior a 0,05 implica que modelo ajustado proporciona un mejor ajuste que el modelo nulo.

En este caso el estadístico G de la ecuación (14) se escribe como:

$$G = -2 \ln \left[\frac{\text{Verosimilitud del modelo solo intercepción}}{\text{Verosimilitud del modelo ajustado}} \right] \quad (55)$$

Los resultados se muestran en la tabla 13, donde puede apreciarse (p -valor = 0,000), por lo que el estadístico es significativo, indicado que el modelo proporciona una mejora significativa con respecto al modelo de referencia o modelo nulo, el modelo sin ninguna variable explicativa.

Esto se ha programado en R mediante la siguiente línea de código:

```
anova(modelo.nulo, modelo.final, test = "Chisq")
```

Tabla 13. Información de ajuste de los modelos

Modelo	-2Logaritmo de la verosimilitud	Chi-cuadrado	gl	Valor P
--------	---------------------------------	--------------	----	---------

Sólo intersección	989,7080	121,535	9	0,000
Final	868.1726			

Fuente: Elaboración Propia

En segundo lugar, se debe comprobar el cumplimiento de la hipótesis de paralelismo para todos los coeficientes del modelo (Odds Proporcionales). Para esto se realiza la prueba de Brant, la cual nos permite comprobar la hipótesis nula de que los coeficientes de regresión son los mismos para todas las categorías de la variable respuesta. De tal manera que lo que se desea es aceptar la hipótesis nula, por lo tanto se requiere que el p-valor sea superior a 0,05. La prueba de Brant se llevó a cabo mediante el paquete “brant” de R, obteniéndose los resultados que se muestran en la tabla 14, confirmándose el cumplimiento de dicha hipótesis.

Tabla 14. Prueba de Brant (Lineas Paralelas)

H₀: Parallel Regression Assumption holds

Fuente: Elaboración	Prueba para	X ²	gl	Valor P	Propia
Como se ha Regresión Logística definición de residual de Regresión Lineal. residuos ampliamente	Omnibus	7,37	9	0,60	mencionado antes, en Ordinal no existe una como en los modelos En su lugar los utilizados para validar
	Facultad	0,07	1	0,79	
	ViveCon	0,03	1	0,87	
	TienePersonasACargo	0,01	1	0,94	
	CosteaEducacion	0,59	1	0,44	
	InterrumpidoEstudios	1,29	1	0,26	
	HaRealizadoOtrosEstudios	0,03	1	0,86	
	RespetoOrientacionProfes	1,26	1	0,26	
	ComplementariaEstudios	0,93	1	0,33	
	SentimientoFamilia	1,37	1	0,24	

el modelo de RLO son los Residuos de Pearson y los Residuos de Devianza. Por ello, en tercer lugar, se procedió validar la bondad de ajuste del modelo a través de las pruebas *chi-cuadrado de Pearson*, *chi-cuadrado de devianza* y *chi-cuadrado de Hosmer-Lemeshow* presentadas en la sección 3.2.2.3.1. En estas pruebas, como en toda prueba de bondad de ajuste, lo que se desea es que no haya significación, es decir aceptar dichas hipótesis nulas de que el modelo propuesto se ajusta a lo observado. Los resultados de estas tres pruebas se muestran en la tabla 15.

Los estadísticos para las pruebas *chi-cuadrado de Pearson* y *devianza* fueron obtenidos en SPSS, mientras que *Hosmer-Lemeshow* se obtuvo con la función *logitgof* del paquete ‘*generalhoslem*’ en R, pudiéndose confirmar que el modelo ajustado es adecuado para modelar los datos de Rendimiento Académico, es decir, se acepta la hipótesis nula de que el modelo ajustado se ajusta a lo observado en los datos.

Tabla 15. Pruebas de Bondad de Ajuste del Modelo

Fuente:	Prueba	Chi-cuadrado	gl	Valor P	Elaboración Propia
	Pearson	278,026	309	0,897	
	Devianza	297,069	309	0,677	
	Hosmer-Lemeshow	16,299	17	0,503	

Los resultados del estudio de los residuales de Pearson y los de la Devianza de la tabla anterior indican la no existencia de problemas con dichos residuales y que el modelo se ajusta suficientemente bien a los datos.

Finalmente, se procedió a estimar los estadísticos Pseudo R^2 , que se obtuvieron mediante la función *pseudoR2* del paquete *DescTools* en R. Los resultados se muestran en la tabla 16 siguiente.

Tabla 16. Pseudo R^2

	McFadden	CoxSnell	Nagelkerke
Fuente: Elaboración Propia	0.123	0.2344	0.2644

Los valores de los estadísticos Pseudo R^2 mostrados en la tabla anterior dan la impresión de un pobre ajuste, dada la interpretación del R^2 en los modelos de regresión lineal. Sin embargo, como ya se comentó previamente, según Faraway (2016), los valores Pseudo R^2 no se deben interpretar en los mismos términos que en una regresión normal, pues es muy frecuente encontrar valores pequeños para estos Pseudo R^2 aun cuando el modelo sea bueno, siendo mejor evitar este estadístico, excepto con el fin de comparar modelos.

Tras haber comprobado la significación del modelo global, se procedió a la interpretación de los coeficientes del modelo, reemplazando los valores obtenidos para los coeficientes del modelo ajustado en la ecuación (50), obteniéndose los siguientes modelos:

$$g_{Básico}(x) = \ln \left[\frac{Pr(\text{Rendimiento} \leq \text{Básico} | x)}{Pr(\text{Rendimiento} > \text{Básico} | x)} \right] = \ln \left[\frac{\phi_{Básico}(x)}{\phi_{Alto}(x) + \phi_{Superior}(x)} \right] \quad (56)$$

$$g_{Básico}(x) = 1,137 - 1,298 * Facultad + 0,440 * ViveCon + 0,588 * TienePersonasACargo - 0,573 * CosteaEduca$$

$$g_{Alto}(x) = \ln \left[\frac{Pr(\text{Rendimiento} \leq \text{Alto} | x)}{Pr(\text{Rendimiento} > \text{Alto} | x)} \right] = \ln \left[\frac{\phi_{Básico}(x) + \phi_{Alto}(x)}{\phi_{Superior}(x)} \right] \quad (57)$$

$$g_{Alto}(x) = 2,669 - 1,298 * Facultad + 0,440 * ViveCon + 0,588 * TienePersonasACargo - 0,573 * CosteaEduca$$

Utilizando la ecuación (24) obtenemos que,

$$Pr(\text{Rendimiento} \leq \text{Básico} | x) = \frac{1}{1 + e^{-g_{Básico}(x)}} \quad (58)$$

$$Pr(\text{Rendimiento} = \text{Básico} | x) = \frac{1}{1 + e^{-g_{\text{Básico}}(x)}} \quad (59)$$

$$Pr(\text{Rendimiento} \leq \text{Alto} | x) = \frac{1}{1 + e^{-g_{\text{Alto}}(x)}} \quad (60)$$

Ya que tenemos tres (3) categorías para la variable respuesta, $Pr(\text{Rendimiento} \leq \text{Superior}) = 1$

De aquí que:

$$Pr(\text{Rendimiento} \leq \text{Alto} | x) = Pr(\text{Rendimiento} = \text{Básico} | x) + Pr(\text{Rendimiento} = \text{Alto} | x) \quad (61)$$

$$Pr(\text{Rendimiento} = \text{Alto} | x) = Pr(\text{Rendimiento} \leq \text{Alto} | x) - Pr(\text{Rendimiento} \leq \text{Básico} | x) \quad (62)$$

$$Pr(\text{Rendimiento} = \text{Alto} | x) = \frac{1}{1 + e^{-g_{\text{Alto}}(x)}} - \frac{1}{1 + e^{-g_{\text{Básico}}(x)}} \quad (63)$$

$$Pr(\text{Rendimiento} = \text{Superior}) = 1 - Pr(\text{Rendimiento} \leq \text{Alto}) = 1 - \frac{1}{1 + e^{-g_{\text{Alto}}(x)}} = \frac{1}{1 + e^{g_{\text{Alto}}(x)}} \quad (64)$$

Según los parámetros del modelo ajustado mostrados en la tabla 12, el coeficiente de regresión para la categoría 2 de Facultad (la categoría 1 es la categoría de referencia) es $\beta_{\text{Facultad}} = 1,298$ y $e^{\beta_{\text{Facultad}}} = 3,661 = \zeta_{\text{Facultad}}$. El Odds Ratio de Facultad significa que para los estudiantes de las facultades de Educación, Salud, Empresariales y Humanidades (categoría 2 de la variable Facultad), la probabilidad de un (mejor) rendimiento (más alto) es 3,66 veces que para los estudiantes de las facultades de Ciencias Básicas o de Ingeniería, manteniendo constantes todas las otras variables.

El coeficiente para la variable que hace referencia a la pregunta *¿Con quién vives?*, que ha sido categorizado en dos grupos (*Familia de Origen* y *Otros*, siendo esta la categoría de referencia), el coeficiente de regresión para la categoría *Familia de Origen* es $-0,440$ y el Odds Ratio es $0,644$, lo que significa que para los estudiantes que viven con su familia de origen la probabilidad de un mejor rendimiento es $0,644$ veces que para los estudiantes que viven con otros, manteniendo constantes todas las otras variables. Es decir, que los estudiantes que viven con otros (que no viven con su familia de origen) tienen mejor desempeño académico, manteniendo constantes todas las otras variables.

En cuanto a la variable que hace referencia a si el estudiante *tiene personas a su cargo* que ha sido categorizada en dos grupos, (quienes tienen personas a cargo y quienes no tienen personas a cargo, siendo esta última la categoría de referencia), el coeficiente de regresión para la categoría “Si tiene personas a Cargo” es -0,588 y el Odds Ratio es 0,556, indicándonos que para los estudiantes que tienen personas a su cargo la probabilidad de un mejor rendimiento es 0,556 veces que para los estudiantes que no las tienen, manteniendo constantes todas las otras variables. Es decir que los estudiantes que no tienen personas a cargo tienen mejor desempeño académico, manteniendo constantes todas las otras variables. Dicho de otra manera, los estudiantes que tienen personas a su cargo la probabilidad de tener un peor rendimiento es 1,798 veces la de los estudiantes que no tienen personas a cargo.

Respecto de la variable referida al *financiamiento de los costos educativo del estudiante* cuya categoría de referencia es *Recursos Propios*, el coeficiente de regresión para los estudiantes que reciben ayuda de *otros* para costear sus estudios es 0,573 y su Odds Ratio es 1,774, lo que significa que las probabilidades de un rendimiento más alto aumentan en un factor de 1,774 para los estudiantes que reciben ayuda de otros para costear sus estudios en comparación con quienes deben pagar ellos mismos los costos de sus estudios.

En el caso de la variable que hace referencia a la pregunta *¿Por qué motivo ha interrumpido sus estudios universitarios?*, la cual se ha categorizado en dos grupos, los que *Han Interrumpido Estudios* y quienes *No han Interrumpido Estudios* (categoría de referencia), por lo que el coeficiente de regresión para los estudiantes que *Han Interrumpido Estudios* es -0,923 y su Odds Ratio es 0,397, lo que significa que las probabilidades de un rendimiento más alto aumentan en un factor de 2,519(1/0,397) para los estudiantes que *No han Interrumpido Estudios* en comparación con quienes *Han Interrumpido Estudios*.

En el caso de la variable si el estudiante *Ha Realizado Otros Estudios*, la categoría de referencia es quienes *No Han Realizado Otros Estudios*, por lo que el coeficiente de regresión para los estudiantes que *Sí Han Realizado Otros Estudios* es 0,678 y su Odds Ratio es 1,969, esto significa que las probabilidades de un rendimiento más alto aumentan en un factor de casi dos para los estudiantes que han realizado otros estudios en comparación con quienes no han estudiado previamente (en una universidad).

Para la variable referida a si el estudiante *Se Siente Respetado y Orientado por sus Profesores*, que se ha categorizado en dos grupos, (con las respuestas *Casi Siempre* o *Siempre*, y *Nunca* u *Ocasionalmente*, siendo esta última la categoría de referencia), el coeficiente de regresión corresponde a los estudiantes que *Casi Siempre* o *Siempre* se sienten respetados y orientados por sus profesores es 0,789 y el OR es de 2,201, lo que significa que las probabilidades de un mejor rendimiento aumentan en un factor de 2,201 para los estudiantes que se sienten respetados y orientados por sus profesores, en comparación con quienes sienten que *Nunca* u *Ocasionalmente* son respetados y orientados por sus profesores.

En lo referente a la pregunta, *¿Cómo complementaría su formación universitaria?*, representada en la variable *ComplementariaEstudios* que se ha categorizado en dos grupos, (un grupo conformado por los estudiantes que prefieren la autoformación o cursos de actualización, siendo ésta la categoría de referencia, con $\beta=0$ y el otro por estudiantes que desean hacer investigación, otra carrera, o estudiar un postgrado, a los que ha etiquetado como “Si”), se obtuvo un coeficiente de regresión de 0,549 y el OR 1,731, lo que indica que las probabilidades de un mejor rendimiento aumentan en un factor de 1,731 para los estudiantes que desean seguir complementando su formación profesional de una manera formal mediante investigación, estudiar otra carrera o un postgrado, en comparación con quienes desean hacer cursos de actualización o la autoformación.

Respecto a la pregunta *¿Cómo se siente usted en su relación con respecto a su familia?*, representada por la variable *SentimientoFamilia*, que se ha categorizado en dos grupos, (un grupo conformado por los estudiantes que sienten que su relación con su familia está *Mal* o *Regular*, que se ha denominado “*No Bien*” y es la categoría de referencia, y el otro por estudiantes que sienten que su relación con su familia está *Bien* o *Muy Bien*, a los que ha etiquetado como “*Bien*”), se obtuvo un coeficiente de regresión $-0,935$ y el OR es 0,392; indicándonos que las probabilidades de un mejor rendimiento aumentan en un factor de 2,551 ($1/0,392$) para los estudiantes que se sienten *Bien* o *Muy Bien* con su familia, en comparación con quienes en su relación familiar se sienten *Mal* o *Regular*.

Finalmente, en la tabla 17 se muestran los resultados de clasificación del modelo ajustado obteniéndose un porcentaje de correcta clasificación del 52,7%. Pese a que

puede dar impresión de una baja tasa de correcta clasificación (TCC), es de recordar que la finalidad del presente estudio es explicativa y no predictiva, de manera que pudiésemos conocer aquellos factores demográficos, sociales, afectivos, económicos, académicos, recreativos y ambientales que más influyen en el rendimiento académico de los estudiantes de la Universidad del Magdalena. A pesar de ello, comentar que el valor de la tasa de clasificación correcta, TCC puede deberse a que la clasificación es muy sensible al tamaño de los grupos de los componentes, favoreciendo siempre la clasificación de los grupos grandes (Hosmer et al., 2013).

Tabla 17. Tabla de Clasificación o Matriz de Confusión

Fuente: Elaboración	Rendimiento Predicho			Suma	Propia	
	Básico	Alto	Superior			
A este respecto,	Básico	136	21	27	184	Sainani (2014), realizar estudios multivariantes, el
manifiesta que al	Alto	66	18	51	135	
con modelos	Superior	36	14	86	136	
	Tasa de correcta clasificación				0,527	

investigador debe tener bien claro si sus objetivos son explicativos o predictivos, pues la investigación explicativa tiene como objetivo determinar los factores relacionados causalmente con un resultado, mientras que la predictiva busca encontrar los factores que mejor predigan un diagnóstico actual o un evento futuro, igualmente considera que la elección de uno u otro enfoque afecta a los aspectos de construcción y evaluación del modelo.

6 DISCUSIÓN

Teniendo en cuenta que el efecto de las variables explicativas sobre la variable respuesta, en este caso, el Rendimiento Académico, es medido a través del valor del correspondiente Odds Ratio para cada una de las variables que resultaron significativas en el estudio, las podemos ordenar según su importancia como se muestra a continuación, donde se recuerda la formulación original de las preguntas en la encuesta:

- 1) *¿En qué facultad estudias?*
- 2) *¿Cómo se siente usted en su relación con respecto a su familia?*
- 3) *¿Ha interrumpido sus estudios universitarios?*
- 4) *¿Se siente respetado y orientado por sus profesores?*
- 5) *¿Ha realizado otros estudios?*
- 6) *¿Tiene personas a su cargo?*

- 7) *¿Cómo financia los costos educativos?*
- 8) *¿Cómo complementaría su formación universitaria?*
- 9) *¿Con quién vives?*

Respecto a la primera pregunta, encontramos que los estudiantes matriculados en programas académicos de las facultades de Ciencias de la Educación, Ciencias de la Salud, Ciencias Empresariales y Económicas, y Humanidades (categoría 2), tienen mejor desempeño que los estudiantes matriculados en programas académicos de las facultades de Ingeniería y Ciencias Básicas. Esto se podría explicar debido a la presencia de un contenido curricular más exigente en matemáticas en los programas de estas dos últimas facultades. Se tiene conocimiento de las dificultades en los procesos de enseñanza de las matemáticas en el sistema educativo colombiano. En tal sentido Murcia y Heno (2015) manifiestan que estas dificultades en las enseñanzas de las matemáticas quedan directamente evidenciadas en los procesos de aprendizaje de los niños y jóvenes cuya formación está enmarcada en estos niveles y que sin lugar a dudas perdura o se extiende hasta la educación terciaria.

En cuanto a la relación que tiene el estudiante con su familia, se encontró que los estudiantes que no tienen problemas en sus relaciones familiares tienen un mejor desempeño académico. Esto concuerda con lo encontrado por Medranda y Romero (2015), quienes concluyen que los problemas familiares presentes en los estudiantes es un factor que en un porcentaje alto provoca el bajo rendimiento académico, y que los alumnos tengan que repetir o desertar del sistema educativo. A conclusiones similares llega la investigación realizada por Torres y Rodríguez (2006).

Respecto al hecho de si un estudiante *ha interrumpido o no sus estudios universitarios* se encontró que aquellos estudiantes que no han sufrido algún tipo de interrupción en sus estudios presentan un mejor desempeño académico que aquellos que han tenido que interrumpir sus estudios. Se procedió a analizar si dicha interrupción es debida precisamente a dificultades académicas o debida a otras causas. Observando los posibles motivos por los cuales los estudiantes interrumpen sus estudios encontramos que solo 14 estudiantes dentro de los 455 estudiantes de la muestra han interrumpido sus estudios debido a problemas académicos (3,1% de la muestra), mientras que la causa principal por la que los estudiantes interrumpen sus estudios es debido a dificultades económicas (casi 21% de los estudiantes de la muestra). Respecto a esto último, se encontró en el

estudio realizado por de Miguel (2001), que una de las causas del bajo rendimiento académico son los factores familiares relativos al nivel económico. Panera, Davalillo y Panera (2010), afirman que las causas del abandono son, en todos los países latinoamericanos, mayoritariamente económicas en un 38%, pasando a un 31% de dificultades de rendimiento académico.

Respecto a la variable que hace referencia a si los estudiantes *Se Siente Respetado y Orientado por sus Profesores*, encontramos que los estudiantes que sienten que sus profesores le respetan y orientan tienen mejor desempeño que aquellos que no lo consideran de este modo. Al respecto Álvaro et al., (1990) manifiesta que:

“la función del profesor, el papel que desempeña en un sistema educativo, influye en gran medida en el rendimiento que obtienen los alumnos. Su manera de comunicarse, las relaciones que establece con el alumno y las actitudes que adopta hacia el mismo juegan un papel determinante tanto en su comportamiento como en su aprendizaje”. (p.91)

Según Marín (1969), citado por Álvaro et al., (1990), "la suerte de un sistema educativo se juega sobre todo en la carta de sus profesores. Planes, programas, organización escolar, métodos y material, cobran todas sus virtualidades en manos de un buen equipo docente". Por su parte Garbanzo (2013) encontró un efecto significativo de la satisfacción de los estudiantes con sus profesores y el Rendimiento Académico.

Garbanzo (2007), expone que las expectativas que el estudiante tiene sobre las relaciones con sus profesores y con sus compañeros de clase son factores importantes que intervienen en los resultados académicos. Al respecto, Castejón y Pérez (1998), citado en Garbanzo (2007) hacen referencia a que el estudiante desea encontrar en el profesor tanto una relación afectiva, como didáctica y que ello tiene repercusiones en el Rendimiento Académico.

En este mismo sentido Abarca, Gómez, y Covarrubias (2015) afirman que la comunicación entre profesor–alumno es clave en el logro del éxito académico, una buena relación genera un ambiente agradable y de confianza para trabajar colaborativamente y lograr los objetivos.

En lo que respecta a la pregunta si el estudiante *Ha Realizado Otros Estudios*, se encontró que los estudiantes que previamente han realizado otros estudios universitarios

tienen un mejor rendimiento académico que quienes no lo han realizado. Esto lo podríamos explicar desde la perspectiva de la madurez del estudiante, es decir, estudiantes que previamente han iniciado el estudio de otras carreras universitarias, y que quizás ahora están estudiando la carrera que realmente les gusta y desean, o quizás estudiantes que están en una nueva etapa universitaria luego de haberse retirado de sus estudios previos y que están ante una nueva oportunidad de llevar a cabo y culminar su formación profesional.

Respecto al hecho de si el estudiante *Tiene personas a su cargo*, la investigación arrojó que los estudiantes que no tienen personas que dependan económicamente de él muestran mejor desempeño académico que quienes si lo tienen. Esto se podría ser explicado aludiendo al hecho de estos últimos poseen responsabilidades en el sostenimiento de una o más personas, además de sí mismo, deben preocuparse por la consecución de los recursos financieros que le permitan cumplir con tal propósito, además de que posiblemente deban dedicar tiempo a velar por la salud, seguridad, estudios, entre otros aspectos de las personas a su cargo. Ello implicaría que la dedicación a sus estudios sea menor. Serrano (2013) manifiesta en su estudio, que los factores más frecuentes de riesgo asociados a la vulnerabilidad social, la cual influye en el rendimiento académico, son los relacionados con las cargas familiares, bien por tener hijos u otros familiares a su cargo.

Respecto a la manera de como el estudiante financia sus estudios universitarios representados en la pregunta, *¿Cómo costea sus gastos educativo?*, se encontró que los estudiantes que reciben algún tipo de ayuda para cubrir los costos de sus estudios universitarios muestran mejor desempeño que quienes lo hacen con sus propios recursos. Esto se podría explicar aludiendo al hecho de estos últimos tienen la obligatoriedad de buscar los recursos económicos que le permitan realizar sus estudios, lo cual implica que la dedicación a sus estudios sea menor. Similar a lo explicado en el caso de tener personas a su cargo.

En lo que respecta a la pregunta de *¿Cómo complementarías su formación universitaria?*, se encontró que aquellos estudiantes que desean continuar su formación profesional realizando otra carrera, un postgrado o dedicarse a la investigación tienen mejores resultados en su desempeño académico que quienes desean realizar cursos de actualización o algún tipo de autoformación. Esto denota que los estudiantes que de una

u otra forma se interesan por continuar estudiando de manera formal es porque se sienten a gusto en sus estudios, lo que quizás se deba a lo que Garbanzo (2007) llama *motivación académica intrínseca del estudiante*, al afirmar que está ampliamente demostrado que la orientación motivacional del estudiante juega un papel significativo en el desempeño académico. Algunos autores como Salanova, Martínez, Bresó, Llorens Gumbau S., Gumbau Grau R. (2005), se refieren a este campo como el *engagement*, definido como “un estado psicológico relacionado con los estudios que es positivo y significativo” (p.173). El *engagement* es caracterizado por vigor, dedicación y absorción.

Según Schaufeli y Bakker (2010), el vigor se refiere a altos niveles de energía y esfuerzo mental con la que el estudiante desarrolla sus actividades académicas, es por ello que las personas se esfuerzan en sus labores incluso cuando existen dificultades, la dedicación es la alta implicación en sus estudios junto con un sentimiento de reto y orgullo por el trabajo y la absorción implica que el tiempo pasa rápido y que el estudiante está totalmente entregado al estudio, vinculando de esta manera al estudiante con sus labores académicas.

Garbanzo (2007), señala que estas condiciones se aprecian en aquellos estudiantes para quienes el estudio significa más que una tarea, un disfrute por el saber, pueden pasar largas horas desarrollando actividades académicas con una gran disposición hacia lo que hacen, debido a que tienen la sensación de que el tiempo pasa demasiado rápido y muestran una alta capacidad de compromiso y concentración académica. Por lo general, estas personas manifiestan sentir felicidad al realizar las tareas académicas, pues el estudio es un disfrute.

En este mismo orden de ideas Pérez, Ramón y Sánchez (2000), citado en Garbanzo (2007), señalan que en la *autoeficacia percibida* se dan casos de estudiantes que por distintas razones carecen de autoeficacia. Esta condición se presenta cuando hay ausencia de un estado de motivación intrínseca que permita al estudiante cumplir con un desempeño académico aceptable. Se asocia con estados de agotamiento, desinterés y falta de proyección con sus estudios, y es conocido como *burnout*, que es la fatiga o la sensación de estar “quemado” por las actividades académicas. La motivación y el compromiso de los estudiantes con el logro académico son fundamentales en sus resultados.

En relación a la pregunta *con quien está viviendo el estudiante* se encontró, paradójicamente, que los estudiantes *Que No Viven con su Familia de Origen* tienen mejor desempeño que los estudiantes *Que Viven con su Familia de Origen*. En la muestra hay 271 (casi 60%) estudiantes *Que Viven con su Familia de Origen* y 184 *Que No Viven con su Familia de Origen*, de estos últimos, 130 estudiantes (casi 71%) son procedentes de otras ciudades del país, es decir, estudiantes que han debido trasladarse de lugar de residencia para realizar sus estudios, mientras que 54 estudiantes son procedentes de la ciudad de Santa Marta, ciudad sede de la Universidad del Magdalena, y por ende, no trasladan su lugar de residencia para llevar a cabo sus estudios universitarios.

De esta manera, se observa que la mayoría de los estudiantes *Que No Viven con su Familia de Origen* son estudiantes que se han trasladado de su municipio de origen hacia la ciudad de Santa Marta para poder realizar sus estudios. Respecto al Rendimiento Académico de los estudiantes universitarios que deben cambiar su lugar de residencia, Di Gresia, Porta y Ripani (2002) afirman que:

“Si el alumno cambia de residencia (se muda) para poder asistir a la universidad, ese cambio tiene un efecto positivo sobre el rendimiento, sugiriendo que el efecto del mayor tiempo libre y la más rápida integración al ambiente universitario supera al efecto de desarraigo y de separación de la familia.” (p.22)

Sin embargo, en el caso de la investigación desarrollada por Sarmiento y Zamora (2015) se concluye que no existe relación estadísticamente significativa entre el Rendimiento Académico del estudiante que se ha trasladado de residencia, contrastado con el educando que no se ha trasladado de residencia para realizar sus estudios universitarios.

Durante la etapa universitaria hay jóvenes adultos que se separan de sus padres, Fasick (1984), citado por Rice (1997), considera que esta separación es un paso muy importante en el proceso de convertirse en adulto. “Los adolescentes buscan en sus padres compañía, satisfacción emocional y orientación, este paso les ayuda a romper los cercanos vínculos con los padres. Establecer una residencia separada, sea en un dormitorio escolar o en un apartamento en la comunidad, ayuda también a la autonomía”. (Rice, 1997, p.471). Aún más que la separación física, afirman Thornton, Young- De Marco y Goldscheider (1993) es lograr la autonomía emocional, la tarea

consistente en romper los vínculos de dependencia que han sido formados y reorientados sobre la base de la equidad, de adulto a adulto.

7 CONCLUSIÓN

Tal y como se mencionó al inicio del presente trabajo, el objetivo fundamental del mismo es determinar los factores demográficos, familiares, afectivos, académicos, personales, socio-económicos y ambientales que influyen en el rendimiento académico de los estudiantes, debido a que el rendimiento académico de los estudiantes universitarios es un indicador clave para determinar si una institución está alcanzando sus objetivos educativos, según lo establece la Commission on Higher Education, 1995.

Un gran número de los factores que influyen en el Rendimiento Académico del estudiante universitario están fuera del alcance de los administradores responsables de los programas académicos y de bienestar universitario. Sin embargo, existen algunos factores que pueden ser modificados desde el interior de la Universidad con el fin de brindarles a los estudiantes las condiciones más adecuadas para la mejora del desempeño estudiantil y que han sido identificados en el presente trabajo.

Entre los factores que tienen cierta relación con el Rendimiento Académico y que pueden ser intervenidos desde el interior de la Universidad del Magdalena, se han identificados los siguientes:

La *Facultad* en la que estudia el estudiante, encontrándose que quienes estudian una carrera vinculada a la Facultades de Ingeniería y Ciencias Básicas tienen peor desempeño académico. Respecto a este aspecto la Universidad podría tratar de identificar las verdaderas causas del por qué los estudiantes de estas facultades presentan un menor desempeño y a partir de ahí formular programas que busquen mejorar las dificultades presentadas. En este trabajo se piensa que el peor rendimiento académico de las carreras técnicas puede ser debido al mayor número de asignaturas relacionadas con las matemáticas que presentan dichos programas. De ser así, la medida de mejora que se recomienda a la Universidad es que plantee cursos de reforzos en esta área desde los primeros semestres académicos, y reforzar los programas de monitorias académicas en esta área o su total reformulación.

En relación a la *Interrupción de los Estudios Universitarios*, se encontró que los problemas de carácter financiero (casi 21% de los estudiantes de la muestra) es la

principal causa. Al respecto, nuestra recomendación para la Universidad podría ser aumentar las plazas disponibles en el *Programa de Almuerzos y Refrigerios* (este consiste en brindar, según algún criterio, que consideramos debe ser la necesidad del estudiante, *Almuerzos* o *Refrigerios* en días específicos a estudiantes de bajos recursos), o reestructurar la forma de asignación de dichas plazas. Puede observarse de la última asignación de éstos (semestre 2018 – 2), que algunos estudiantes se le brinda *Refrigerio* solo un día de la semana, mientras a otros *Almuerzo* completo los cinco días de la semana cubriendo un total de 4.601 estudiantes con este beneficio, un número bastante amplio, aunque deben considerarse que las necesidades alimenticias son todos los días y quizás esta estrategia de cubrir con una gran cantidad de estudiantes no está cumpliendo debidamente las necesidades de los más necesitados de manera integral. Es obvio que es imposible ofrecerles esta ayuda a todos los estudiantes de la comunidad universitaria, pero sí podría mejorarse la asignación para ofrecérselos a los que más lo necesitan.

Puede consultarse el listado de estudiantes beneficiados con la asignación de Almuerzos y Refrigerios en la siguiente dirección web.

https://www.unimagdalena.edu.co/Content/ArchivosPublicaciones/Publicacion_8377/documento_2_20180926105430.048.pdf

En relación con el hecho de si el estudiante se *Siente Respetado y Orientado por sus Profesores*, este es un factor en el cual la Universidad del Magdalena puede intervenir de manera proactiva mediante la concienciación del cuerpo docente y con la implementación de cursos de formación cuyo objetivo sea mejorar la comunicación del profesor con sus estudiantes, propiciando un verdadero respeto y orientación.

En cuanto a los demás factores que se han identificado que pueden incidir en los resultados del Rendimiento Académico de los estudiantes de la Universidad del Magdalena, es decir, *la relación con su familia, el tener personas a cargo, el haber realizado otros estudios universitarios, cubrimiento de los costos educativos, el deseo del estudiante por continuar su formación profesional por medio de postgrados, y el vivir con personas diferentes a su familia de origen*, la Dirección de Bienestar Universitario de la Universidad del Magdalena podría implementar programas para identificar la problemática particular de los estudiantes y su posterior tratamiento mediante la orientación y acompañamiento por profesionales. Dichos programas deben cubrir aspectos como el manejo de conflictos familiares, la gestión del tiempo,

organización y coordinación adecuada entre las actividades académicas con las responsabilidades laborales y/o familiares, hábitos de estudios, concienciación de la importancia de los niveles superiores de educación como maestrías y doctorados, entre otros. Estos programas deben estar apoyados por los profesores, ya que son éstos quienes están en contacto directo con los estudiantes y la vinculación de los estudiantes a este programa debe ser de carácter obligatorio para aquellos que presenten alta tasa de repeticiones de materias y/o bajo Rendimiento Académico.

Podemos concluir, comentado que, los resultados obtenidos en la presente investigación, constituyen el primer eslabón para guiar el proceso de identificación de los factores que están afectando el Rendimiento Académico de los estudiantes de la Universidad del Magdalena, contribuyendo a la mejora de los procesos y programas administrados por la Dirección de Bienestar Universitario, en busca de disminuir el bajo rendimiento académico y deserción en los estudiantes de la Universidad.

8 LINEAS FUTURAS DE INVESTIGACIÓN

La presente investigación constituye la primera fase de un estudio más ambicioso en el análisis de las relaciones que puedan existir entre los factores antes mencionados y factores de otra índole (institucionales, pedagógicos y psicosociales) para estudiar su posible relación causal sobre el Rendimiento Académico del estudiante de la Universidad del Magdalena mediante modelos más complejos de ecuaciones estructurales, ya sea mediante modelos de covarianza (SEM) o modelos basados en la varianza (SEM-PLS), de manera que se puedan analizar las relaciones causales entre, por un lado, un conjunto de variables observables y por otro, variables tanto observables como no observables, desde una perspectiva confirmatoria.

Otra posible futura línea de investigación sería modelar el Rendimiento Académico mediante un modelo de regresión de mínimos cuadrados parciales o Partial Least Squares Regression (PLS Regression), la cual se utilizaría para hallar las relaciones fundamentales entre las variables explicativas y la(s) variable(s) respuesta(s), mediante un enfoque de variables latentes que permite modelar la estructura de covarianza en estos dos espacios, de tal modo que podría abordarse la inclusión de alguna(s) otra(s) posible(s) variable respuesta excluida en estos momentos por su alta correlación con alguna otra (por ejemplo, en nuestro caso “si tenía hijos” y “si tenía personas a su cargo”), en el análisis. La regresión de mínimos cuadrados parciales es especialmente adecuada cuando la matriz de predictores tiene más variables que observaciones, o cuando hay multicolinealidad entre las variables explicativas.

Por supuesto, hubo limitaciones en este estudio. Como se indicó anteriormente, se utilizó una muestra de estudiantes con un cuestionario específico, no realizándose ningún tipo de diseño experimental, con lo que los resultados obtenidos deberán ser interpretados con cautela y contrastados en futuras investigaciones con otras técnicas estadística multivariantes más potentes, como las mencionadas anteriormente.

9 BIBLIOGRAFÍA

- Abarca, M., Gómez, M., y Covarrubias, M. (2015). Análisis de los factores que contribuyen al éxito académico en estudiantes universitarios: estudio de cuatro casos de la Universidad de Colima. *Revista Internacional de Educación y Aprendizaje*, 3(2), 125-136.
- Álvaro, M., Bueno, M. J., Calleja, J. A., Cerdán, J., Echeverría M. J., García C., Gaviria J. L., Gómez C., Jiménez, S. C., López B., Martín-Javato L., Mínguez A. L., Sánchez A., y Trillo, C. (1990). *Hacia un modelo causal del rendimiento académico*. Madrid, España. Centro de Investigación, Documentación y Evaluación – C.I.D.E.
- Camarero Rioja, L., Llorente, A. A., y Ramírez, B. M. (2015). *Regresión Logística: Fundamentos y aplicación a la investigación sociológica*. Departamento de Sociología I, UNED.
- Castejón, C., Pérez, S. (1998). Un modelo causal-explicativo sobre la influencia de las variables psicosociales en el rendimiento académico. *Bordón. Revista de Pedagogía*. Sociedad Española de Pedagogía. 2(50), 170-184.
- Chakrabarti, A., & Ghosh, J. K. (2011). AIC, BIC, and recent advances in model selection. *Handbook of the philosophy of science*, 7, 583-605.
- Commision On Higher Education MSA: Framework for Outcomes Assessment. Middel State Association, 1995
- Cortez Bailón, F. M., Tutiven Campos, J. L., y Villavicencio Morejon, M. N. (2017). Determinantes del rendimiento académico universitario. *Revista Publicando*, 4(10), 284-296.
- Cruz, Z., Medina, J., Vázquez, J., Espinosa, E., y Antonio A. (2014). Influencia del nivel socioeconómico en el rendimiento académico de los alumnos del programa educativo de ingeniería industrial en la Universidad Politécnica de Altamira. *Ciencias Administrativas y Sociales Handbook TV*, 5, 24-38.

- de Miguel Díaz M, Apodaca Urquijo P, Arias Blanco J. M., Escudero Escorza T., Rodríguez Espinar S., y Vidal García Javier (2002). Evaluación del Rendimiento Académico en la Enseñanza Superior. Comparación de resultados entre alumnos procedentes de la LOGSE y del COU. *Revista de Investigación Educativa*, 2(20), 357-383
- de Miguel, C. R. (2001). Factores familiares vinculados al bajo rendimiento. *Revista complutense de educación*, 12(1), 81 – 113.
- Di Gresia, L. M., Porto, A., y Ripani, L. (2002). Rendimiento de los estudiantes de las universidades públicas argentinas. Documentos de Trabajo. Recuperado de <http://www.depeco.econo.unlp.edu.ar/wp/wp-content/uploads/2017/05/doc45.pdf>
- Dziak, J. J., Coffman, D. L., Lanza, S. T., y Li, R. (2012). Sensitivity and specificity of information criteria (Methodology Center Technical Report 12-119). University Park, PA: The Methodology Center, The Pennsylvania State University. Recuperado de <http://methodology.psu.edu/media/techreports/12-119.pdf>
- Ebenuwa-Okoh E. E. (2010): Influence of Age, Financial Status, and Gender on Academic. *Journal of Psychology*, 1 (2): 99-103
- Fajardo Bullón, F., Maestre Campos, M., Felipe Castaño, E., León del Barco, B. y Polo del Río, M.I. (2017). Análisis del rendimiento académico de los alumnos de Educación Secundaria Obligatoria según las variables familiares. *Educación XXI*, 20(1), 209-232, doi: 10.5944/educXXI.14475
- Faraway, J. J. (2016). *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. Second Edition. CRC press.
- Fasick, F. A. (1984). Parents, peers, youth culture and autonomy in adolescence. *Adolescence*, 19(73), 143.
- Fernández Aráuz, A. (2015). Aplicación del análisis factorial confirmatorio a un modelo de medición del rendimiento académico en lectura. *Ciencias Económicas*, 33(2). 39-66.

- Garbanzo Vargas, G. M. (2013). Factores asociados al rendimiento académico en estudiantes universitarios desde el nivel socioeconómico: Un estudio en la Universidad de Costa Rica. *Revista Electrónica Educare*, 17(3).
- Garbanzo Vargas, G. M. (2007). Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde la calidad de la educación superior pública. *Educación*, 31(1). 43-63
- García Aretio, L. (1989). Factores que inciden en el rendimiento académico de los alumnos de la Universidad Nacional de Educación a Distancia (UNED) española (elaboración de un índice). *Revista de Tecnología Educativa*. 11(1). 69-95
- García Jiménez, M. V., Alvarado Izquierdo, J. M., y Jiménez Blanco, A. (2000). La predicción del rendimiento académico: regresión lineal versus regresión logística. *Psicothema*, 12(Suplemento), 248-252.
- Gómez Sánchez, D., Oviedo Marín, R., y Martínez López, E. (2011). Factores que influyen en el rendimiento académico del estudiante universitario. *TECNOCIENCIA Chihuahua*, 5 (2). 90 – 97.
- Greenwell, B. M., McCarthy, A. J., Boehmke, B. C., & Liu, D. (2018). Residuals and Diagnostics for Binary and Ordinal Regression Models: An Introduction to the *sure* Package. *The R Journal*. The R Journal. 10(1). 381 - 394
- Hair, J. F., Anderson, R. E., Tatham, R. L., y Black, W. C. (1999). *Análisis multivariante*. Madrid: Prentice Hall. 5ª ed
- Heredia, J., Rodríguez, A., y Vilalta, J. A. (2012). Empleo de la regresión logística ordinal para la predicción del rendimiento académico. *Investigación Operacional*, 33(3), 252-267.
- Hosmer, D. W., Lemeshow, S., y Sturdivant, R. X. (2013). *Applied logistic regression*. Recuperado de <https://ebookcentral.proquest.com>
- Ibarra, M., y Michalus, J. (2010). Análisis del rendimiento académico mediante un modelo Logit. *Revista Ingeniería Industrial*, 9(2).
- Instituto Internacional de la UNESCO para la Educación Superior en América Latina y el Caribe (IESALC) (2008). *Declaración de la Conferencia Regional de*

Educación Superior en América Latina y Caribe. Políticas Educativas, 6(1). Recuperado de https://www.mineducacion.gov.co/cvn/1665/articles-162147_archivo_pdf.pdf

Kim, J. E., & Nembhard, D. A. (2017). Modeling academic performance and procrastination: A structural equation approach. In IIE Annual Conference. Proceedings (pp. 537-542). Institute of Industrial and Systems Engineers (IISE).

Ley 715 de 2001. Artículo 20. República de Colombia. Recuperado de https://www.mineducacion.gov.co/1621/articles-86098_archivo_pdf.pdf

Ley 142 de 1994. Artículo 102. República de Colombia. Recuperado de http://www.sic.gov.co/sites/default/files/normatividad/Ley_142_1994.pdf

Marín Ibáñez, R. (1969). La formación y selección del profesorado universitario. Revista de educación, (203), 14-22. Recuperado de <https://www.mecd.gob.es/dctm/revista-de-educacion/1969-203/1969re203estudios02.pdf?documentId=0901e72b818397ee>

Medranda R. F., y Romero S. (2018). Los problemas familiares y el rendimiento académico de los y las estudiantes de 3er. nivel de la carrera de trabajo social de la facultad de ciencias humanísticas y sociales. Revista Caribeña de Ciencias Sociales. En línea: <https://www.eumed.net/rev/caribe/2018/05/estudiantes-trabajo-social.html//hdl.handle.net/20.500.11763/caribe1805estudiantes-trabajo-social>

Montero Rojas, Eiliana., Villalobos Palma, Jeannette y Valverde Bermúdez, Astrid. (2007). “Factores institucionales, pedagógicos, psicosociales y sociodemográficos asociados al rendimiento académico en la Universidad de Costa Rica: Un análisis multinivel”. *Relieve*. 13(2). 215-234. Recuperado de <http://www.redalyc.org/articulo.oa?id=91613205>

Mora García, R. T. (2015). Factores que intervienen en el rendimiento académico universitario: Un estudio de caso. *Opción*, 31(6). 1041-1063

Murcia, M. E. y Henao, J. C. (2015). Educación matemática en Colombia, una perspectiva evolucionaria. *Entre Ciencia e Ingeniería*, 9(18). 23 – 30

- Panera, F., Davalillo A., y Panera A. (2010). Evaluación de la educación. Cómo mejorar el rendimiento académico en la enseñanza superior: una perspectiva de gestión por procesos. Congreso Iberoamericano de Educación – Metas 2021. Buenos Aires, República Argentina, 13 – 15 septiembre de 2010. Recuperado de http://www.adeepa.org.ar/congresos/Congreso%20IBEROAMERICANO/EVALUACION/RLE2537_Panera.pdf
- Pérez, A., Ramón, J., y Sánchez, J. (2000). Análisis exploratorio de las variables que condicionan el rendimiento académico. Sevilla, España: Universidad Pablo de Olavide, 34, 15-20.
- Real Decreto 1125/2003, de 5 de septiembre, por el que se establece el sistema europeo de créditos y el sistema de calificaciones en las titulaciones universitarias de carácter oficial y validez en todo el territorio nacional. Boletín Oficial del Estado. Madrid, 18 de septiembre de 2003. núm. Pp. 224. 1-4. Recuperado de <https://www.boe.es/buscar/pdf/2003/BOE-A-2003-17643-consolidado.pdf>
- Redondo, L., Corrás, T., Novo, M., y Fariña, F. (2017). El rendimiento académico: La influencia de las expectativas, el optimismo y la autoeficacia. Revista de Estudios e Investigación en Psicología y Educación, (10), 104-108.
- Rice, F. P. (1997). Desarrollo humano: estudio del ciclo vital. Pearson Educación.
- Rodríguez, Gustavo (2014). Determinantes del desempeño académico universitario. El caso de la Región Caribe colombiana. Estudios sobre calidad de la educación en Colombia. Working paper. ICFES 2014. Disponible en <http://www.icfes.gov.co/docman/investigadores-y-estudiantes-de-posgrado/resultados-de-investigaciones/equidad/989-determinantes-del-desempeno-academico-universitario-el-caso-region-caribe-colombiana>
- Rodríguez Ayán, M. N., y Ruiz Díaz, M. Á. (2011). Indicadores de rendimiento de estudiantes universitarios: calificaciones versus créditos acumulados. Revista de Educación, 355. Mayo-agosto 2011, 467-492.
- Sainani, K. L. (2014). Explanatory versus predictive modeling. PM&R, 6(9), 841-844.
- Salanova Soria, M., Martínez Martínez, I. M., Bresó Esteve, E., Llorens Gumbau, S., y Grau Gumbau, R. (2005). Bienestar psicológico en estudiantes universitarios:

facilitadores y obstaculizadores del desempeño académico. *Anales de psicología*, 21(1).

Schaufeli, W. B., y Bakker, A. B. (2010). Defining and measuring work engagement: Bringing clarity to the concept. *Work engagement: A handbook of essential theory and research*, 10-24.

Sarmiento, M. J., y Zamora, J. (2015). Relación entre el criterio de traslado de residencia y el rendimiento académico en los estudiantes de la carrera de Fonoaudiología de la Universidad de La Frontera en el primer año. *Revista de Educación en Ciencias de la Salud*. 12(2). 157-161. Publicación oficial de ASOFAMECH y SOEDUCSA CONCEPCIÓN-CHILE. Recuperado de <http://www2.udec.cl/ofem/recs/antiores/vol1222015/artinv12215h.pdf>

Serrano Ruíz, L. (2013). La influencia de los factores socioeconómicos en el rendimiento académico. Trabajo Fin de Máster. Universidad de Almería. Almería. España. Recuperado de <http://repositorio.ual.es/bitstream/handle/10835/2335/Trabajo.pdf?sequence=>

Shahzadi, E., y Ahmad, Z. (2011). A study on academic performance of university students. *Proc. 8th International Conference on Recent Advances in Statistics*, 255, 268

Shalizi, Cosma (2015). Lecture 26: Variable Selection. 36-401, Fall 2015, Section B. Recuperado de <https://www.stat.cmu.edu/~cshalizi/mreg/15/lectures/26/lecture-26.pdf>

Silva, L. C., y Barroso, I. M. (2004). *Regresión Logística*. La Muralla. Salamanca.

Smith, T., y McKenna C. (2013). A Comparison of Logistic Regression Pseudo R² Indices. *Multiple Linear Regression Viewpoints*, 39(2). 17 - 26

Thornton, A., Young-DeMarco, L., y Goldscheider, F. (1993). Leaving the parental nest: The experience of a young white cohort in the 1980s. *Journal of Marriage and the Family*, 216-229.

Tomás-Miquel, J. V., Expósito-Langa, M., y Sempere-Castelló, S. (2014). Determinantes del rendimiento académico en los estudiantes de grado. Un estudio

en administración y dirección de empresas. *Revista de investigación educativa*, 32(2), 379-392.

Torres, L.E. y Rodríguez, N. Y. (2006). Rendimiento académico y contexto familiar en estudiantes universitarios. *Enseñanza e investigación en psicología*. 11(2). 255-270.

Tourón Figueroa, Javier. (1985). La predicción del rendimiento académico: Procedimientos, resultados e implicaciones. *Revista Española de Pedagogía*. 43(169-170). 473-495. Disponible en <http://www.jstor.org/stable/23764399>

Universidad del Magdalena. Reglamento Estudiantil. Recuperado de https://www.unimagdalena.edu.co/Content/Public/Docs/Entrada_Direccion8/adjunto_1023-20180509151728_256.pdf

Verdugo Saldivar, M. E., y Urias Montes, V. P. (2007). Antecedentes a nivel nacional, internacional y proceso de acreditación de una institución educativa Mexicana. El papel de la universidad en la transformación de la sociedad. Puebla: Sexto Congreso Internacional "Retos y expectativas de la Universidad".

Unesco. (1998). Conferencia mundial sobre la educación superior: La educación superior en el siglo XXI: Visión y Acción. Recuperado de <http://unesdoc.unesco.org/images/0011/001163/116345s.pdf>

Upton, G. J. (2017). *Categorical data analysis by example*. John Wiley & Sons.

10 ANEXOS

Tabla 18. Estudiantes entrevistados según facultad

Facultad	Frecuencia	Porcentaje	Porcentaje acumulado
Facultad de Ciencias Básicas	7	1,5	1,5
Facultad de Ciencias de la Educación	75	16,5	18,0
Facultad de Ciencias de la Salud	60	13,2	31,2
Facultad de Ciencias Empresariales y Económicas	73	16,0	47,3
Facultad de Humanidades	43	9,5	56,7
Facultad de Ingeniería	197	43,3	100,0
Total	455	100,0	

Fuente: Elaboración Propia

Tabla 19. Estudiantes entrevistados según programa académico

Programa Académico	Frecuencia	Porcentaje	Porcentaje acumulado
	a	e	
Administración de empresas	20	4,4	4,4
Antropología	11	2,4	6,8
Biología	7	1,5	8,4
Cine y audiovisuales	7	1,5	9,9
Contaduría pública	19	4,2	14,1
Derecho	25	5,5	19,6
Economía	4	,9	20,4
Enfermería	12	2,6	23,1
Ingeniería agronómica	8	1,8	24,8
Ingeniería ambiental y sanitaria	28	6,2	31,0
Ingeniería civil	25	5,5	36,5
Ingeniería de sistemas	33	7,3	43,7
Ingeniería electrónica	11	2,4	46,2
Ingeniería industrial	86	18,9	65,1
Ingeniería pesquera	6	1,3	66,4
Licenciatura en educación básica con énfasis en informática	40	8,8	75,2
Licenciatura en preescolar	35	7,7	82,9
Medicina	18	4,0	86,8
Negocios internacionales	23	5,1	91,9
Odontología	10	2,2	94,1
Psicología	20	4,4	98,5
Tecnología en administración hotelera y turística	7	1,5	100
Total	455	100	

Fuente: Elaboración Propia

Tabla 20. Estudiantes entrevistados según estrato socio-económico

Estrato Socioeconómico			
Estrato	Frecuencia	Porcentaje	Porcentaje acumulado
1	191	42,0	42,0
2	170	37,4	79,3
3	80	17,6	96,9
4	12	2,6	99,6
5	1	,2	99,8
6	1	,2	100,0
Total	455	100,0	

Fuente: Elaboración Propia

Tabla 21. Estudiantes entrevistados según tipo de colegio (Público/Privado)

Tipo colegio	Frecuencia	Porcentaje	Porcentaje acumulado
Privado	115	25,3	25,3
Público	340	74,7	100,0
Total	455	100,0	

Fuente: Elaboración Propia

Tabla 22. Estudiantes entrevistados según provengan de un municipio certificado en educación

Municipio Certificado	Frecuencia	Porcentaje	Porcentaje acumulado
No	108	23,7	23,7
Si	347	76,3	100,0
Total	455	100,0	

Fuente: Elaboración Propia

Tabla 23. Estudiantes entrevistados según género

Sexo	Frecuencia	Porcentaje	Porcentaje acumulado
Femenino	218	47,9	47,9
Masculino	237	52,1	100,0
Total	455	100,0	

Fuente: Elaboración Propia

Tabla 24. Estudiantes entrevistados según la clasificación ordinal del rendimiento académico

Promedio Académico	Frecuencia	Porcentaje	Porcentaje acumulado
Básico	184	40,44	40,44
Alto	135	29,67	70,11
Superior	136	29,89	100
Total	455	100,0	

Fuente: Elaboración Propia

Tabla 25. Estudiantes entrevistados según jornada de su programa académico

¿Cuál es la jornada de su programa académico?	Frecuencia	Porcentaje	Porcentaje acumulado
Diurno	214	47,0	47,0
Mixto	167	36,7	83,7
Nocturno	74	16,3	100,0
Total	455	100,0	

Fuente: Elaboración Propia

Tabla 26. Estudiantes entrevistados según su estado civil

¿Cuál es su estado civil?	Frecuencia	Porcentaje acumulado
Casado	33	7,3
Separado	3	7,9
Soltero	369	89,0
Unión Libre	49	99,8
Viudo	1	100,0
Total	455	

Fuente: Elaboración Propia

Tabla 27. Estudiantes entrevistados según estado civil (Soltero/Otros)

Estado Civil	Frecuencia	Porcentaje	Porcentaje acumulado
Soltero	369	81,1	81,1
Otros	86	18,9	100
Total	455	100	

Fuente: Elaboración Propia

Tabla 28. Estudiantes entrevistados según edad

¿Cuál es su edad?	Frecuencia	Porcentaje	Porcentaje acumulado
17	1	,2	0,2
18	7	1,5	1,8
19	37	8,1	9,9
20	59	13,0	22,9
21	81	17,8	40,7
22	65	14,3	54,9
23	46	10,1	65,1
24	39	8,6	73,6
25	25	5,5	79,1
26	14	3,1	82,2
27	17	3,7	85,9
28	15	3,3	89,2
29	8	1,8	91,0
30	10	2,2	93,2
31	2	,4	93,6
32	7	1,5	95,2
33	5	1,1	96,3
34	2	,4	96,7
35	4	,9	97,6
36	2	,4	98,0
37	2	,4	98,5
39	1	,2	98,7
43	2	,4	99,1
44	2	,4	99,6
47	2	,4	100,0
Total	455	100,0	

Fuente: Elaboración Propia

Tabla 29. Estudiantes entrevistados según su número de hijos

¿Cuántos hijos tiene?	Frecuencia	Porcentaje	Porcentaje acumulado
0	373	82,0	82,0
1	58	12,7	94,7
2	15	3,3	98,0
3	4	,9	98,9
4	5	1,1	100
Total	455	100	

Fuente: Elaboración Propia

Tabla 30. Estudiantes entrevistados según número de personas que dependen económicamente de él

¿Cuántas personas tiene a su cargo?	Frecuencia	Porcentaje	Porcentaje acumulado
0	336	73,8	73,8
1	59	13,0	86,8
2	31	6,8	93,6
3	21	4,6	98,2
4	6	1,3	99,6
5	2	,4	100,0
Total	455	100,0	

Fuente: Elaboración Propia

Tabla 31. Estudiantes entrevistados según con quien viven mientras estudian

Actualmente vive con:	Frecuencia	Porcentaje	Porcentaje acumulado
Alguien diferente a su familia	40	8,8	8,8
Con su cónyuge	53	11,6	20,4
La familia de su cónyuge	6	1,3	21,8
Parientes	48	10,5	32,3
Sólo	25	5,5	37,8
Su familia de origen	271	59,6	97,4
Un compañero(a) o amigo(a)	12	2,6	100,0
Total	455	100,0	

Fuente: Elaboración Propia**Tabla 32. Estudiantes entrevistados según ocupación de su padre**

¿Cuál es la ocupación de su padre?	Frecuencia	Porcentaje	Porcentaje acumulado
Desempleado	77	16,9	16,9
Empleado	157	34,5	51,4
Hogar	10	2,2	53,6
Independiente	176	38,7	92,3
Pensionado	35	7,7	100,0
Total	455	100,0	

Fuente: Elaboración Propia**Tabla 33. Estudiantes entrevistados según ocupación de su madre**

¿Cuál es la ocupación de su madre?	Frecuencia	Porcentaje	Porcentaje acumulado
Desempleado	32	7,0	7,0
Empleado	102	22,4	29,5
Hogar	224	49,2	78,7
Independiente	82	18,0	96,7
Pensionado	15	3,3	100,0
Total	455	100,0	

Fuente: Elaboración Propia**Tabla 34. Estudiantes según nivel educativo logrado por su padre**

¿Seleccione el nivel educativo alcanzado por su padre?	Frecuencia	Porcentaje	Porcentaje acumulado
Bachillerato completo	90	19,8	19,8
Bachillerato incompleto	56	12,3	32,1
Diplomado	3	,7	32,7
Doctorado	2	,4	33,2
Especialización	16	3,5	36,7
Maestría	6	1,3	38,0
Ninguno	27	5,9	44,0
Primaria completa	29	6,4	50,3
Primaria incompleta	68	14,9	65,3
Técnico	57	12,5	77,8
Tecnólogo	30	6,6	84,4
Universitaria Completa	56	12,3	96,7
Universitaria Incompleta	15	3,3	100,0
Total	455	100,0	

Fuente: Elaboración Propia

Tabla 35. Estudiantes según nivel educativo logrado por su madre

¿Seleccione el nivel educativo alcanzado por su madre?	Frecuencia	Porcentaje	Porcentaje acumulado
Bachillerato completo	102	22,4	22,4
Bachillerato incompleto	68	14,9	37,4
Diplomado	6	1,3	38,7
Doctorado	1	,2	38,9
Especialización	16	3,5	42,4
Maestría	6	1,3	43,7
Ninguno	12	2,6	46,4
Primaria completa	35	7,7	54,1
Primaria incompleta	57	12,5	66,6
Técnico	60	13,2	79,8
Tecnólogo	27	5,9	85,7
Universitaria Completa	47	10,3	96,0
Universitaria Incompleta	18	4,0	100,0
Total	455	100,0	

Fuente: Elaboración Propia

Tabla 36. Estudiantes según el origen del dinero para costear sus estudios

¿Cómo costea sus gastos educativos?	Frecuencia	Porcentaje	Porcentaje acumulado
Auxilio empresarial	2	,4	,4
Ayuda familiar	226	49,7	50,1
Beca	67	14,7	64,8
Otra forma	22	4,8	69,7
Préstamos educativos	49	10,8	80,4
Propio sueldo	89	19,6	100,0
Total	455	100,0	

Fuente: Elaboración Propia

Tabla 37. Estudiantes según si desempeñan o no alguna actividad que les represente ingresos económicos

¿Desempeña alguna actividad que represente remuneración?	Frecuencia	Porcentaje	Porcentaje acumulado
No	278	61,1	61,1
Si	177	38,9	100,0
Total	455	100,0	

Fuente: Elaboración Propia

Tabla 38. Estudiantes entrevistados según el tiempo transcurrido entre la finalización del bachillerato y el ingreso a la universidad

Al terminar el bachillerato usted ingresó a la universidad:	Frecuencia	Porcentaje	Porcentaje acumulado
Inmediatamente	168	36,9	36,9
Un semestre después	109	24,0	100,0
Un año después	59	13,0	76,0
Más de un año	119	26,2	63,1
Total	455	100,0	

Fuente: Elaboración Propia

Tabla 39. Estudiantes entrevistados según el tipo de bachillerato del cual se graduó

Su título de bachiller es:	Frecuencia	Porcentaje	Porcentaje acumulado
Académico	273	60,0	60,0
Comercial	64	14,1	74,1
Industrial	11	2,4	76,5
Otro	10	2,2	78,7
Pedagógico	29	6,4	85,1
Técnico	65	14,3	99,3
Validación	3	,7	100,0
Total	455	100,0	100,0

Fuente: Elaboración Propia

Tabla 40. Estudiantes entrevistados según la jornada del colegio donde recibió su título de bachiller

¿Cuál es la jornada del colegio donde recibió el título de bachiller?	Frecuencia	Porcentaje	Porcentaje acumulado
Diurna	446	98,0	98,0
Nocturna	9	2,0	100,0
Total	455	100,0	

Fuente: Elaboración Propia

Tabla 41. Estudiantes entrevistados según si han o no interrumpido sus estudios universitarios

¿Por qué motivo ha interrumpido sus estudios universitarios?	Frecuencia	Porcentaje	Porcentaje acumulado
Desastre natural	3	,7	,7
Desplazamiento forzado	2	,4	1,1
Dificultad laboral	15	3,3	4,4
Dificultad personal	24	5,3	9,7
Dificultades académicas	14	3,1	12,7
Dificultades económicas	94	20,7	33,4
No ha interrumpido sus estudios	303	66,6	100,0
Total	455	100,0	

Fuente: Elaboración Propia

Tabla 42. Estudiantes entrevistados según si han realizado o no otros estudios

¿Ha realizado otros estudios?	Frecuencia	Porcentaje	Porcentaje acumulado
No	272	59,8	59,8
Si	183	40,2	100,0
Total	455	100,0	

Fuente: Elaboración Propia

Tabla 43. Estudiantes entrevistados según si realizan o no otros estudios

¿Está realizando otros estudios?	Frecuencia	Porcentaje	Porcentaje acumulado
No	409	89,9	89,9
Si	46	10,1	100,0
Total	455	100,0	

Fuente: Elaboración Propia

Tabla 44. Estudiantes entrevistados según el grado en que se sientes respetados y orientados por sus profesores

¿Se siente respetado y orientado por sus profesores?	Frecuencia	Porcentaje	Porcentaje acumulado
Casi siempre	221	48,6	48,6
Nunca	4	,9	49,5
Ocasionalmente	57	12,5	62,0
Siempre	173	38,0	100,0
Total	455	100,0	

Fuente: Elaboración Propia

Tabla 45. Estudiantes entrevistados según sus intereses de continuar estudiando luego de adquirir su título universitario

De acuerdo con sus posibilidades e intereses, ¿cómo complementaría su formación universitaria?	Frecuencia	Porcentaje	Porcentaje acumulado
Auto-formación	16	3,5	3,5
Cursos de actualización	62	13,6	17,1
Investigaciones	44	9,7	26,8
Otra carrera	34	7,5	34,3
Postgrados	299	65,7	100,0
Total	455	100,0	

Fuente: Elaboración Propia

Tabla 46. Estudiantes entrevistados según a quien acuden en caso de conflictos personales

¿Cuándo tiene conflictos personales a quién acude?	Frecuencia	Porcentaje	Porcentaje acumulado
Amigos	109	24,0	24,0
Cónyuge	30	6,6	30,5
Madre	183	40,2	70,8
Novio(a)	51	11,2	82,0
Otro familiar	25	5,5	87,5
Padre	41	9,0	96,5
Profesor	1	,2	96,7
Psicólogo	7	1,5	98,2
Sacerdote	8	1,8	100,0
Total	455	100,0	

Fuente: Elaboración Propia

Tabla 47. Estudiantes entrevistados según su relación con su familia

¿Cómo se siente usted en su relación con respecto a su familia?	Frecuencia	Porcentaje	Porcentaje acumulado
Bien	157	34,5	34,5
Mal	3	,7	35,2
Muy bien	254	55,8	91,0
Regular	41	9,0	100,0
Total	455	100,0	

Fuente: Elaboración Propia

Tabla 48. Estudiantes entrevistados según su relación con sus amigos

¿Cómo se siente usted en su relación con respecto a sus amigos?	Frecuencia	Porcentaje	Porcentaje acumulado
Bien	207	45,5	45,5
Mal	1	,2	45,7
Muy bien	215	47,3	93,0
Regular	32	7,0	100,0
Total	455	100,0	

Fuente: Elaboración Propia

Tabla 49. Estudiantes entrevistados según su relación con la Universidad del Magdalena

¿Cómo se siente usted en su relación con respecto a la universidad?	Frecuencia	Porcentaje	Porcentaje acumulado
Bien	236	51,9	51,9
Mal	5	1,1	53,0
Muy bien	158	34,7	87,7
Regular	56	12,3	100,0
Total	455	100,0	

Fuente: Elaboración Propia**Tabla 50. Estudiantes entrevistados según su relación con la ciudad de Santa Marta**

¿Cómo se siente usted en su relación con respecto a Santa Marta?	Frecuencia	Porcentaje	Porcentaje acumulado
Bien	221	48,6	48,6
Mal	13	2,9	51,4
Muy bien	142	31,2	82,6
Regular	79	17,4	100,0
Total	455	100,0	

Fuente: Elaboración Propia**Tabla 51. Estudiantes entrevistados según su relación respecto a Colombia**

¿Cómo se siente usted en su relación con respecto al país?	Frecuencia	Porcentaje	Porcentaje acumulado
Bien	226	49,7	49,7
Mal	15	3,3	53,0
Muy bien	99	21,8	74,7
Regular	115	25,3	100,0
Total	455	100,0	

Fuente: Elaboración Propia**Tabla 52. . Estudiantes entrevistados según las actividades desarrolladas en su tiempo libre**

Seleccione las actividades que realiza en sus ratos libres	Frecuencia	Porcentaje	Porcentaje acumulado
Actividades físicas	7	1,5	1,5
Bailar	10	2,2	3,7
Caminatas	1	,2	4,0
Dormir	7	1,5	5,5
Hacer deporte	31	6,8	12,3
Ir a cine	34	7,5	19,8
Leer	168	36,9	56,7
Oír música	107	23,5	80,2
Otra	13	2,9	83,1
Pasear	11	2,4	85,5
Realizar actividades artísticas (Danza, Coro, Teatro, etc.)	14	3,1	88,6
Ver televisión	25	5,5	94,1
Videojuegos	11	2,4	96,5
Visitar amigos o familiares	16	3,5	100,0
Total	455	100,0	

Fuente: Elaboración Propia**Tabla 53. Estudiantes entrevistados según su pertenencia a grupos**

Pertenece usted a grupos:	Frecuencia	Porcentaje	Porcentaje acumulado
Académicos	39	8,6	8,6

Culturales	19	4,2	12,7
Deportivos	53	11,6	24,4
Ninguno	253	55,6	80,0
Religiosos	63	13,8	93,8
Sociales	28	6,2	100,0
Total	455	100,0	

Fuente: Elaboración Propia

Tabla 54. Apreciación hacia la calidad de los servicios de bienestar estudiantil

Fuente: Elaboración	Frecuencia	Porcentaje	Porcentaje acumulado	Propia
Buena	410	90,1	90,1	
No buena	45	9,9	100	
Total	455	100		