Universidad Nacional Mayor de San Marcos
Universidad del Perú. Decana de América
Facultad de Ciencias Matemáticas
Escuela Profesional de Estadística

Factores asociados al rendimiento académico en los cursos de Matemática Básica y Cálculo I de los alumnos ingresantes de la FCM-UNMSM utilizando regresión logística binaria

TESIS

Para optar el Título Profesional de Licenciado en Estadística

AUTOR

Adolfo Aldo ANCCO CAYLLAHUA

ASESOR

Olga Lidia SOLANO DÁVILA

Lima, Perú

2017



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

https://creativecommons.org/licenses/by-nc-sa/4.0/

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Ancco, A. (2017). *Factores asociados al rendimiento académico en los cursos de Matemática Básica y Cálculo I de los alumnos ingresantes de la FCM-UNMSM utilizando regresión logística binaria*. Tesis para optar el título de Licenciado en Estadística. Escuela Profesional de Estadística, Facultad de Ciencias Matemáticas, Universidad Nacional Mayor de San Marcos, Lima, Perú.

HOJA DE METADATOS COMPLEMENTARIOS

Código ORCID del autor	0000-0001-6555-7246
DNI o pasaporte del autor	44861679
Código ORCID del asesor	0000-0002-7178-0126
DNI o pasaporte del asesor	09118212
Grupo de investigación	Modelos Cuantitativos Aplicados a la investigación Científica (MOCA)
Agencia financiadora	Financiamiento Propio
Ubicación geográfica donde se desarrolló la investigación	País: Perú / Ciudad: Lima/ Distrito: Lima Ciudad Universitaria - UNMSM, Av. República de Venezuela 3400 Coordenadas: -12.060333, -77.082111
Disciplinas OCDE	Estadísticas, Probabilidad http://purl.org/pe-repo/ocde/ford#1.01.03

UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
(Universidad del Perú, DECANA DE AMÉRICA)
FACULTAD DE CIENCIAS MATEMÁTICAS

ESCUELA PROFESIONAL DE ESTADÍSTICA

ACTA DE SUSTENTACIÓN DE TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE LICENCIADO EN ESTADÍSTICA

En la Ciudad Universitaria, Facultad de Ciencias Matemáticas, siendo las ..17:00.. horas del día ..13.. de ..Enero... del año 2017 se reunieron los docentes designados como miembros del Jurado:

Mg. Ana María Cárdenas Rojas (Presidenta)

Lic. Geraldine Judith Vigo Chacón (Miembro)

Mg. Olga Lidia Solano Dávila (Miembro Asesor)

Para la sustentación de la Tesis intitulada **"FACTORES ASOCIADOS AL RENDIMIENTO ACADÉMICO DE LOS CURSOS DE MATEMÁTICA BÁSICA Y CÁLCULO I DE LOS ALUMNOS INGRESANTES DE LA FCM – UNMSM UTILIZANDO REGRESIÓN LOGÍSTICA BINARIA"**, presentada por el Bachiller Adolfo Aldo ANCCO CAYLLAHUA, para obtener el Título Profesional de Licenciado en Estadística.

Luego de la exposición de la Tesis, la Presidenta invitó al expositor a dar respuesta a las preguntas formuladas.

Realizada la evaluación correspondiente por los miembros del jurado, el expositor mereció la aprobación.....Sobresaliente....., con un calificativo promedio de .....Dieciocho.....18.....(letras y números).

A continuación los miembros del jurado, dan manifiesto que el participante Bachiller Adolfo Aldo ANCCO CAYLLAHUA, en virtud de haber aprobado la sustentación de su tesis, será propuesto para que se le otorgue el Título Profesional de Licenciado en Estadística.

Siendo las ..18:00.. horas, se levantó la Sesión, firmando para constancia la presente Acta en tres (3) copias originales.

Lic. Geraldine Judith Vigo Chacón
Miembro

Mg. Olga Lidia Solano Dávila
Miembro Asesor

Mg. Ana María Cárdenas Rojas
Presidenta

By Aldo Adolfo Ancco Cayllahua

Lima, Setiembre 2017

Asesora: Mg Olga Solano Dávila

Título Obtenido: Licenciado en Estadística

La presente investigación está basada en identificar los factores que influyen en el rendimiento académico de los alumnos ingresantes de la Facultad de Ciencias Matemática (F.C.M) de la Universidad Nacional Mayor de San Marcos (UNMSM) en los cursos de Matemática Básica y Cálculo I utilizando el modelo estadístico de Regresión Logística binaria; para identificar los principales factores influyentes se realizó una encuesta dirigida a los alumnos que ingresan a la Facultad de Ciencias Matemáticas matriculados en el semestre 2015-I en los cursos de Matemática Básica y Cálculo I.

PALABRAS CLAVES: Rendimiento académico**,** Regresión Logística Binaria, Matemática Básica, Cálculo I, Encuesta a estudiantes.

SUMMARY

FACTORS ASSOCIATED WITH ACADEMIC PERFOMANCE IN MATHEMATICS BASIC AND CALCULUS I COURSES OF THE STUDENTS INGRESANTS OF THE FCM-UNMSM USING BINARY LOGISTIC REGRESSION

By Aldo Adolfo Ancco Cayllahua

Lima, September 2017

Advisor: Mg Olga Solano Dávila

Degree obtained: Bachelor's degree in statistics

The present research is based on identifying the factors that influence the academic performance of incoming students of the Faculty of Mathematical Sciences of the National University of San Marcos in the Basic Mathematics and Calculus I courses using the logistic regression model; in order to identify the determining factors, a survey was conducted for the students enrolled in the Faculty of Mathematical Sciences enrolled in the semesters 2015-I in the courses of Basic Mathematics and Calculus I.

KEY WORDS: Academic achievement, binary logistic regression, Basic Mathematics, Calculus I, Survey students.

INDICE

INTRODUCCIÓN 6

CAPÍTULO I: 8

MARCO REFERENCIAL 8

1.1 PLANTEAMIENTO DEL PROBLEMA 9

1.2 OBJETIVO GENERAL 10
1.2.1 OBJETIVOS ESPECÍFICOS 10

1.3 ANTECEDENTES DEL PROBLEMA 10
1.3.1 ANTECEDENTES A NIVEL NACIONAL 10
1.3.2 ANTECEDENTES A NIVEL INTERNACIONAL 13

1.4 METODOLOGÍA 10
1.4.1 TIPO DE ESTUDIO 17
1.4.2 MUESTRA O TRATAMIENTO DE LOS DATOS 17
1.4.3 DESCRIPCIÓN DE LAS VARIABLES 19

CAPÍTULO II: 20

EL MODELO DE REGRESIÓN LOGÍSTICA 20

2. EL MODELO DE REGRESIÓN LOGÍSTICA 21

2.1 INTRODUCCIÓN 21

2.2 EL MODELO DE REGRESIÓN LOGÍSTICA SIMPLE 22

2.3 AJUSTE DEL MODELO DE REGRESIÓN LOGÍSTICA SIMPLE 24

2.4 PRUEBA DE HIPÓTESIS PARA LOS PARÁMETROS 26

2.5 ESTIMACIÓN POR INTERVALO DE CONFIANZA. 29

2.6 PRUEBA DE HOSMER – LEMESHOW 30

3. EL MODELO DE REGRESIÓN LOGÍSTICA MÚLTIPLE 31

3.1 AJUSTE DEL MODELO DE REGRESIÓN LOGÍSTICA MÚLTIPLE 32

3.2 PRUEBA PARA LA SIGNIFICANCIA DEL MODELO 33

3.3 ESTIMACIÓN POR INTERVALOS DE CONFIANZA 34

CAPÍTULO III 36

ANÁLISIS DE DATOS 36

3. ANÁLISIS DE DATOS 37

3.1 ANÁLISIS EXPLORATORIO DE DATOS 37

3.2 ANÁLISIS EXPLORATORIO DE LAS VARIABLES CUALITATIVAS CON RESPECTO AL RESULTADO FINAL DEL CURSO DE MATEMÁTICA BÁSICA 37

3.3 ANÁLISIS EXPLORATORIO DE LAS VARIABLES CUALITATIVAS CON RESPECTO AL RESULTADO FINAL DEL CURSO DE CÁLCULO I 55

4 AJUSTE DEL MODELO DE REGRESIÓN LOGÍSTICA 71

4.1 ANÁLISIS DE RESULTADOS 72

4.1.1AJUSTE DEL MODELO MEDIANTE EL MÉTODO DIRECTO 72

4.2 AJUSTE DEL MODELO MEDIANTE EL MÉTODO DE WALD HACIA ADELANTE (FORWARD STEPWISE (WALD)) PARA LA SELECCIÓN DE VARIABLES 77

CAPÍTULO IV CONCLUSIONES Y RECOMENDACIONES 88

CONCLUSIONES 89

RECOMENDACIONES 91

BIBLIOGRAFÍA 92

ANEXOS 94

ANEXO 1 95

ANEXO 2 97

ANEXO 3 103

ANEXO 4 116

INTRODUCCIÓN

A lo largo de los años en la Facultad de Ciencias Matemáticas de la Universidad Nacional Mayor de San Marcos se observa que una gran cantidad de alumnos ingresantes reprueban los cursos de Matemática Básica y Cálculo I (Dirección Academia de la Facultad de Ciencia Matemáticas), cursos que son llevados en el primer semestre de estudio y que son considerados básicos en su formación académica ya que de estos dependen la comprensión y éxito de los cursos posteriores de su formación académica. Los cursos requieren que los estudiantes tengan habilidades matemáticas sólidas y una buena comprensión de lógica matemática, conjuntos, sistemas numéricos, matrices, determinantes y polinomios.

Un indicador del bajo performance en estos cursos es por ejemplo la nota media final en el semestre 2015 – I de los estudiantes que ingresan a la F.C.M en los cursos de Matemática Básica fue de 8 y en el curso de Cálculo I fue 5.

Es por tal motivo que este bajo performance genera muchas interrogantes e hipótesis, en algunos casos se determina como causa la falta de preocupación del alumno por estudiar el curso o en otros casos se cuestiona si el profesor tiene la metodología adecuada para llegar al alumno.

Para iniciar este estudio partimos de la necesidad de entender y comprender como algunas variables pueden influir en el rendimiento académico de los alumnos y posteriormente su éxito en el desarrollo de los cursos.

Definimos Rendimiento académico como la capacidad que tienen los alumnos a responder a metas a educativas establecidas previamente. Si el estudiante tiene un nivel bajo podemos concluir que no ha captado de manera adecuada y completa los conocimientos. Así mismo no contiene las herramientas y habilidades básicas para la resolución de problemas referente al material de estudio.

En primer lugar, podemos partir de la formación del alumno previo al ingreso a la universidad marcada por su etapa escolar (que será medida mediante características como Institución Educativa donde estudió, etc.) y la preparación pre-universitaria que tiene como

principal objetivo dotar al alumno de determinados conocimientos con el fin de rendir con éxito la prueba de admisión a la universidad.

En esta investigación se intenta lograr la más clara explicación sobre el rendimiento académico (si aprueba o no el curso) del alumno en función a un conjunto de variables seleccionadas de diversos estudios que abordan el tema del rendimiento académico de los alumnos.

Este trabajo de investigación está compuesto de cuatro Capítulos, el primero comprende el Marco referencial y contiene el planteamiento del problema, objetivo, antecedentes y metodología, el Capítulo II contiene el marco teórico del modelo de Regresión Logística, el Capítulo III presenta el análisis de datos de los resultados al aplicar la encuesta a los alumnos ingresantes que estaban estudiando los cursos de Matemática Básica y Cálculo I durante el Semestre 2015-I y finalmente el Capítulo IV que comprende las conclusiones y recomendaciones.

CAPÍTULO I:

MARCO REFERENCIA

1.1 PLANTEAMIENTO DEL PROBLEMA

El presente trabajo de tesis nace de la observación realizada sobre el fenómeno que se repite año tras año en la Facultad de Ciencias Matemáticas de la Universidad Nacional Mayor de San Marcos: la gran cantidad de alumnos ingresantes que desaprueban los cursos de Matemática Básica y Cálculo I, cursos que son percibidos como absolutamente necesarios para el éxito de los futuros cursos de la formación académica de los estudiantes.

Según la información brindada por la Oficina de Matricula de la Facultad de Ciencias Matemáticas indican que por ejemplo en la Escuela Académica Profesional de Estadística en el Semestre 2010-I el 57% de los alumnos matriculados en el curso de Matemática Básica desaprobó el curso, así mismo en el semestre 2011-I el porcentaje de desaprobados fue de 82%, en el semestre 2012-I fue de 88%, en el semestre 2013-I fue de 83% y en el semestre 2014-I el porcentaje de desaprobados fue de 73% (Dirección Academia de la Facultad de Ciencia Matemáticas).

Así mismo en el curso de Cálculo I los porcentajes de desaprobados son igual de altos, en el semestre 2010-I el 76% de alumnos desaprobaron el curso, en el semestre 2011-I el porcentaje fue de 93%, en el semestre 2012-I fue de 81%, en el semestre 2013-I el porcentaje fue de 91% y en el semestre 2014-I el porcentaje de desaprobados fue de 92%(Dirección Academia de la Facultad de Ciencia Matemáticas).

Es por tal motivo que se pretende investigar qué factores inciden en el bajo rendimiento académico de los alumnos y que trae como consecuencia desaprobar las materias.

Existen diversas investigaciones dirigidas al estudio de factores que influyen en el bajo rendimiento académico de los alumnos universitarios las cuales explican e incorporan posibles factores. Desde factores de carácter personal (género, edad, situación socioeconómica, etc.) como factores académicos (Institución Educativa de procedencia, tipo de preparación, etc.), estos actúan de manera silenciosa por lo que no son fáciles de identificar como responsables directos del bajo rendimiento académico que conlleva al éxito o fracaso en determinadas materias.

1.2 OBJETIVO GENERAL

Identificar cuáles son los factores asociados al rendimiento académico en los cursos de Matemática Básica y Cálculo I de los alumnos ingresantes a la Facultad de Ciencias Matemáticas de la UNMSM utilizando el modelo de regresión logística.

1.2.1 OBJETIVOS ESPECÍFICOS

Identificar cuáles son los factores asociados al rendimiento académico del curso de Matemática Básica de los alumnos que ingresan a estudiar a la F.C.M de la UNMSM utilizando el modelo de regresión logística binario.

Identificar cuáles son los factores académicos asociados al rendimiento académico del curso de Cálculo I de los alumnos que ingresan a estudiar a la F.C.M de la UNMSM aplicando el modelo estadístico de Regresión Logística binario.

1.3 ANTECEDENTES DEL PROBLEMA

1.3.1 ANTECEDENTES A NIVEL NACIONAL

El autor Plasencia, S. (2010) realizó la tesis titulada "*Factores relacionados con el rendimiento académico en matemática en los estudiantes de la Universidad Nacional de Educación Enrique Guzmán y Valle" en el año 2010".* (Tesis de Maestría). Universidad Nacional de Educación "Enrique Guzmán y Valle". Perú. Sostiene que la matemática brinda a los profesionales reglas, técnicas, instrumentos, pues es formativa; por lo tanto, su aprendizaje es importante para los universitarios.

Además, afirma que existen diversas causas que influyen en el aprendizaje de esta ciencia, algunos de ellos relacionados al estudiante como la inteligencia, ansiedad, antecedentes escolares, salud, motivación u otras causas internas. Otro factor puede ser en relación al docente como la preparación académica, metodología, didáctica, experiencia o satisfacción laboral.

Otro factor, se relaciona con el Currículo y la programación de la enseñanza; es decir, la forma cómo se planifican los objetivos, el contenido temático, las estrategias, los recursos, la forma de evaluación, etc. También, otro aspecto importante es relación entre docente y

estudiante, específicamente sobre la estrategia de trabajo que aplica el docente durante el proceso de enseñanza.

Un estudio realizado por Álvarez, J. (2011) En su artículo titulado "Causas endógenas y exógenas del rendimiento académico de los estudiantes de matemática de la Facultad de Ciencias de la Educación de la UNJBG de Tacna. Ciencia y Desarrollo. Perú – Tacna." Señala que las academias preuniversitarias influyen negativamente en los estudiantes, ya que se centran en mecanizarlo y entrenarlo para ingresar a una universidad, sin desarrollar capacidades como el análisis y el razonamiento.

Otra de las causas que influyen en el bajo nivel de los estudiantes es que estas academias no desarrollan en los estudiantes hábitos de estudio, proporcionan una preparación deficiente, no promueven el pensamiento abstracto, analítico y la generalización. Además, existe una brecha entre el programa curricular de la Educación Básica Regular del nivel secundario con el contenido que se exige en la universidad.

La diversidad de contenidos del curso de Matemática que se ofrecen a los universitarios del primer ciclo no es suficiente para brindar un conocimiento básico sobre dicho curso. Con respecto a la metodología, el docente no utiliza un método que optimice el aprendizaje, hay una falta de conexión entre docente y estudiante, pues este no está motivado, ya que el docente sólo centra su atención en la enseñanza y no en el aprendizaje; también influye la falta de material bibliográfico, equipos y herramientas tecnológicas.

Con respecto a la evaluación, el docente desconoce los modelos de construcción de criterios e instrumentos de evaluación de carácter cualitativo. Otra de las razones del bajo nivel de los estudiantes universitarios, es porque el docente no se involucra ni centra en el aprendizaje de los estudiantes; además, se debe cambiar el proceso de evaluación, programa y cambiar la visión que se tiene sobre el área de matemática.

Hernández, D. (2011) en su tesis titulada "Impacto de la didáctica en el rendimiento académico universitario". (Tesis de Licenciatura). Universidad Nacional Federico Villarreal. Perú. Afirma que el bajo rendimiento académico tiene causas externas e internas. Entre los factores externos encontramos los conflictos familiares como falta de atención, violencia o familias permisivas, afectan en la conducta y desarrollo académico.

La crisis social, como la violencia o falta de valores, así como las relaciones sociales influyen en el nivel académico de los estudiantes, pues no permiten la formación de un ciudadano con actitudes positivas que se desenvuelva adecuadamente. Además, la universidad es otro factor, pues los estudiantes pasan la mayor parte de su tiempo allí y están rodeados de distintas realidades que pueden influir en su aprendizaje.

La plana docente puede influir del fracaso de los estudiantes, ya que pueden transmitir inseguridad con respecto al curso, su falta de motivación o atención hacia los estudiantes, inclusive la falta de un método de enseñanza, son factores a considerar. Las causas internas que agravan esta situación, son la falta de afecto familiar o problemas conductuales como hiperactividad, déficit de atención, problemas de autoestima y psicológicos.

También, se puede identificar causas físicas relacionados a la vista, problemas auditivos, del lenguaje o de carácter crónico; además de factores intelectuales, desbalance entre la edad y el desarrollo intelectual, estudiantes con necesidades especiales o superdotados que se cansan de los contenidos pues les parecen insuficientes.

El bajo nivel en la Matemática se debe a diversas razones, entre las cuales se pueden mencionar la enseñanza de una malla curricular deficiente, el constante uso de herramientas como la calculadora y la formación del docente de Matemática, estas causas afectan seriamente al aprendizaje de esta disciplina.

Ocaña, Y. (2012) en su artículo "variables académicas que influyen en el rendimiento académico de los estudiantes universitarios". Investigación Educativa. Universidad Nacional Mayor de San Marcos. Perú. Sostiene que para asegurar una educación superior de calidad es necesario reconocer y estudiar las razones que afectan al rendimiento académico,

Además, el problema se le puede atribuir a una preparación escolar deficiente, especialmente de instituciones estatales, pues los estudiantes poseen habilidades; sin embargo, se necesitan cambios en el proceso de admisión, el currículo y el método de enseñanza, con el fin de optimizar el rendimiento académico y disminuir el porcentaje de desaprobados.

Uno de los factores de mayor relevancia es la variable académica, ya que se relaciona con el rendimiento académico previo del estudiante y al programa curricular. Así mismo, el otro factor de carácter educativo es la variable pedagógica que se relaciona con el esfuerzo del docente y es de carácter interno.

Dentro de las variables académicas encontramos las características de las instituciones educativas de donde proviene el estudiante, el rendimiento escolar, el rendimiento en las pruebas de aptitud y de admisión, entre otros factores académicos que permitirían un resultado positivo en el estudiante, si son usados de manera adecuada.

1.3.2 ANTECEDENTES A NIVEL INTERNACIONAL

La autora Roque, M. (2005) en su tesis titulada "Factores que influyen en el rendimiento de la matemática en el estudiante del ciclo básico, del Instituto Oficial Mixto Básico Leonídas Méncos Ávila. Tiquisate – Escuintla". (Tesis de Licenciatura) Universidad de San Carlos de Guatemala. Guatemala. Sostiene que las causas que impactan en el rendimiento académico pueden ser de carácter positivo que permiten el logro de un alto nivel académico de los estudiantes y otros de carácter negativo.

Afirma que el área de Matemática es el curso en el que se presenta mayores dificultades y con resultados desaprobatorios. Entre las causas de carácter positivo tenemos docentes competentes, una metodología de enseñanza adecuada, apoyo familiar, buena salud física y mental, gusto por la matemática, adecuada preparación y uso de material bibliográfico adecuado.

Las causas de carácter negativo que influyen en el bajo rendimiento son: docentes incompetentes, sin especialización, metodología de enseñanza deficiente, indisciplina, falta de apoyo familiar o problemas de salud.

El autor Izar, J. (2011) en su trabajo titulado "Factores que afectan el desempeño académico de los estudiantes del nivel superior en Rioverde, San Luis de Potosí, México". Universidad de Veracruz. México. Señala que las dificultades que afronta educación superior de México es el alto nivel de desaprobados, el abandono y el bajo nivel de competencia de los egresados, producto de un plan curricular insuficiente, metodología desfasada, una brecha entre la teoría y la práctica, docentes que desconocen las necesidades de aprendizaje y orientación insuficiente. Además, la falta de talleres que refuerce las habilidades de los estudiantes, la situación socioeconómica y la orientación vocacional, son causantes de un bajo rendimiento.

Según Barahona, P. (2012) en su tesis titulada "Factores determinantes del rendimiento académico de los estudiantes de la Universidad de Atacama". (Tesis de Maestría). Universidad de Atacama. Chile. Sostiene que el nivel académico de los estudiantes depende

de diversas causas que pueden ser factores personales y sociales. También se concluyó que el nivel de escolaridad de los padres no es una variable significativa que afecta al desempeño del estudiante.

La prueba Second Internacional Adult Literacy Survey (SIALS) realizada en Chile en el año 1998, evidenció que solo el 8% de profesionales obtuvieron el nivel de logro al comprender y desarrollar problemas de aritmética básica. Es así que la educación universitaria no garantiza que un profesional promueva un ambiente intelectual y lleve al éxito de sus hijos en su etapa universitaria, especialmente si provienen de niveles sociales medios o bajos como la Universidad de Atacama; por tal motivo, el nivel o formación profesional de los padres no es una variable que influya en el rendimiento académico.

Las variables que consideran aspectos verbales y matemáticos, ha demostrado que pueden influir en el rendimiento académico. Además, con relación a los coeficientes, se debe considerar que indican en cuantas unidades varía la variable dependiente cuando la variable dependinte varía en una unidad.

Además, existen otras investigaciones internacionales que también han demostrado que existe una relación entre los resultados obtenidos para ingresar a la universidad con el rendimiento académico del universitario. Así mismo, nos permite reconocer que las diversas capacidades de interpretación de textos y las capacidades matemáticas pueden influenciar positivamente en el rendimiento del estudiante.

Esta situación se ha identificado en los estudiantes universitarios chilenos, pues al ingresar en los primeros años muestran deficiencias en el campo de las matemáticas que suelen determinar su rendimiento a futuro.

Según las autoras Martínez, H.; Ramírez, G. & Zalazar, L. (2011) en su tesis titulada "Análisis del bajo rendimiento en matemática de los ingresantes de la Facultad de Ciencias Económicas." (Tesis Maestría) Universidad Nacional del Nordeste. Argentina. Sostienen que, en la Facultad de Ciencias Económicas, el proceso de enseñanza de esta disciplina no se desarrolla de una manera adecuada debido a la gran diversidad de estudiantes en cuestión de habilidades, formación académica previa, hábitos de estudio; inclusive por las características propias de esta asignatura, cuyos conceptos exigen una base de conocimientos

para asi lograr un aprendizaje significativo. Lan totalidad de estas causas influyen significativamente en el rendimiento académico del área de matemática.

Los autores Tejedor, J. y Valcárcel, A. (2011) en su artículo titulado "Causas del bajo rendimiento del estudiante universitario (en la opinión de profesores y alumnos) Propuestas de mejora en el marco de la EEES *(Espacio Europeo de Educación Superior)".* (Tesis de Grado) Universidad de Salamanca. España. Afirman que España se ubique en el primer lugar con respecto a la cantidad de estudiantes universitarios, debido a la gran demanda.

Esta situación que demuestra una tasa alta de universitarios, se ha visto perjudicado y ha llevado al fracaso académico, pues da como resultado un alto nivel de competitividad para alcanzar una oportunidad laboral.

Además, el alto índice de demanda universitario no ha permitido una adecuada organización y estructuración de la educación, afectando así a la calidad educativa. También, sostienen es difícil poder identificar y delimitar de manera precisa las causas que afectan al rendimiento académico, pues están entrelazadas entre sí.

Según el punto de vista en que se analicen estos factores, se identifican los relacionados a los rasgos de personalidad, actitud, nivel social, formación académica, estilos y métodos de aprendizaje, metodología del docente, etc.

Por tal motivo, se proponen cinco variables: Variables de identificación que abarcan el género y la edad; variables psicológicas relacionadas con aspectos intelectuales, de personalidad, métodos de aprendizaje, etc; variables académicas enfocadas en la formación académica y tipo de preparación para ingresar a una universidad; variables pedagógicas con respecto al método y estrategias de enseñanza, proceso de evaluación y las variables relacionadas al aspecto familiar.

1.4 METODOLOGÍA

El presente trabajo propone el análisis de los factores asociados al rendimiento académico de los estudiantes que ingresan a la F.C.M de la UNMSM mediante la utilización de una de las técnicas del análisis multivariante, la Regresión Logística.

Para el cálculo de la variable dependiente "Resultado Final del Curso" se utilizaron los datos de los promedios finales de los estudiantes que se matricularon en los cursos de Matemática

Básica y Cálculo I, pudiendo obtener como resultado al finalizar el semestre aprobar o desaprobar los cursos.

Para determinar el promedio final el profesor del curso realiza el cálculo de la siguiente manera:

Promedio Final = (Examen parcial + Examen Final + Promedio de prácticas)/3.

Se considera que el alumno aprobó el curso si obtiene un promedio final mayor o igual a 10,5.

La información de los promedios finales del curso de Matemática Básica y Cálculo I fue brindada por la Dirección Académica de la F.C.M, al finalizar el Semestre Académico 2015-I.

Para determinar las variables independientes se manejaron distintas variables propuestas en diferentes investigaciones y que están relacionadas con el desarrollo del alumno en la universidad tales como el lugar donde estudia después de clase, el tiempo que le toma llegar a la Universidad, el lugar donde ingiere sus alimentos, si el alumno estudia otra carrera o cursos en paralelo o si trabaja en la actualidad, también se consideraron variables independientes relacionadas con la características del hogar del alumno y su entorno familiar tales como el tipo de vivienda, la ubicación, si cuenta con internet en el hogar, con quien convive, la condición laboral y nivel educativo de los padres, sin embargo durante el análisis del modelo de regresión logística no todas fueron significativas por lo que algunas no fueron consideradas.

Para la recopilación de la información de las variables independientes se elaboró un cuestionario (ver Anexo 4) que fue aplicado entre la primera y segunda semana del mes de Junio del año 2015. Para la localización de los alumnos matriculados en el semestre 2015-I en los cursos de Matemática Básica y Cálculo I, se solicitó la información respectiva a la Dirección Académica de la Facultad de Ciencia Matemáticas.

1.4.1 TIPO DE ESTUDIO

En el presente trabajo el tipo de estudio es transversal debido a que se va desarrollar en un solo momento, cuantitativo ya que la información será tomada de encuestas, correlacional debido a que uno de los fines es observar si dos variables están correlacionadas o no y multivariado debido a que se va determinar la contribución de varios factores en un evento o resultado.

1.4.2 MUESTRA O TRATAMIENTO DE LOS DATOS

1.4.2.1 DISEÑO MUESTRAL

Población objetivo. - Alumnos ingresantes que se matricularon en el periodo 2015-I, en la F.C.M de la UNMSM.

Unidad de Análisis. - La unidad de análisis corresponde a cada uno de los alumnos ingresantes matriculados el semestre 2015-I, en la F.C.M de la UNMSM.

Tipo de muestreo. - Muestreo aleatorio simple (M.A.S)

Tamaño de muestra. - Para determinar el tamaño de la muestra se utilizó el muestreo aleatorio simple, el cálculo se hizo mediante la siguiente formula:

$$n = \frac{Np(1-p)}{(N-1)D + p(p-1)}$$

Dónde:

N: 300

n: representa el tamaño de la muestra

p: representa el valor aproximado del porcentaje de estudiantes que aprueban los cursos de Matemática Básica y Cálculo I.

$$D = (\frac{E}{Z})^2$$

E: Margen de error.

Z: representa el valor de la abscisa en la distribución normal.

Así mismo se consideró las siguientes precisiones:

- Los resultados fueron obtenidos con un nivel de confianza del 95% (Z=1.96),
- p=0.5
- Un error relativo de 7.7%.

Para una población de 300 alumnos, con un margen de error de 0.077, y un nivel de confianza del 95%, se obtiene una muestra de 105 estudiantes a encuestar.

Para seleccionar a los estudiantes se utilizó el marco muestral de alumnos matriculados que estén cursando paralelamente el curso de Matemática Básica y Cálculo I, de tal forma que con la variable dependiente ("resultado final del curso") que se obtiene para cada uno de los cursos, se elaboren dos bases de datos y se desarrolle un modelo logístico independiente para cada curso.

1.4.3 DESCRIPCIÓN DE LAS VARIABLES

En el siguiente cuadro resumen se presentan las variables consideradas para el análisis

Cuadro 1.4.3: Variables consideradas en la investigación

Variable	Tipo	Nivel de medición	Categorías
Resultado Final del curso	Cualitativa	Nominal	0: Desaprobado
			1: Aprobado
Sexo	Cualitativa	Nominal	0 : Hombre
			1: Mujer
Edad	Cualitativa	Ordinal	0: Menos de 20 años
			1: 20 años o Más
Tipo de Colegio	Cualitativa	Nominal	0: Estatal/Nacional
			1: Particular/Preuniversitario
Lugar de preparación	Cualitativa	Nominal	0: Academia
			1: Pre San Marcos
			2: Casa
			3: Grupo de Estudio
Preferencia por la carrera	Cualitativa	Nominal	0: Si era su preferida
			1: No era su preferida
Condición laboral	Cualitativa	Nominal	0: Si trabaja
			1: No trabaja
Tiempo de estudio en el día después de clase	Cualitativa	Ordinal	0: Menos de una hora
			1: Entre una y tres horas
			2: Más de tres horas
Resultado de examen parcial	Cualitativa	Nominal	0: Aprobado
			1: Desaprobado
Con quien convive el alumno	Cualitativa	Nominal	0: Con padres y hermanos
			1: Sólo con padre/madre
			2: Sólo
			3: Otros familiares
Nivel académico del padre	Cualitativa	Ordinal	0: Sin nivel
			1: Primaria incompleta/completa
			2: Secundaria incompleta/completa
			3: Técnico incompleto/Incompleto
			4: Universitario incompleto/Completo
Nivel académico de la madre	Cualitativa	Ordinal	0: Sin nivel
			1: Primaria incompleta/completa
			2: Secundaria incompleta/completa
			3: Técnico incompleto/Incompleto
			4: Universitario incompleto/Completo

CAPÍTULO II:

EL MODELO DE REGRESIÓN LOGÍSTICA

2. EL MODELO DE REGRESIÓN LOGÍSTICA

2.1 INTRODUCCIÓN

Los modelos de Regresión Lineal asumen que la variable respuesta Y es de tipo cuantitativa, sin embargo, existen muchos casos donde la variable respuesta es de del tipo cualitativa. Por ejemplo, podemos definir como variable de interés el color de los ojos de un conjunto de personas, el resultado de esta variable nos mostrara distintas categorías según el color de los ojos de las personas pudiendo ser marrones, azules o negros.
Bajo ciertas condiciones, el modelo de regresión lineal es una herramienta valiosa para cuantificar los efectos de varias variables explicativas sobre una variable continua dependiente. Para situaciones donde la variable dependiente es cualitativa, sin embargo, se han desarrollado otros métodos. Uno de ellos es el modelo de Regresión Logística.

La Regresión Logística es un modelo estadístico que nos permite calcular la probabilidad de que un conjunto de individuos pueda pertenecer a una clase de interés, para esto el modelo estadístico de Regresión Logística considera que la variable dependiente Y es binaria la cual puede tomar dos posibles valores (0 o 1) es así que cada uno individuo va tener una probabilidad de tomar el valor 0 o 1.

Los modelos de Regresión Logística indican que deben existir una relación por parte de la variable respuesta y el conjunto de variables regresoras, así mismo las variables independientes pueden ser cualitativas o cuantitativas. Otro punto a considera es que las variables independientes no deben guardar relación entre ellas.

Supongamos que tenemos una variable aleatoria binaria que sigue una distribución Bernoulli que puede tomar el valor 1 o 0 con probabilidad $\pi(x)$ o $1-\pi(x)$ respectivamente. De hecho, $\pi(x)$ representa la probabilidad condicional $P(Y=1|\ x)$ dado $y=1$, basados en la variable respuesta binaria utilizamos la distribución logística que se representa de la siguiente manera

$$\pi(x)=\frac{e^{(\beta_0,\beta_1 x_1+\cdots+\beta_p x_p)}}{1+e^{(\beta_0,\beta_1 x_1+\cdots+\beta_p x_p)}} \qquad (2,1,1)$$

Donde $x = (x_1, x_2, \ldots, x_p) \in R^p$ es un vector de p variables explicativas y $\beta_0, \beta_1, .. \beta_p$ parámetros desconocidos.

2.2 EL MODELO DE REGRESIÓN LOGÍSTICA SIMPLE

Los modelos de Regresión Lineal demandan un conjunto de supuestos que generalmente son complicados de ejecutar (variables tipo continuas y lineales), la ventaja del modelo estadístico de Regresión Logística es que no necesita el cumplimiento de algún supuesto en cuanto a la distribución de la probabilidad de las variables independientes de este modo las variables independientes pueden tener o no una distribución normal o las variables pueden ser del tipo cualitativo. (Hosmer &Lemeshow, 2000).

Si se posee dos conjuntos de individuos, de tal forma que un individuo puede formar parte de uno de los conjuntos (conjunto de los estudiantes que aprueban y los que no aprueban, por ejemplo), el modelo estadístico de Regresión Logística estima que tan probable es que un individuo pertenezca a uno de los conjuntos.

Sea Y la variable dependiente (binaria para la presente investigación) y un conjunto de variables llamémosla X. Siendo Y una variable binaria se le puede asignar el valor de ‘‘0’’ en caso el hecho no suceda y ‘‘1’’ caso contrario. Las variables independientes podrían ser del tipo cuantitativas o cualitativas. La probabilidad de que Y=1 es igual a p.

En los modelos de regresión nuestro valor principal es el valor medio de la variable respuesta determinado por el conjunto de variables regresoras.
A este valor se le denomina “media condicional”’ y se expresa E(Y/x) donde “Y” y “X” hacen referencia a los valores de la variable dependiente e independiente respectivamente. En los modelos de regresión lineal se asume que el valor medio se puede expresar en forma de una ecuación lineal en función a x:

$$E(Y/x) = \beta_0 + \beta_1 x \qquad (2.2.1)$$

Esta fórmula determina que es factible que E(Y/x) puede asumir algún valor de x y este se encuentre dentro del rango -∞ y +∞ .

Con el fin de hacer más simple la notación, se utilizará $\pi(x) = E(\ Y\ /x)$ para expresar el valor medio de Y/x cuando se utiliza la distribución logística. Por lo tanto, la expresión propia del modelo de Regresión Logística que se utilizara es:

.

$$P(y = 1) = \pi(x) = \frac{e^{\beta_0 + \beta_1 X}}{1 + \ e^{\beta_0 + \beta_1 X}} \qquad (2.2.2)$$

Un reajuste de $\pi(x)$ que es muy importante cuando se utiliza la regresión logística es la transformación Logit. Este reajuste este definido en función a $\pi(x)$ de la siguiente forma:

$$g(x) = \ln\left[\frac{\pi(x)}{1 + \ \pi(x))}\right] \qquad (2.2.3)$$

$$= \beta_0 + \beta_1 X$$

El alcance de este reajuste indica que la función g(x) posee varios atributos propios de la regresión lineal.

Otra diferencia clave que muestran los modelos de regresión lineal y logística hacen referencia a la distribución condicional de la variable dependiente. Los modelos de lineales se supone que un valor de la variable dependiente se puede expresar como $y = E(Y/x) + \varepsilon$. El valor de ε se le denomina error y nos muestra la desviación de un valor de la media condicionada. La conjetura más característica es que ε posee una distribución de Gauss o también denominada Normal con media cero y algunas variaciones que es constante a través de los niveles de la variable independiente.

Se Infiere que la distribución condicionada de la variable dependiente dado x será normal con una media E(Y/x), y una varianza que será constante.

Para el caso que la variable dependiente es binaria se puede expresar este valor dado x como $y = E(Y|x) + \varepsilon$. En este caso el valor de ε puede asumir uno de dos probables valores. Si $\varepsilon = 1 - \pi(x)$ con probabilidad $\pi(x)$ entonces Y=1. Por consiguiente, ε tendrá una distribución con valor medio igual a cero y varianza equivalente a $\pi(x)$[1- $\pi(x)$].

En definitiva, la distribución condicional de la variable dependiente tendrá una distribución binomial, con una probabilidad de distribución propuesta por la media condicional $\pi(x)$. (Hosmer &Lemeshow, 2000).

2.3 AJUSTE DEL MODELO DE REGRESIÓN LOGÍSTICA SIMPLE

Asumamos que se tiene un conjunto con "n" atributos que no guardan relación entre ellas(independientes) del par *(x_i, y_i), donde i puede ir de 1, 2,..., n.* , y_i asume el valor de la variable respuesta dicotómica y x_i *toma* el valor de la variable respuesta para el *i-esimo* individuo. Así mismo, adjudicamos que la variable respuesta puede tomar el valor de 0 o 1 donde 0 indica que un sujeto posee determinada característica y 1 caso contrario. Para realizar el ajuste del modelo de Regresión Logística que se representa en la función *(2.2.2)* para un conjunto de datos hace falta que se estimen los valores de β_0 y β_1, los cuales son desconocidos.

En el caso de la Regresión logística el método que comúnmente se utiliza para la estimación de parámetros que no se conocen es el llamado "Mínimos Cuadrados".
Al utilizar el método de Mínimos Cuadros en la Regresión Lineal los estimadores cumplen determinadas propiedades estadísticas las cuales no se cumplen en caso la variable de salida sea binaria como es el caso de la Regresión Logística

El método que utiliza la regresión Logística para la estimación de los parámetros se le denomina Máxima Verosimilitud. Este método entrega valores a los parámetros que no se conocen con la mayor probabilidad de tener como resultado el conjunto de datos que se observa. El método de Máxima Verosimilitud parte de la construcción de la función, denominada la función de verosimilitud.

En caso se tipifique como 0 o 1 la expresión de $\pi(x)$ determinada en la ecuación (2.2.2) nos indica (por los valores arbitrarios de $\beta = (\beta_0, \beta_1)$, el par de parámetros) la probabilidad condicional de que Y es equivalente a 1 dado un valor de x . Esto se expresa como $P(Y = 1/x)$. De ello se infiere que $1 - \pi(x_i)$ de la probabilidad condicional de Y determinado por un valor de x es igual a cero, $P(Y = 0 \ / \ x)$.

De tal forma que para los distintos pares (x_i, y_i), donde $y_i = 1$, el aporte a la función de probabilidad es $\pi(x_i)$, y en el caso donde los pares $y_i = 0$, el aporte a la función de probabilidad es 1- $\pi(x_i)$, donde la cantidad $\pi(x_i)$ expresa los valores de $\pi(x)$ calculada en x_i . Otro modo adecuado de expresar el aporte a la función de probabilidad para el par (x_i, y_i) es a través de la función .

$$\pi(x_i)^{y_i} = [1 - \pi(x_i)]^{1-y_i} \tag{2.3.1}$$

Puesto que se asumimos que las observaciones no guardan relaciones entre sí, la función de verosimilitud se calcula multiplicando los componentes de la ecuación (2.3.1) de la siguiente forma:

$$l(\beta) = \prod_{i=1}^{n} \pi(x_i)^{y_i}[1 - \pi(x_i)]^{1-y_i} \tag{2.3.2}$$

El método de máxima verosimilitud indica que la expresión de la ecuación (2.3.2) es maximizada por los valores que tome β . No obstante, esta expresión es matemáticamente más sencilla de trabajar utilizando el logaritmo de la ecuación (2.3.2). La expresión de logaritmo de verosimilitud está definida bajo la siguiente estructura:

$$L(\beta) = \ln [l(\beta)] = \sum_{i=1}^{n}\{y_i \ln [\pi(x_i)] + (1 - y_i) \ln [1 - \pi(x_i)]\} \tag{2.3.3}$$

Para determinar los valores de β que maximicen $L(\beta)$ se calcula la diferencia $L(\beta)$ con respecto a β_0 y β_1 , y posteriormente se igualan a cero. Las expresiones calculadas son llamadas ecuaciones de verosimilitud y se representa de la siguiente forma:

$$\sum_{i=1}^{n} [y_i - \pi(x_i)] = 0 \tag{2.3.4}$$

Y

$$\sum_{i=1}^{n} x_i * [y_i - \pi(x_i)] = 0 \tag{2.3.5}$$

Los resultados de β obtenidos por la solución de las expresiones (2.3.4) y (2.3.5) se les denomina estimaciones de máxima verosimilitud y se le denota como β. La estimación de estos valores no es sencilla, para eso se existen distintos métodos numéricos, uno de ellos es el método de Newton Raphson.

2.4 PRUEBA DE HIPÓTESIS PARA LOS PARÁMETROS

Una vez estimado los parámetros del modelo nuestro interés se centra en la evaluación del nivel de importancia que tienen las variables dentro del modelo. Para realizar esta evaluación implica formular test de hipótesis y mediante estos test determinar la relación que existe entre la variable dependientes y las independientes

.

En la Regresión Lineal existe un método generar para determinar la significancia de las variables, el cual puede utilizarse en la Regresión Logística. Este método evalúa la significancia del coeficiente (β_1) para posteriormente estructurar una matriz de varianzas. La matriz estructurara separa la suma de cuadrados respecto a su media en dos componentes: SSE y SSR.

La fórmula del SSE (Suma de Cuadrados respecto a la regresión) se define de la siguiente forma:

$$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad (2.4.1)$$

Partiendo del modelo que no posee la variable respuesta, y trabajando con β_0, y $\widehat{\beta_0} = \bar{y}$ que es equivalente a la media de la variable dependiente, se asume que , $\hat{y}_i = \bar{y}$ así mismo SSE es equivalente a la Varianza Total. Al incorporar la variable independiente dentro del modelo alguna reducción en SSE es producto de que β_0 que acompaña a la variable independiente es distinto de cero. La variación del valor de SSE es causado por la variabilidad de la regresión, que se denota como SSR y se representa de la siguiente forma

$$SSR = [\ \sum_{i=1}^{n}(y_i - \bar{y}_i)^2] - [\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad (2.4.2)$$

En los modelos de Regresión Lineal nos interesa que tan alto puedo ser el valor de SSR, un valor alto nos indica que la variable independiente es significativa dentro del modelo,

mientras que un valor menor nos indica que esta variable no es útil para predecir la variable dependiente.

El principio utilizado en la Regresión Lineal es el mismo que se utiliza para la Regresión Logística, es decir compara los valores reales de la variable dependiente con los valores predecidos que se obtuvieron de los modelos con la variable independiente y si la misma. En los modelos de Regresión Logística esta comparación entre los valores reales y predecidos se define en la función (2.2.6).

La prueba para medir el supuesto de que el β_0 de la variable dependiente es altamente significativo es el prueba de razón de verosimilitud que se presenta en la siguiente formula:

$$D = -2\sum_{i=1}^{n}[\, y_i \ln\left(\frac{\hat{\pi}_i}{y_i}\right) + (1 - y_i)\ln(\frac{1 - \hat{\pi}_i}{1 - y_i})] \qquad (2.4.3)$$

Donde $\hat{\pi}_i = \hat{\pi}(x_i)$.

La estadística, D, en la ecuación (2.4.3) se llama Desvianza, el papel que asume la Desvianza en la Regresión Logística es el mismo que tiene la suma residual de cuadrados en el caso de la Regresión Lineal. Es así que las ecuaciones mostradas en (2.4.3), cuando es calculada para la Regresión Lineal, es equivalente a SSE.
Este valor estadístico puede ser expresado de la siguiente forma:

$$D = -2\ln(Verosimilitud\ del\ modelo\ ajustado \qquad (2.4.4)$$

Para evaluar el nivel de significancia de alguna variable regresora podemos comparar el resultado de D con la variable independiente y sin la misma.
Un incremento o reducción del valor de D ser resultado de la inserción de la variable regresora dentro del modelo, se puede calcular de la siguiente forma:

$$G = D(Modelo\ \text{sin}\ variable) - D\ (Modelo\ con\ la\ variable) \qquad (2.4.5)$$

El valor de G cumple un papel parecido en la regresión logística como en el numerador de la prueba F parcial de la Regresión Lineal. El valor de la estadística G se expresa bajo la siguiente estructura:

$$G = -2\ln\left[\frac{Verosimilitud\ sin\ la\ variable}{Verosimilitud\ con\ la\ variable}\right] \tag{2.4.6}$$

En el caso de una variable independiente es sencillo demostrar que en caso la variable no forme parte del modelo , el estimador de máximo verosimilitud de β_0 es equivalente a $\ln(\frac{n1}{n0})$ donde $n_1 = \sum y_i$, $n_0 = \sum(1 - y_i)$ y el valor predecido es constante, n_1/n. Bajo estos supuestos el valor de G se presenta de la siguiente:

$$G = 2\ln\left[\frac{\left(\frac{n_1}{n}\right)^{n_1}\left(\frac{n_0}{n}\right)^{n_0}}{\prod_{i=1}^{n}\widehat{\pi_i}^{y_i}(1-\hat{\pi}_i)^{(1-y_i)}}\right] \tag{2.4.7}$$

O

$$G = 2\left\{\sum_{i=1}^{n}[y_i\ \ln(\hat{\pi}_i) + (1-y_i)\ln(1-\ \hat{\pi}_i)]\right.$$
$$\left. -\ [n_i\ln(n_i) + n_0\ln(n_0) - nln(n)]\right\} \tag{2.4.8}$$

Bajo el supuesto que β_i es equivalente a cero, el valor del estadístico G va mantener una distribución Chi-cuadrado con solo un grado de libertad.

Otras pruebas que son estadísticamente iguales son la prueba de Wald y la prueba de Score. Los supuestos que se utilizan para estas pruebas son los mismos de la prueba de razón de verosimilitud.

La prueba de Wald calcula contrastando la estimación máxima verosimilitud de $\hat{\beta}_1$ y la estimación de su error estándar. La fracción resultante, asumiendo el supuesto que β_1 es equivalente a cero, posee una distribución normal estándar. El valor del estadístico de Wald para los modelos de Regresión Logística es expresado de la siguiente forma:

$$W = \frac{\hat{\beta}_1}{\widehat{SE}(\hat{\beta}_1)} \tag{2.4.9}$$

W sigue una distribución normal estándar.

El test de Score se basa en la teoría de la distribución de probabilidad del logaritmo de verosimilitud. En general, este es un test multivariante que requiere cálculos matriciales. En el caso univariante, este test se basa en la distribución condicional de la derivada en la ecuación (2.3.5).

En este caso, podemos escribir una expresión para el test de Score. El test utiliza el valor de la ecuación (2.3.5) utilizando $\beta_0 = \ln\left(\frac{n_1}{n_0}\right)$ y β_1=0, bajo estos parámetros, $\hat{\pi} = \frac{n_1}{n_0} = \bar{y}$. Se puede demostrar que la varianza estimada es $\sum_{i=1}^{n} x_i(y_i - \bar{y})$. El estadístico de prueba para el test de Score es:

$$ST = \frac{\sum_{i=1}^{n} x_i(y_i - \bar{y})}{\sqrt{\bar{y}(1-\bar{y})\sum_{i=1}^{n}(x_i - \bar{x})^2}} \qquad (2.4.10)$$

2.5 ESTIMACIÓN POR INTERVALO DE CONFIANZA.

El estimador calculado utilizando intervalos de confianza para el valor β_1 y el intercepto se fundamente en la prueba de Wald. Por lo tanto para β_1 se define su intervalo de confianza al $100\,(1-\alpha)\%$ de la siguiente forma:

$$\hat{\beta}_1 \pm Z_{1-\alpha/2}\widehat{SE}(\hat{\beta}_1) \qquad (2.5.1)$$

Y para el intercepto es

$$\hat{\beta}_0 \pm Z_{1-\alpha/2}\widehat{SE}(\hat{\beta}_0) \qquad (2.5.2)$$

Donde $Z_{1-\alpha/2}$ es el cuantil al $100\,(1-\alpha)\%$ de una distribución de Gauss estándar y $\widehat{SE}(.)$ hace referencia al estimador del error estándar de β_1 dado por el modelo calculado.

La logit es la componente lineal de los modelos de Regresión Logística y su estimación se calcula de siguiente forma:

$$\hat{g}(x) = \hat{\beta}_0 + \hat{\beta}_1 x \qquad (2.5.3)$$

El estimador de la varianza de $\hat{g}(x)$ necesita calcularse a través de la suma. En tal caso es equivalente a:

$$\widehat{VAR}[\,\hat{g}(x)] = \widehat{VAR}[\hat{\beta}_0\,] + x^2\widehat{VAR}[\hat{\beta}_1] + 2x\widehat{COV}[\hat{\beta}_0, \hat{\beta}_1\,] \qquad (2.5.4)$$

En sentido general la varianza de la suma es equivalente a la suma de la varianza de cada término y la covarianza de cada posible valor compuesto por los elementos de la suma del intervalo de confianza al $100\,(1-\alpha)\%$ según la prueba de Wald para la logit es

$$\hat{g}(x) \pm Z_{1-\alpha}\widehat{SE}\big(\hat{g}(x)\big) \qquad (2.5.5)$$

Donde $\widehat{\mathrm{SE}}(\hat{\mathrm{g}}(\mathrm{x}))$es igual a la raíz cuadrada del estimador de la varianza dada en la ecuación (2.5.4). (Hosmer &Lemeshow, 2000).

2.6 PRUEBA DE HOSMER – LEMESHOW

Para validar el ajuste del modelo de Regresión Logística, Hosmer–Lemeshow evalúa que el modelo propuesto puede corroborar lo que se observa. El estadístico que utiliza es el Chi-cuadrado de Pearson el cual se basa en contrastar lo observado y lo esperado.

Hosmer–Lemeshow plantea las siguientes hipótesis:

H_0: El modelo es apropiado.

H_1: El modelo no es apropiado.

La estadística de prueba es:

$$\hat{C} = \sum_{n=1}^{g} \frac{(O_n - n_n\overline{\pi}_n)^2}{\overline{\pi}_n n_n(1-\overline{\pi}_n)} \qquad (2.6.1)$$

Donde $O_s = \sum_{j=1}^{C_n} y_j$ es la cantidad de respuestas entre las covariables, donde n_n define la cantidad de covariables, en el *n*-ésimo decil, C_n el total de sujetos en el n-ésimo grupo y

$$\overline{\pi}_k = \sum_{j=1}^{C_k} \frac{m_j\hat{\pi}_j}{n_j} \qquad (2.6.2)$$

es la probabilidad del valor medio estimada.

El valor del estadístico $\hat{C}$ se aproxima a una distribución Chi-cuadrado con g – 2 grados de libertad, dado por la hipótesis nula. A un grado de significancia α se rechaza Ho si:

$$\hat{C} > X^2_{1-\alpha}\,(g-2) \tag{2.6.3}$$

y se concluye que el modelo no es el apropiado. (Hosmer &Lemeshow, 2000).

3. EL MODELO DE REGRESIÓN LOGÍSTICA MÚLTIPLE

Sea Y que representa la variable respuesta y un conjunto de *t* variables independientes representadas como $X' = (x_1, x_2, \dots\dots, x_t)$. Definimos la probabilidad condicional como $P(Y=1|x) = \pi(x)$, por lo tanto, los modelos de Regresión Logística Múltiple expresan el logit de la siguiente forma:

$$g(x) = \beta_0 + \beta_1 x_1 + \ \dots + \beta_t x_t \tag{3.1}$$

De esta forma el modelo de Regresión Logística es:

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}} \tag{3.2}$$

En el caso que dentro de nuestro modelo estemos introduciendo variables independientes de origen discreto o de escala nominal (sexo, grupo o raza, por ejemplo), no es apropiado acoplarlos como si fueran de escala inteverlar ya que los números que muestran son más identificadores y no representan un valor con significancia cuantitativa

A este tipo de variables se les denomina '*variables dummy*'.

Sea a el caso que la variable de escala nominal posea *n* probables valores, en consecuencia, se necesitara *n-1* '*variables dummy*'. Para mostrar este concepto asumamos que la *k-ésima* variable independiente x_k posee k_k niveles. El grupo de k_{k-1} variables dummy se identifica como D_{jl} y los β_0 para este grupo de variables dummy se expresaran como β_{jl}, *l=1,2,....,* k_{j-1}

de esta forma, la logit para un modelo con *n* variables y con la *j-ésima* variable discreta se define como

$$g(x) = \beta_0 + \beta_1 + \ldots + \sum_{i=1}^{k_j-1} \beta_{jl} D_{jl} + \beta_n x_n \qquad (3.3)$$

3.1 AJUSTE DEL MODELO DE REGRESIÓN LOGÍSTICA MÚLTIPLE

Asumamos que se tiene un conjunto de n observaciones independientes (x_i, y_i), donde $i=1,2,\ldots,n$. Al igual que en el caso univariante, un ajuste del modelo demanda que calculemos los estimadores del vector $\beta' = (\beta_0, \beta_1, \ldots, \beta_p)$. En el caso de lo de los modelos univariantes la estimación de estos parámetros se realiza mediante la máxima verosimilitud, para el caso multivariante utilizaremos la misma técnica.

La función de verosimilitud es casi similar a la presentada en la ecuación (2.3.2) la diferencia radica en que $\pi(x)$ se define como en la ecuación (3.2).

Las funciones de verosimilitud resultantes pueden se expresan de la siguiente forma:

$$\sum_{i=1}^{n} [y_j - \pi(x_i)] \qquad (3.1.1)$$

Y

$$\sum_{i=1}^{n} x_{ij} [y_j - \pi(x_i)] \qquad (3.1.2)$$

Definamos que $\hat{\beta}$ que indica la solución de estas funciones. De esta forma, los valores ajustados para el modelo de Regresión Logística Múltiple será $\hat{\pi}(x)$, equivalente al valor de la ecuación (3.2) calculado usando $\hat{\beta}$ y x_i

La derivada parcial se expresar la siguiente forma general:

$$\frac{\partial^2 l(\beta)}{\partial \beta_j^2} = -\sum_{i=1}^{n} x_{ij}^2 \pi_i (1-\pi_i) \quad (3.1.3)$$

Y

$$\frac{\partial^2 l(\beta)}{\partial \beta_j^2 \partial \beta_l} = -\sum_{i=1}^{n} x_{ij}\, x_{jl} \pi_i (1-\pi_i) \quad (3.1.4)$$

Para j y l que pueden tomar valores desde 0 hasta p y donde π_i equivale $\pi(x_i)$. Puesto que la matriz [p+1] *[p+1] que posee los negativos de los términos calculados en las ecuaciones (3.1.3) y (3.1.4) esto es denotado como I(β). A esta matriz se le denomina "matriz de información observada". De esta forma tanto la varianza como la covarianza de los coeficientes estimados se obtendrán calculando la inversa de la matriz y se denotara como Var(β)=I-1(β).

Utilizaremos la notación Var(β_j) para indicar el j-ésimo componente de la diagonal de la matriz, que es equivalente a la varianza de $\widehat{\beta_j}$ y Cov(β_j, β_l) para precisar un componente arbitrario de la diagonal, que es igual a la covarianza de $\widehat{\beta_j}$ y $\widehat{\beta_l}$.

Los estimadores tanto de la varianza y la covarianza se denotara por $\widehat{\mathrm{Var}}(\widehat{\beta_j})$ y $\widehat{\mathrm{Cov}}(\widehat{\beta_j}, \widehat{\beta_l})$ respectivamente, donde j y l que pueden tomar valores desde 0 hasta p y para precisar los valores de esta matriz utilizaremos las expresiones $\widehat{\mathrm{Var}}(\beta_j)$ y $\widehat{\mathrm{Cov}}(\beta_j, \beta_l)$,
j, l=1, 2…p para precisar que valores puede tomar esta matriz. Posteriormente se tendrá la oportunidad de utilizar únicamente la estimación del error estándar de los coeficientes calculados, que se denotará como:

$$SE = [\widehat{Var}(\widehat{\beta_j})]^{1/2} \qquad j = 0, 1, 2, \ldots, p \quad (3.1.5)$$

Se utilizará esta representación en el desarrollo del test de coeficientes y la estimación utilizando intervalos de confianza.

3.2 PRUEBA PARA LA SIGNIFICANCIA DEL MODELO

Al igual que en el caso univariante, una vez que tenemos el modelo ajustado, proseguiremos con la evaluación de la significancia de las variables dentro del modelo. el test de máxima verosimilitud que se utiliza en le Regresión Logística Univariada se utilizada de manera similar para el caso de la Regresión Logistica Multiple.

El test se fundamenta en el cálculo del estadístico *G dado* en la ecuación, la diferencia radica en que los valores ajustados de $\hat{\pi}$ bajo el modelo de Regresión Logistica Mutiple se fundamenta sobre el vector que posee *p+1* parámetros, $\hat{\beta}$, , bajo H_0 = hipótesis nula, que indica que los "p" valores de los coeficientes de las covariables en el modelo son equivalente a cero, *G* tendrá una distribución Chi- Cuadrado con p grados de libertad.

El símil multivariable de la prueba de Wald es calcular el valor del siguiente vector matriz.

$$W = \hat{\beta}'[\widehat{Var}(\hat{\beta}_j)]^{-1}\hat{\beta} \tag{3.2.1}$$

$$W = \hat{\beta}'[\mathrm{XVX}]^{-1}\hat{\beta} \tag{3.2.2}$$

Que bajo el supuesto que cada uno de los *p+1* coeficientes es equivalente a cero esta va poseer una distribución Chi-cuadrado con *p+1* grados de libertad. (Hosmer &Lemeshow, 2000).

3.3 ESTIMACIÓN POR INTERVALOS DE CONFIANZA

En la sección (2) se revisó como se estiman los parámetros utilizando intervalos de confianza en el caso de la Regresión Logística Simple. Para la Regresión Logística Múltiple la estimación es de manera muy similar.

A partir de las ecuaciones (2.5.1) y (2.5.2) podemos obtener un intervalo de confianza al *100(1-α)%* para los coeficientes de la Regresión Logística Múltiple.

La estimación de la logit por intervalos de confianza es ligeramente más complicada en el caso los modelos con múltiples variables que el resultado obtenido en la ecuación (2.5.5). La idea básica es similar, sin embargo, ahora se tiene más términos implicados en la suma. Partiendo de la ecuación (3.1) una forma general para calcular la estimación del logit que contiene *p* covariantes es:

$$\hat{g}(X) = \widehat{\beta_0} + \widehat{\beta_1}x_1 + \quad \ldots . + \widehat{\beta_p}x_p \tag{3.3.1}$$

Una expresión alternativa para mostrar la estimación de la logit dada en la ecuación (3.3.1) es la utilización de la notación vectorial como $\hat{g}(X) = X'\hat{\beta}$, donde el vector

$$\widehat{\beta'}_0 = (\widehat{\beta_0}, \widehat{\beta_1}, \widehat{\beta_2}, \ldots \ldots, \widehat{\beta_p}) \tag{3.3.2}$$

muestra la estimación de los (p + 1) coeficientes, así mismo $X = (x_0, x_1, x_2, \ldots x_p)$ representa el valor de la constante y el conjunto de valores de los p–covariantes en el modelo donde $x_0 = 1$

Partiendo de la ecuación (2.5.4) donde una forma para expresar la estimación de la varianza del estimador de la logit presentada en la ecuación (3.3.1) *e s* :

$$\widehat{Var}[\hat{g}(X)] = \sum_{j=0}^{p} x_j^2 \, \widehat{Var}[\hat{\beta}_j] \; + \sum_{j=0}^{p} \sum_{k=j+1}^{p} 2x_j \, x_k \widehat{cov}(\widehat{\beta_j}, \widehat{\beta_k}) \qquad (3.3.3)$$

La ecuación anterior se puede representar utilizando matrices

$$\widehat{Var}(\widehat{\beta}) = (\mathrm{X'VX})^{-1} \qquad (3.3.4)$$

Continuando de la ecuación (3.3.3) una forma similar para estimar la ecuación dada en el punto (3.3.4) es:

$$\widehat{Var}[\hat{g}(x)] = x'^{\widehat{Var}}(\hat{\beta})x \qquad (3.3.5)$$

$$\widehat{Var}[\hat{g}(x)] = x'^{(\mathrm{X'VX})^{-1}}x \qquad (3.3.6)$$

Por fortuna, los programas que contiene la Regresión Logística brindan al usuario la opción de poder crear nuevas variables que contienen los valores estimados de la ecuación (3.3.6). (Hosmer &Lemeshow, 2000).

CAPÍTULO III

ANÁLISIS DE DATOS

3. ANÁLISIS DE DATOS

3.1 ANÁLISIS EXPLORATORIO DE DATOS

Previamente a desarrollar el análisis de los datos mediante la Regresión Logística, revisaremos cada una de las variables mediante resúmenes cuantitativos y gráficos estadísticos con el objetivo de identificar patrones que respondan alguna interrogante de interés.

Para observar la relación entre variables cualitativas, se puede optar por un análisis bidimensional y utilizar una tabla de doble entrada donde a cada división la corresponde su frecuencia conjunta que representa el número de datos que forman parte de la modalidad i-ésima de la primera variable y a la modalidad j-ésima de la segunda. Con esta tabla resumen podemos revisar si las dos variables son o no independientes.

Para determinar la dependencia de las variables se optará por utilizar "el test de independencia de Chi-Cuadrado de Pearson" que contrasta la hipótesis de que las variables son independientes, frente a la hipótesis alternativa de que una variable se distribuye de modo diferente para diversos niveles de la otra.

3.2 ANÁLISIS EXPLORATORIO DE LAS VARIABLES CUALITATIVAS CON RESPECTO AL RESULTADO FINAL DEL CURSO DE MATEMÁTICA BÁSICA

La revisión de las variables se realizará a través de las tablas de doble entrada o también denominadas tablas de contingencia donde se demostrará el grado de relación entre la variable dependiente y las variables regresoras con sus respectivos test de Chi-cuadrado

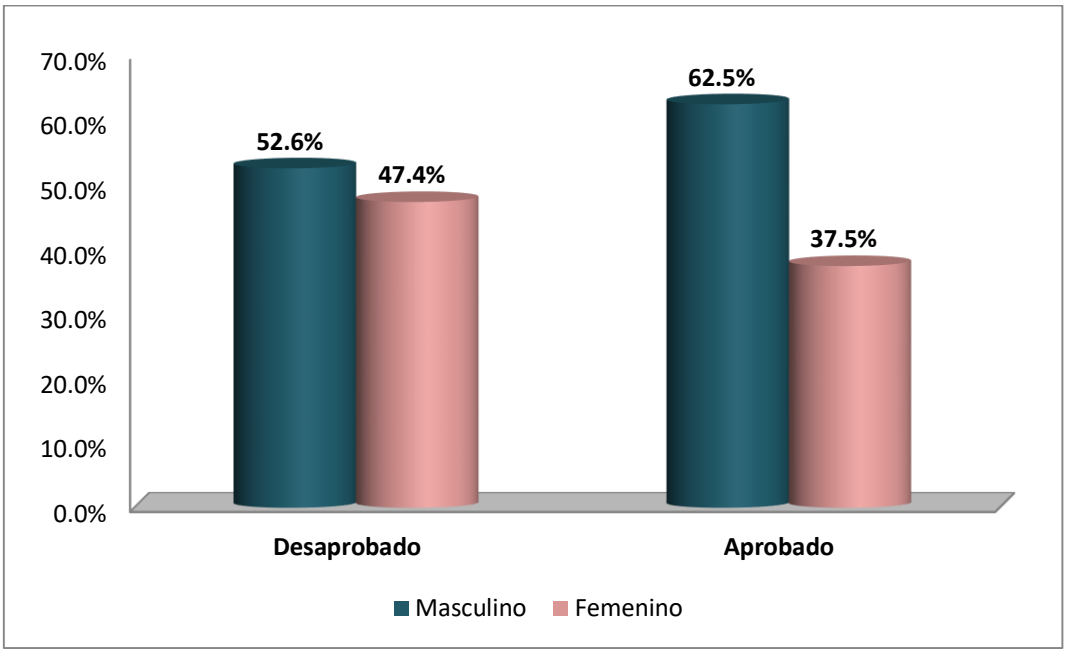
Cuadro 3.2.1

Resultado Final en Matemática Básica y Sexo del alumno

			Género		Total
			Masculino	Femenino	
Resultado final Matemática Básica	Desaprobado	Recuento	30	27	57
		% dentro de Resultado final Matemática Básica	52,6%	47,4%	100,0%
	Aprobado	Recuento	30	18	48
		% dentro de Resultado final Matemática Básica	62,5%	37,5%	100,0%
Total		Recuento	60	45	105
		% dentro de Resultado Final Matemática Básica	57,1%	42,9%	100,0%

Gráfico 3.2.1

Resultado Final en Matemática Básica y Sexo del alumno



En la tabla resumen (3.2.1) se contempla que de los estudiantes que aprueban la asignatura de Matemática Básica el 62.5% son de género masculino, así mismo el 37.5% son de género femenino.

Cuadro 3.2.2

Test de Chi cuadrado

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	1,036[a]	1	,309		
Corrección por continuidad[b]	,672	1	,412		
Razón de verosimilitudes	1,040	1	,308		
Estadístico exacto de Fisher				,329	,206
Asociación lineal por lineal	1,026	1	,311		
N de casos válidos	105				

El "test" Chi-cuadrado (cuadro 3.2.2) para ver la dependencia de las variables, observamos que se obtiene un p_*valor* igual a 0.309 lo cual nos indica que las variables "Resultado final del curso de Matemática Básica y "Sexo del alumno" no están relacionadas con un nivel de significancia del 5%.

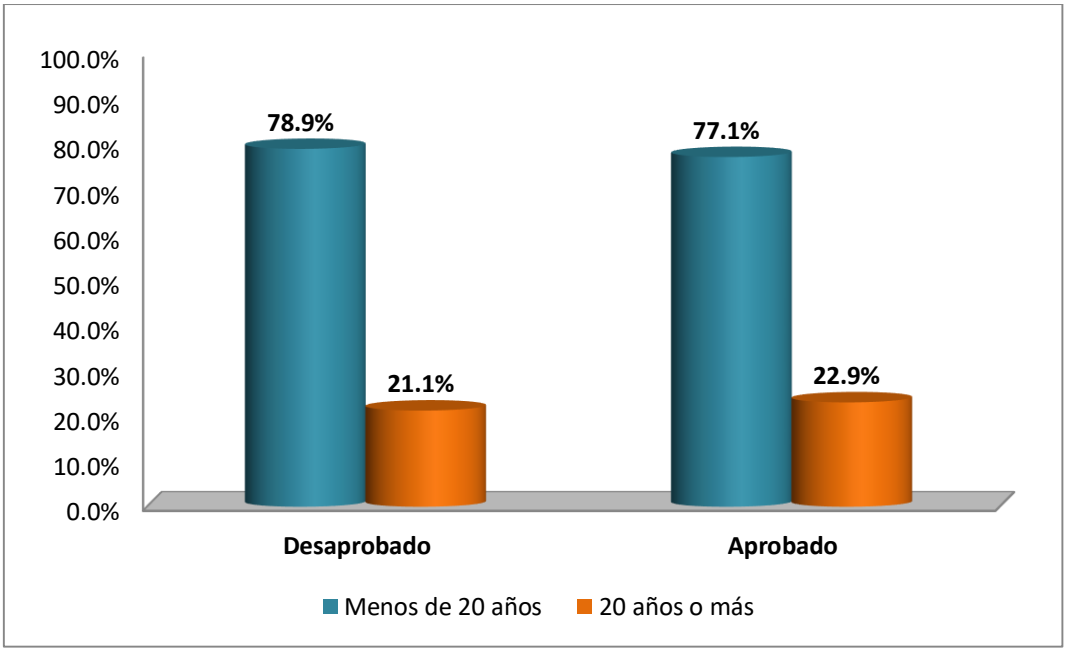
Cuadro 3.2.3

Resultado Final en Matemática Básica y Grupo Etario

			Grupo Etario		Total
			Menos de 20 años	20 años o más	
Resultado final Matemática Básica	Desaprobado	Recuento	45	12	67
		% dentro de Resultado final Matemática Básica	78,9%	21,1%	100,0%
	Aprobado	Recuento	37	11	48
		% dentro de Resultado final Matemática Básica	77,1%	22,9%	100,0%
Total		Recuento	82	23	105
		% dentro de Resultado final Matemática Básica	78,1%	21,9%	100,0%

Gráfico 3.2.2

Resultado Final en Matemática Básica y Grupo Etario



En la tabla resumen (3.2.3) se contempla que de los estudiantes que reprueban la asignatura de Matemática Básica el 21.1% tienen edad mayor a 20 años, este porcentaje se incrementa a un 22.9% en el conjunto de los aprobados.

Cuadro 3.2.4

Test de Chi cuadrado

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	,053a	1	,818		
Corrección por continuidad[b]	,000	1	1,000		
Razón de verosimilitudes	,053	1	,818		
Estadístico exacto de Fisher				,818	,501
Asociación lineal por lineal	,052	1	,819		
N de casos válidos	105				

El "test" Chi-cuadrado (cuadro 3.2.4) para ver la dependencia de las variables, observamos que se obtiene un valor de 0.818, esto nos indica que las variables "Resultado final del curso de Matemática Básica" y "Grupo Etario" no están relacionadas con un nivel de significancia del 5%.

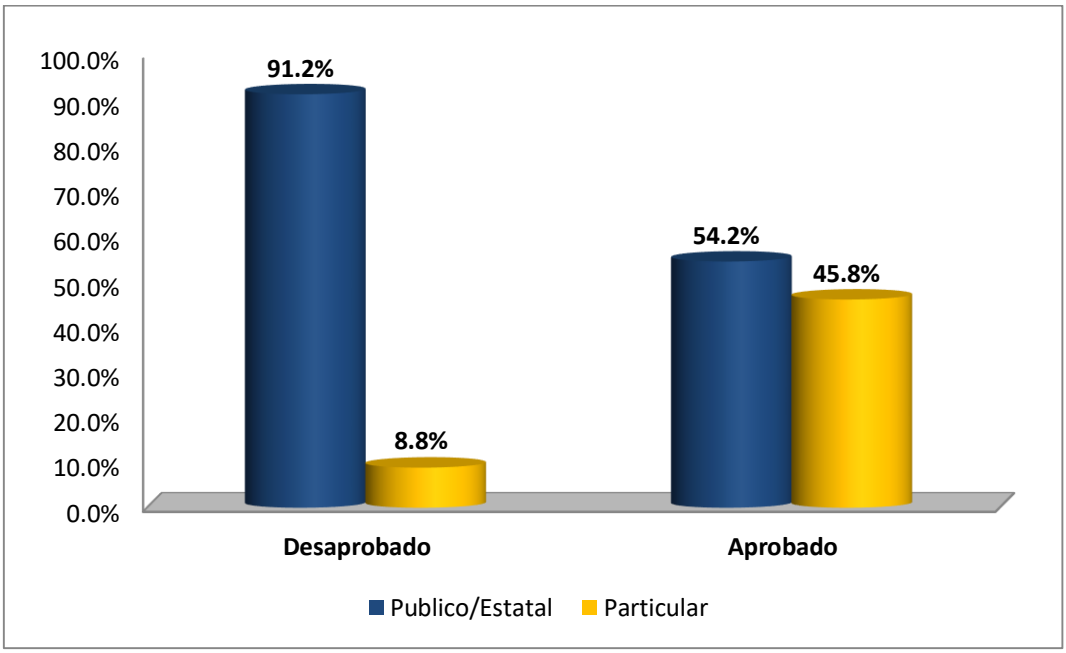
Cuadro 3.2.5

Resultado Final en Matemática Básica y Tipo de Colegio

			Tipo de colegio		Total
			Publico/Estatal	Particular	
Resultado final Matemática Básica	Desaprobado	Recuento	52	7	59
		% dentro de Resultado final Matemática Básica	91,2%	8,8%	100,0%
	Aprobado	Recuento	26	22	48
		% dentro de Resultado final Matemática Básica	54,2%	45,8%	100,0%
Total		Recuento	78	27	105
		% dentro de Resultado final Matemática Básica	74,3%	25,7%	100,0%

Gráfico 3.2.3

Resultado Final en Matemática Básica y Tipo de Colegio



En la tabla resumen (3.2.5), se destaca que del conjunto de estudiantes que reprueban la asignatura de Matemática Básica el 8.8 % proviene de un colegio Particular, este porcentaje se incrementa para este mismo conjunto en un 45.8% en el caso de los aprobados.

Se puede visualizar ampliamente la diferencia con los alumnos provenientes de un colegio Nacional que en gran porcentaje (91.2%) reprueban la asignatura.

Cuadro 3.2.6

Test de Chi cuadrado

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	18,737[a]	1	,000		
Corrección por continuidad[b]	16,847	1	,000		
Razón de verosimilitudes	19,617	1	,000		
Estadístico exacto de Fisher				,000	,000
Asociación lineal por lineal	18,558	1	,000		
N de casos válidos	105				

En el cuadrado (cuadro 3.2.6) para ver la dependencia de las variables se obtiene un *p_valor* de 0.00, quiere decir que las variables "Resultado final del curso de Matemática Básica" y "Tipo de colegio" están relacionadas con un nivel de significancia del 5%.

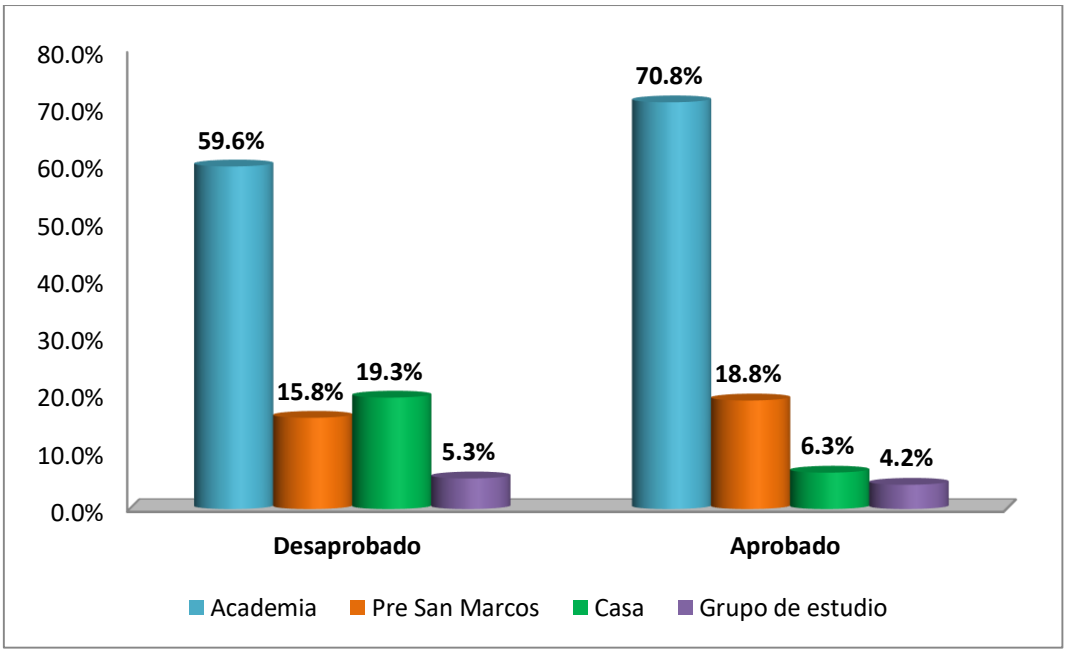
Cuadro 3.2.7

Resultado Final en Matemática Básica y Lugar de Preparación

			Lugar de preparación				Total
			Academia	Pre San Marcos	Casa	Grupo de estudio	
Resultado final Matemática Básica	Desaprobado	Recuento	34	9	11	3	57
		% dentro de Resultado final Matemática Básica	59,6%	15,8%	19,3%	5,3%	100,0%
	Aprobado	Recuento	34	9	3	2	48
		% dentro de Resultado final Matemática Básica	70,8%	18,8%	6,3%	4,2%	100,0%
Total		Recuento	68	18	14	5	105
		% dentro de Resultado final Matemática Básica	64,8%	17,1%	13,3%	4,8%	100,0%

Gráfico 3.2.4

Resultado Final en Matemática Básica y Lugar de Preparación



En la tabla resumen (3.2.7) se visualiza que de los estudiantes que reprueban la asignatura de Matemática Básica el 15.8% realizo su preparación preuniversitaria en el centro Pre-San Marcos, así mismo en el conjunto de los aprobados el porcentaje se incrementa a un 18.8%.

Cuadro 3.2.8

Test de Chi cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	4,030[a]	3	,258
Razón de verosimilitudes	4,289	3	,232
Asociación lineal por lineal	2,265	1	,132
N de casos válidos	105		

El "test" Chi-cuadrado (cuadro 3.2.8) para ver la dependencia de las variables, se observa que el *p_valor* es igual a 0.258, lo cual nos indica que las variables "Resultado final del curso de Matemática Básica" y "Lugar de preparación" no están relacionadas con un nivel de significancia del 5%.

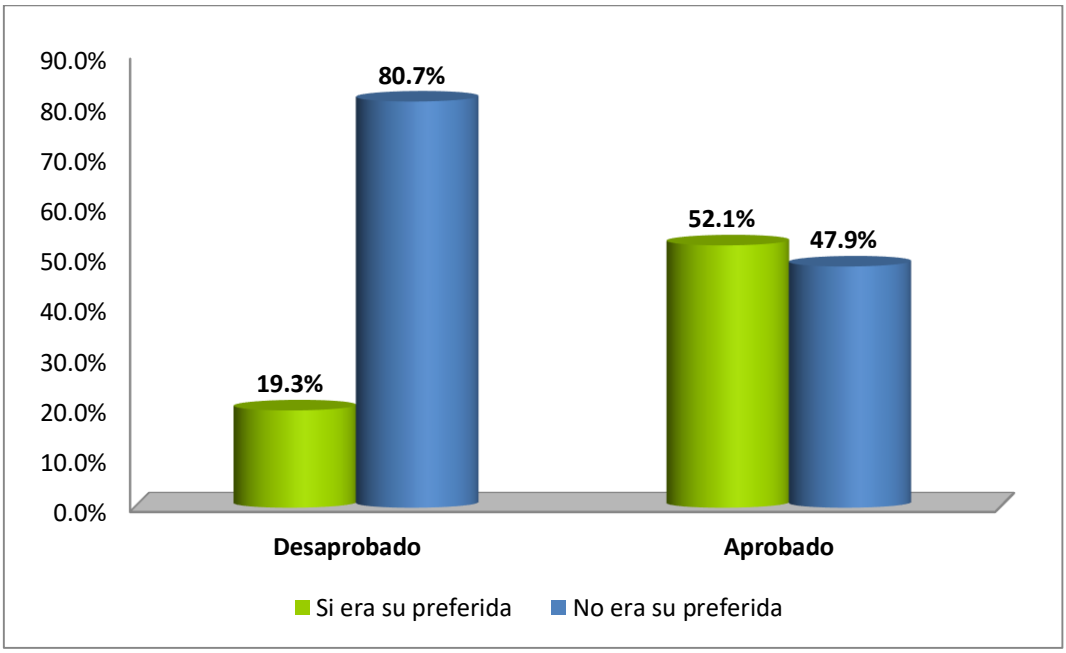
Cuadro 3.2.9

Resultado Final en Matemática Básica y Preferencia por la Carrera

			Preferencia por la carrera		Total
			Si era su preferida	No era su preferida	
Resultado final Matemática Básica	Desaprobado	Recuento	11	46	57
		% dentro de Resultado final Matemática Básica	19,3%	80,7%	100,0%
	Aprobado	Recuento	25	23	48
		% dentro de Resultado final Matemática Básica	52,1%	47,9%	100,0%
Total		Recuento	36	69	105
		% dentro de Resultado final Matemática Básica	34,3%	65.7%	100,0%

Gráfico 3.2.5

Resultado Final en Matemática Básica y Preferencia por la Carrera



En la tabla resumen (3.2.9) se destaca que de los estudiantes que no reprobaron la asignatura de Matemática Básica el 80.7% la carrera que estudia en la actualidad no era su preferida, por otro lado, del conjunto de estudiantes que aprobaron el 52.1% carrera que estudia actualmente si era de su preferencia.

Cuadro 3.2.10

Test de Chi cuadrado

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	12,431[a]	1	,000		
Corrección por continuidad[b]	11,018	1	,001		
Razón de verosimilitudes	12,634	1	,000		
Estadístico exacto de Fisher				,001	,000
Asociación lineal por lineal	12,313	1	,000		
N de casos válidos	105				

El "test" Chi-cuadrado (cuadro 3.2.10) para estudiar la dependencia de las variables, se observa que el *p_valor* es igual a 0.000, esto quiere decir que las variables "Resultado final del curso de Matemática Básica" y "Preferencia por la carrera que estudia" están relacionadas con un nivel de significancia del 5%.

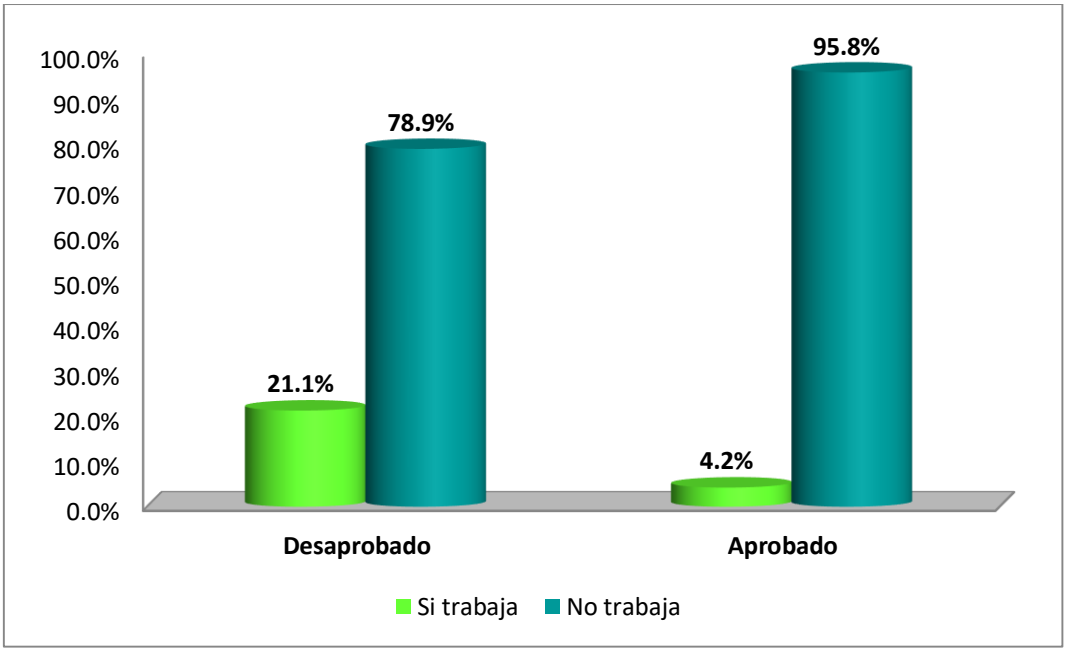
Cuadro 3.2.11

Resultado Final en Matemática Básica y Condición Laboral del Alumno

			Condición Laboral		Total
			Si trabaja	No trabaja	
Resultado final Matemática Básica	Desaprobado	Recuento	12	45	57
		% dentro de Resultado final Matemática Básica	21,1%	78,9%	100,0%
	Aprobado	Recuento	2	46	48
		% dentro de Resultado final Matemática Básica	4,2%	95,8%	100,0%
Total		Recuento	14	91	105
		% dentro de Resultado final Matemática Básica	13,3%	86,7%	100,0%

Gráfico 3.2.6

Resultado Final en Matemática Básica y Condición Laboral del Alumno



En la tabla resumen (3.2.11) se visualiza que del conjunto de alumnos que aprueban la asignatura de Matemática Básica el 95.8% no tenían un trabajo paralelo a sus estudios universitarios. Del conjunto de alumnos que reprueban la asignatura de Matemática Básica el 21.1% si tenían un trabajo paralelo a sus estudios universitarios.

Cuadro 3.2.12

Test de Chi cuadrado

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	6,430[a]	1	,011		
Corrección por continuidad[b]	5,051	1	,025		
Razón de verosimilitudes	7,163	1	,007		
Estadístico exacto de Fisher				,019	,010
Asociación lineal por lineal	6,368	1	,012		
N de casos válidos	105				

El "test" Chi-cuadrado (cuadro 3.2.12) para estudiar la dependencia de las variables, se observa que el *p_valor* es igual a 0.011, esto nos indica que las variables "Resultado final

del curso de Matemática Básica" y "Condición Laboral del Alumno" están relacionadas a un nivel de significancia del 5%.

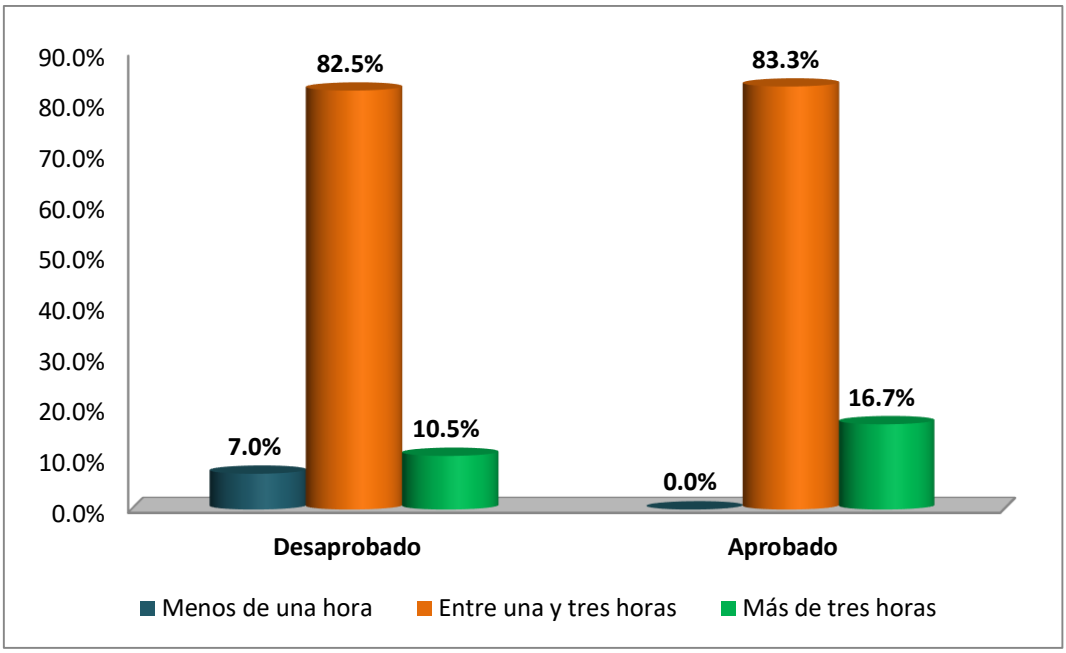
Cuadro 3.2.13

Resultado Final en Matemática Básica y Tiempo de estudio

			Tiempo de estudio			Total
			Menos de una hora	Entre una y tres horas	Más de 3 horas	
Resultado final Matemática Básica	Desaprobado	Recuento	4	47	6	57
		% dentro de Resultado final Matemática Básica	7,0%	82,5%	10,5%	100,0%
	Aprobado	Recuento	0	40	8	48
		% dentro de Resultado final Matemática Básica	,0%	83,3%	16,7%	100,0%
Total		Recuento	4	87	14	105
		% dentro de Resultado Matemática Básica	3,8%	82,9%	13,3%	100,0%

Gráfico 3.2.7

Resultado Final en Matemática Básica y Tiempo de estudio



En la tabla resumen (3.2.13) se visualiza que de los estudiantes que reprueban la asignatura de Matemática Básica el 10.5% posteriormente a las clases estudia tres horas o más, este

porcentaje se incrementa en un 16,7% en el conjunto de los que aprueban la asignatura de Matemática Básica.

Cuadro 3.2.14

Test de Chi cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	4,108[a]	2	,128
Razón de verosimilitudes	5,623	2	,060
Asociación lineal por lineal	2,752	1	,097
N de casos válidos	105		

El "test" Chi-cuadrado (cuadro 3.2.14) para ver la dependencia de las variables, se observa que el *p_valor* es igual a 0.128, este valor nos indica que las variables "Resultado final del curso de Matemática Básica" y "Tiempo de estudio" no están relacionadas, a un nivel de significación del 5%.

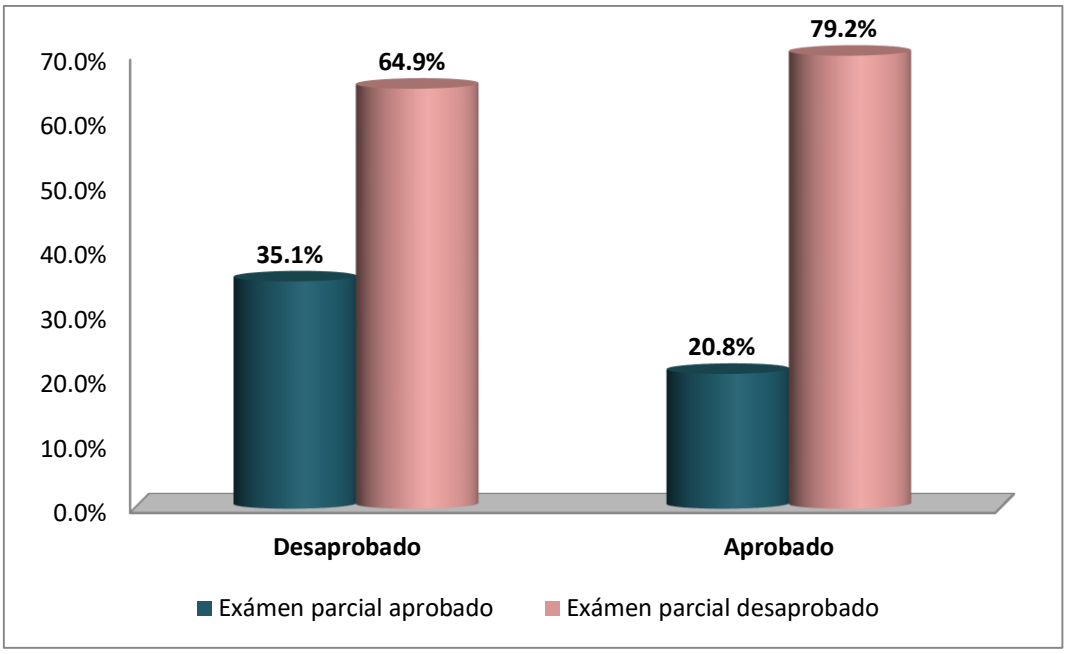
Cuadro 3.2.15

Resultado Final en Matemática Básica y Resultado del Examen Parcial de Matemática Básica

			Resultado del examen parcial de Matemática Básica		Total
			Aprobado	Desaprobado	
Resultado final Matemática Básica	Desaprobado	Recuento	20	37	57
		% dentro de Resultado final Matemática Básica	35,1%	64,9%	100,0%
	Aprobado	Recuento	10	38	48
		% dentro de Resultado final Matemática Básica	20,8%	79,2%	100,0%
Total		Recuento	30	75	105
		% dentro de Resultado final Matemática Básica	28,6%	71,4%	100,0%

Gráfico 3.2.8

Resultado Final en Matemática Básica y Resultado del Examen Parcial de Matemática Básica



En la tabla resumen (3.2.15) se visualiza que del conjunto de estudiantes que reprueban la asignatura de Matemática Básica el 64.9% tuvieron en su examen parcial nota desaprobatoria, esta cantidad se incrementa a un 79.2% para el conjunto de estudiantes que si aprobaron la asignatura de Matemática Básica.

Cuadro 3.2.16

Test de Chi cuadrado

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	2,594a	1	,107		
Corrección por continuidad[b]	1,943	1	,163		
Razón de verosimilitudes	2,639	1	,104		
Estadístico exacto de Fisher				,131	,081
Asociación lineal por lineal	2,570	1	,109		
N de casos válidos	105				

El "test" Chi-cuadrado (cuadro 3.2.16) para ver la dependencia de las variables, se observa que el *p_valor* es igual a 0.107, este valor nos indica que las variables "Resultado final del curso de Matemática Básica" y "Resultado del Examen Parcial de Matemática Básica" no están relacionadas, a un nivel de significación del 5%.

.

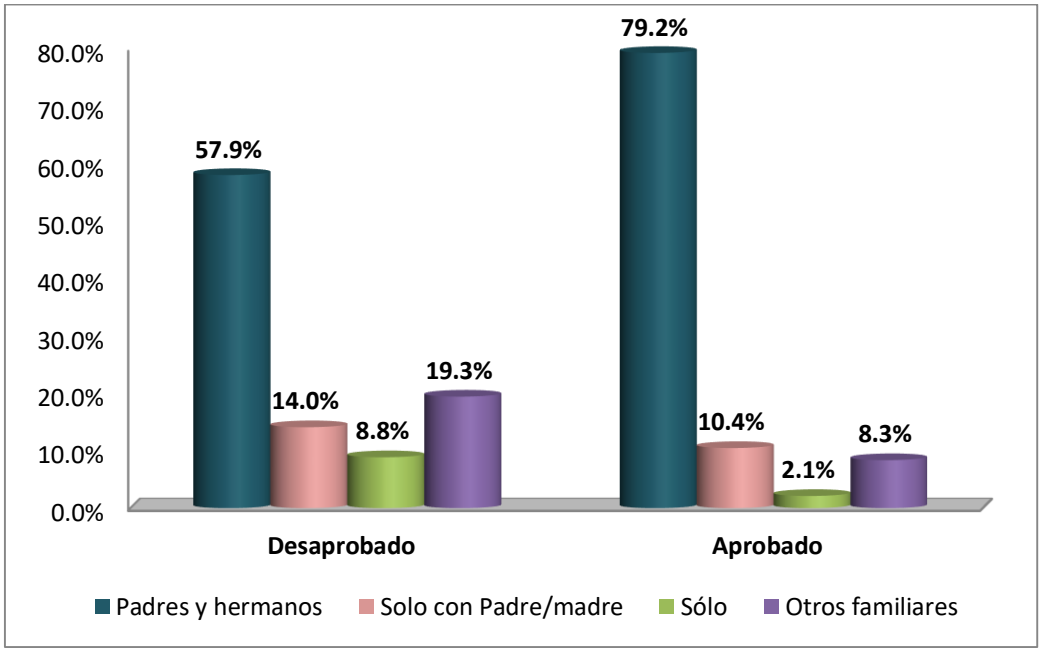
Cuadro 3.2.17

Resultado Final en Matemática Básica y Personas con quien convive el alumno

			Personas con quién vive el alumno				Total
			Padres, hermanos	Solo con Padre/madre	Sólo	Otros familiares	
Resultado final Matemática Básica	Desaprobado	Recuento	33	8	5	11	57
		% dentro de Resultado final Matemática Básica	57,9%	14,0%	8,8%	19,3%	100,0%
	Aprobado	Recuento	38	5	1	4	48
		% dentro de Resultado final Matemática Básica	79,2%	10,4%	2,1%	8,3%	100,0%
Total		Recuento	71	13	6	15	105
		% dentro de Resultado final Matemática Básica	67.6%	12,4%	5,7%	14,3%	100,0%

Gráfico 3.2.9

Resultado Final en Matemática Básica y Personas con quien vive el alumno



En la tabla resumen (3.2.17) se visualiza que de los estudiantes que aprueban la asignatura de Matemática Básica el 79.2% conviven con sus padres y hermanos. De los estudiantes que reprobaron la asignatura el 14.0% convive únicamente con uno de sus padres.

Cuadro 3.2.18

Test de Chi cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	6,252[a]	3	,100
Razón de verosimilitudes	6,587	3	,086
Asociación lineal por lineal	5,382	1	,020
N de casos válidos	105		

El "test" Chi-cuadrado (cuadro 3.2.18) para ver la dependencia de las variables, se observa que el *p_valor* es igual a 0.100, este valor nos indica que las variables "Resultado final del curso de Matemática Básica" y "Personas con quien vive el alumno" no están relacionadas, a un nivel de significación del 5%.

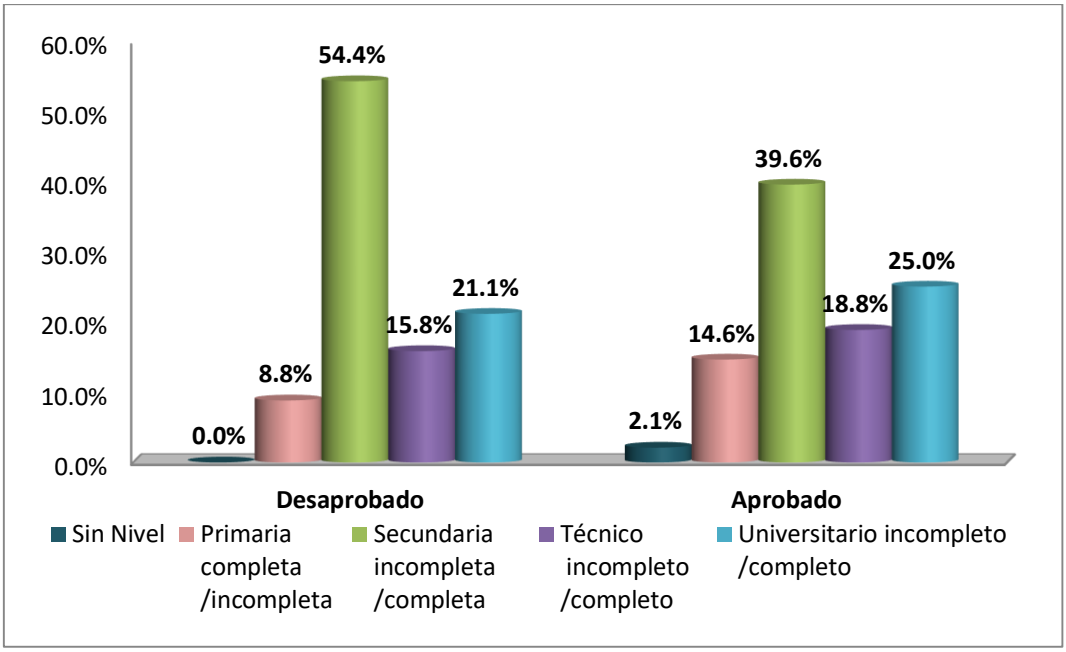
Cuadro 3.2.19

Resultado Final en Matemática Básica y Nivel Académico del Padre

			Nivel académico del padre					Total
			Sin Nivel	Primaria Completa /incompleta	Secundaria Completa/ Incompleta	Técnico Completo/ Incompleto	Universitario Completo/ incompleto	
Resultado final Matemática Básica	Desaprobado	Recuento	0	5	31	9	12	57
		% dentro de Resultado final Matemática Básica	,0%	8,8%	54,4%	15,8%	21,1%	100,0%
	Aprobado	Recuento	1	7	19	9	12	48
		% dentro de Resultado final Matemática Básica	2,1%	14,6%	39,6%	18,8%	25,0%	100,0%
Total		Recuento	1	12	50	18	24	111
		% dentro de Resultado final Matemática Básica	1,0%	11,4%	47,6%	17,1%	22,9%	100,0%

Gráfico 3.2.10

Resultado Final en Matemática Básica y Nivel Académico del Padre



En la tabla resumen (3.2.19) se visualiza que el 25,0% de los estudiantes que aprueban la asignatura de Matemática Básica, sus padres cursaron estudios universitarios completa /incompleta. El 54,4% del conjunto de estudiantes que reprobaron la asignatura de Matemática Básica, sus padres cursaron hasta el nivel de Secundaria.

Cuadro 3.2.20

Test de Chi cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	3,467[a]	4	,483
Razón de verosimilitudes	3,857	4	,426
Asociación lineal por lineal	,002	1	,964
N de casos válidos	105		

El "test" Chi-cuadrado (cuadro 3.2.20) para estudiar la dependencia de las variables se obtiene un *p_valor* de 0.483, quiere decir que las variables "Resultado final del curso de

Matemática Básica" y "Nivel académico del padre" son independientes, a un nivel de significación del 5%.

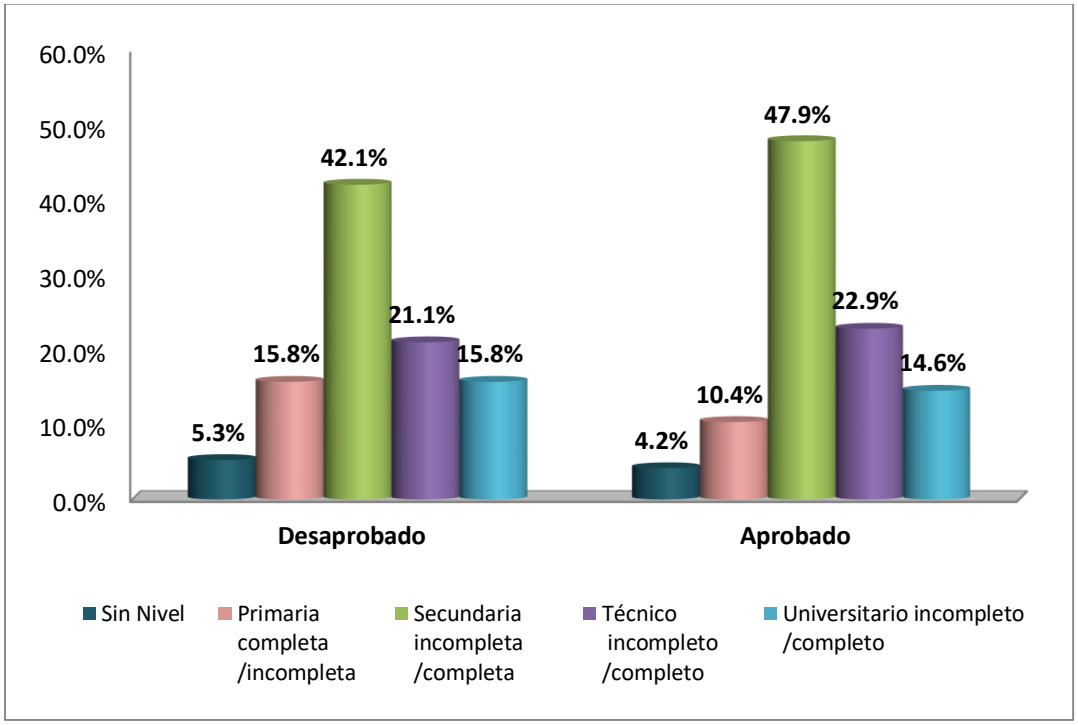
Cuadro 3.2.21

Resultado Final en Matemática Básica y nivel académico de la madre

			Nivel académico de la madre					Total
			Sin Nivel	Primaria completa /incompleta	Secundaria completa /incompleta	Técnico completa /incompleta	Universitario completa /incompleta	
Resultado final Matemática Básica	Desaprobado	Recuento	3	9	24	12	9	57
		% dentro de Resultado final Matemática Básica	5,3%	15,8%	42,1%	21,1%	15,8%	100,0%
	Aprobado	Recuento	2	5	23	11	7	48
		% dentro de Resultado final Matemática Básica	4,2%	10,4%	47,9%	22,9%	14,6%	100,0%
Total		Recuento	5	14	47	23	16	105
		% dentro de Resultado final Matemática Básica	4,8%	13,3%	44,8%	21,9%	15,2%	100,0%

Gráfico 3.2.11

Resultado Final en Matemática Básica y nivel académico de la madre



En la tabla resumen (3.2.21) se visualiza que el 21,1% del conjunto de estudiantes que reprueban la asignatura de Matemática Básica, sus madres estudiaron una carrera Técnica Completa/Incompleta, por otro lado, el 47.9% del conjunto de estudiantes que aprueban la asignatura de Matemática Básica sus madres alcanzaron el grado de Secundaria.

Cuadro 3.2.22

Test de Chi cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	,893[a]	4	,926
Razón de verosimilitudes	,903	4	,924
Asociación lineal por lineal	,119	1	,730
N de casos válidos	105		

El "test" Chi-cuadrado (cuadro 3.2.22) para ver la dependencia de las variables se observa que el *p_valor* es igual a 0.926, este valor nos indica que las variables "Resultado final del curso de Matemática Básica" y "Nivel académico de la madre" son independientes, a un nivel de significación del 5%.

3.3 ANÁLISIS EXPLORATORIO DE LAS VARIABLES CUALITATIVAS CON RESPECTO AL RESULTADO FINAL DEL CURSO DE CÁLCULO I

La revisión de las variables se realizará a través de las tablas de doble entrada o también denominadas tablas de contingencia donde se demostrará el grado de relación entre la variable dependiente y las variables regresoras con sus respectivos test de Chi-cuadrado

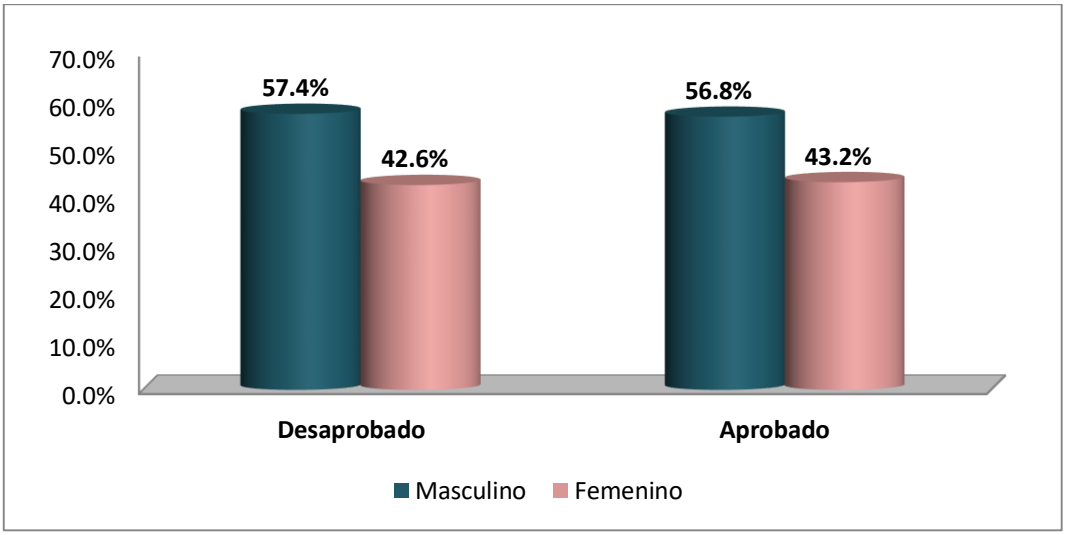
Cuadro 3.3.1

Resultado Final en Cálculo I y Sexo del alumno

			Género		Total
			Masculino	Femenino	
Resultado final Cálculo I	Desaprobado	Recuento	39	29	68
		% dentro de Resultado final Cálculo I	57,4%	42,6%	100,0%
	Aprobado	Recuento	21	16	37
		% dentro de Resultado final Cálculo I	56,8%	43,2%	100,0%
Total		Recuento	60	45	105
		% dentro de Resultado Final Cálculo I	57,1%	42,9%	100,0%

Gráfico 3.3.1

Resultado Final en Cálculo I y Sexo del alumno



En la tabla resumen (3.3.1) se contempla que de los estudiantes que aprueban la asignatura de Cálculo I el 56.8% son de género masculino, así mismo el 43.2% son de género femenino.

Cuadro 3.3.2

Test de Chi cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	**,003[a]**	**1**	**,953**
Corrección por continuidad[b]	,000	1	1,000
Razón de verosimilitudes	,003	1	,953
Estadístico exacto de Fisher			
Asociación lineal por lineal	,003	1	,953
N de casos válidos	105		

El "test" Chi-cuadrado (cuadro 3.3.2) para estudiar la dependencia de las variables, se observa que el *p_valor* es igual a 0.953, este valor nos indica que las variables "Resultado final del curso de Cálculo I" y "Sexo del alumno" no están relacionadas, a un nivel de significación del 5%.

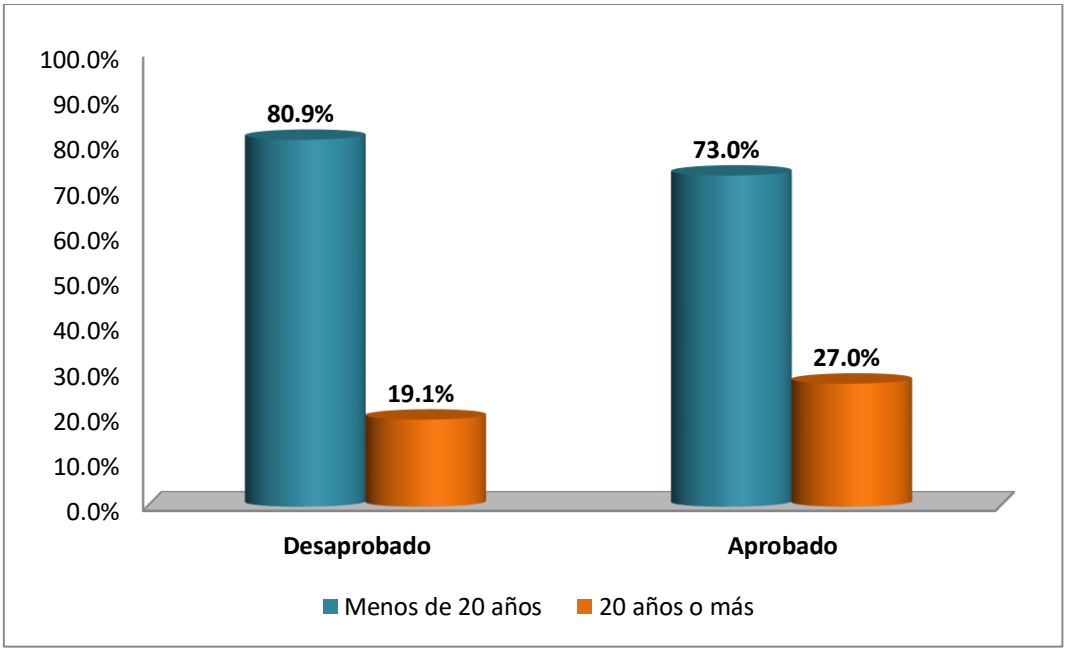
Cuadro 3.3.3

Resultado Final en Cálculo I y Grupo Etario

			Grupo Etario		Total
			Menos de 20 años	20 años o más	
Resultado final Cálculo I	Desaprobado	Recuento	55	13	68
		% dentro de Resultado final Cálculo I	80,9%	19,1%	100,0%
	Aprobado	Recuento	27	10	37
		% dentro de Resultado final Cálculo I	73,0%	27,0%	100,0%
Total		Recuento	82	23	105
		% dentro de Resultado final Cálculo I	78,1%	21,9%	100,0%

Gráfico 3.3.2

Resultado Final en Cálculo I y Grupo Etario



En la tabla resumen (3.3.3) se contempla que de los estudiantes que reprueban la asignatura de Calculo I el 19.1% tienen edad mayor a 20 años, esta ratio porcentual se incrementa a 27.0% en el conjunto de los alumnos que aprueban la asignatura.

Cuadro 3.3.4

Test de Chi cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	**,876[a]**	**1**	**,349**
Corrección por continuidad[b]	,475	1	,491
Razón de verosimilitudes	,858	1	,354
Estadístico exacto de Fisher			
Asociación lineal por lineal	,868	1	,352
N de casos válidos	105		

El "test" Chi-cuadrado (cuadro 3.3.4) para estudiar la dependencia de las variables, se observa que el *p_valor* es igual a 0.349, esto quiere decir que las variables "Resultado final del curso de Cálculo I" y "Grupo Etario" no están relacionadas, a un nivel de significación del 5%.

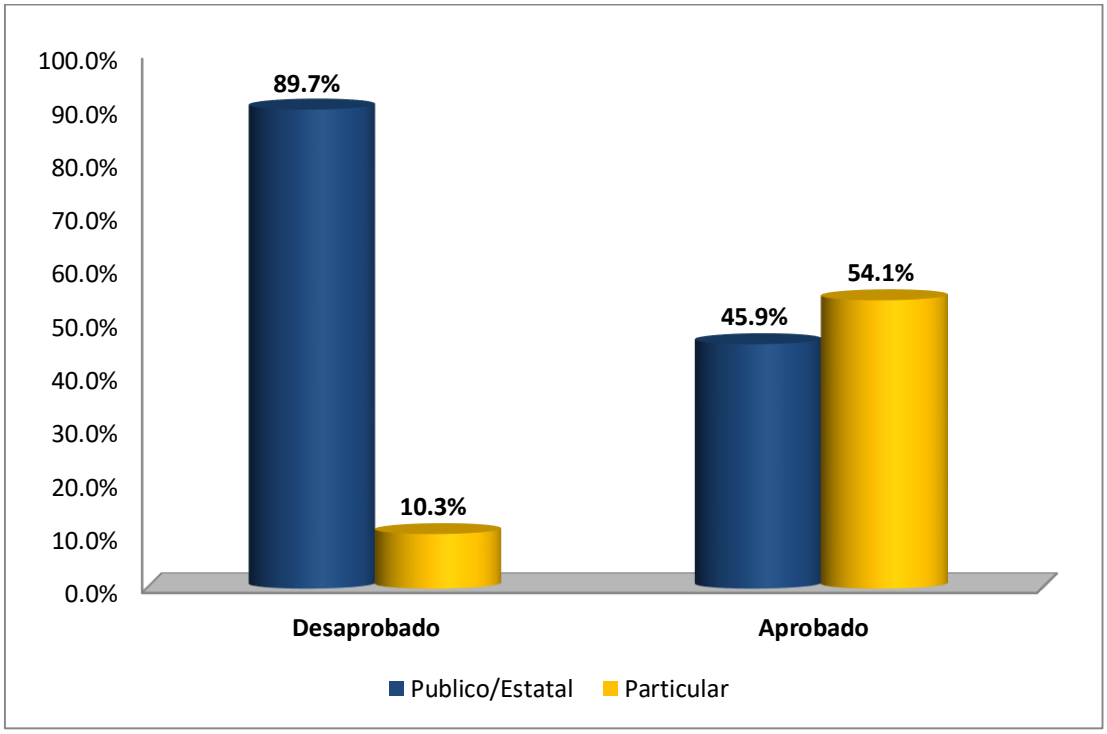
Cuadro 3.3.5

Resultado Final en Cálculo I y Tipo de Colegio

			Tipo de colegio		Total
			Publico/Estatal	Particular	
Resultado final Cálculo I	Desaprobado	Recuento	61	7	68
		% dentro de Resultado final Cálculo I	89,7%	10,3%	100,0%
	Aprobado	Recuento	17	20	37
		% dentro de Resultado final Cálculo I	45,9%	54,1%	100,0%
Total		Recuento	78	27	105
		% dentro de Resultado final Cálculo I	74,3%	25,7%	100,0%

Gráfico 3.3.3

Resultado Final en Cálculo I y Tipo de Colegio



En la tabla resumen (3.3.5), se destaca que del conjunto de estudiantes que reprueban la asignatura de Cálculo I el 10.3% proviene de un colegio Particular, este porcentaje se incrementa para este mismo conjunto en un 54.1% en el caso de los aprobados.

Cuadro 3.3.6

Test de Chi cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	**24,021[a]**	**1**	**,000**
Corrección por continuidad[b]	21,785	1	,000
Razón de verosimilitudes	23,577	1	,000
Estadístico exacto de Fisher			
Asociación lineal por lineal	23,792	1	,000
N de casos válidos	105		

En el cuadrado (cuadro 3.3.6) para ver la dependencia de las variables, se observa que el *p_valor* es igual a 0.00, este valor nos indica que las variables "Resultado final del curso de Cálculo I" y "Tipo de colegio" están relacionadas, un nivel de significancia del 5%.

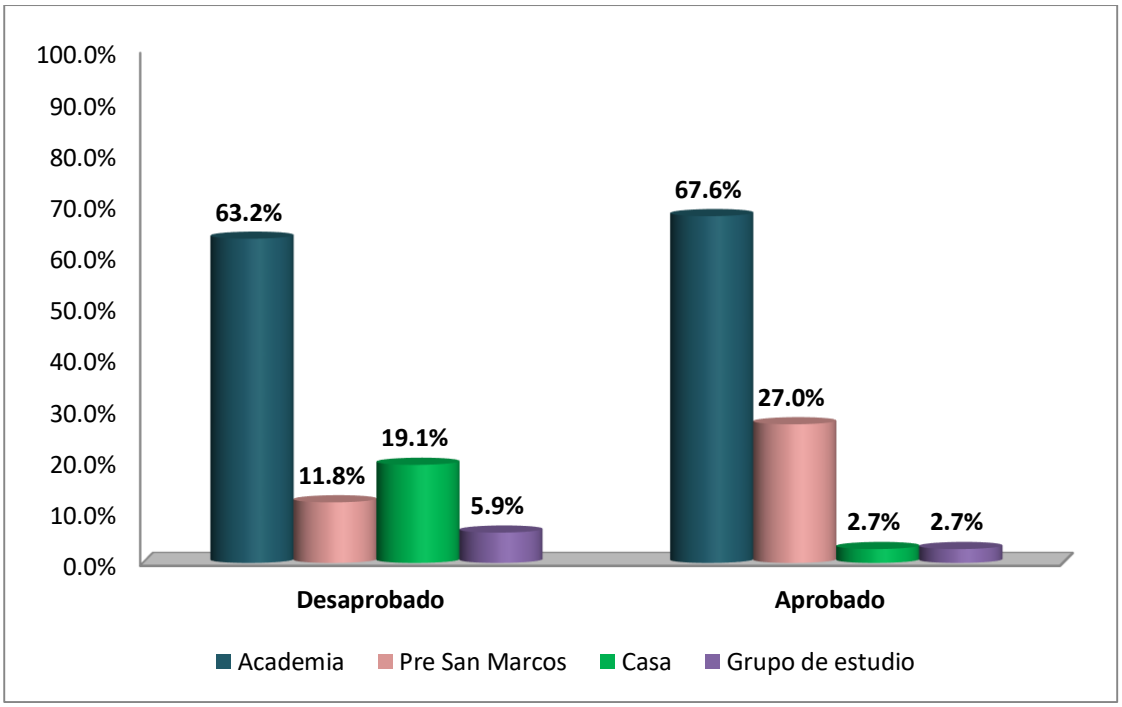
Cuadro 3.3.7

Resultado Final en Cálculo I y Lugar de Preparación

			Lugar de preparación				Total
			Academia	Pre San Marcos	Casa	Grupo de estudio	
Resultado final Cálculo I	Desaprobado	Recuento	43	8	13	4	68
		% dentro de Resultado final Cálculo I	63,2%	11,8%	19,1%	5,9%	100,0%
	Aprobado	Recuento	25	10	1	1	37
		% dentro de Resultado final Cálculo I	67,6%	27,0%	2,7%	2,7%	100,0%
Total		Recuento	66	20	14	5	111
		% dentro de Resultado final Cálculo I	64,8%	17,1%	13,3%	4,8%	100,0%

Gráfico 3.3.4

Resultado Final en Cálculo I y Lugar de Preparación



En la tabla resumen (3.3.7) se visualiza que de los estudiantes que reprueban la asignatura de Cálculo I el 11.8% realizo su preparación preuniversitaria en el centro Pre-San Marcos, así mismo en el conjunto de los aprobados el porcentaje se incrementa a un 27.0%.

Cuadro 3.3.8

Test de Chi cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	**8,677[a]**	**3**	**,034**
Razón de verosimilitudes	9,885	3	,020
Asociación lineal por lineal	2,191	1	,139
N de casos válidos	105		

El “test” Chi-cuadrado (cuadro 3.3.8) para ver la dependencia de las variables, se observa que el *p_valor* es igual a 0.034, este valor nos indica que las variables “Resultado final del curso de Cálculo I” y “Lugar de preparación” están relacionadas, con un nivel de significancia del 5%.

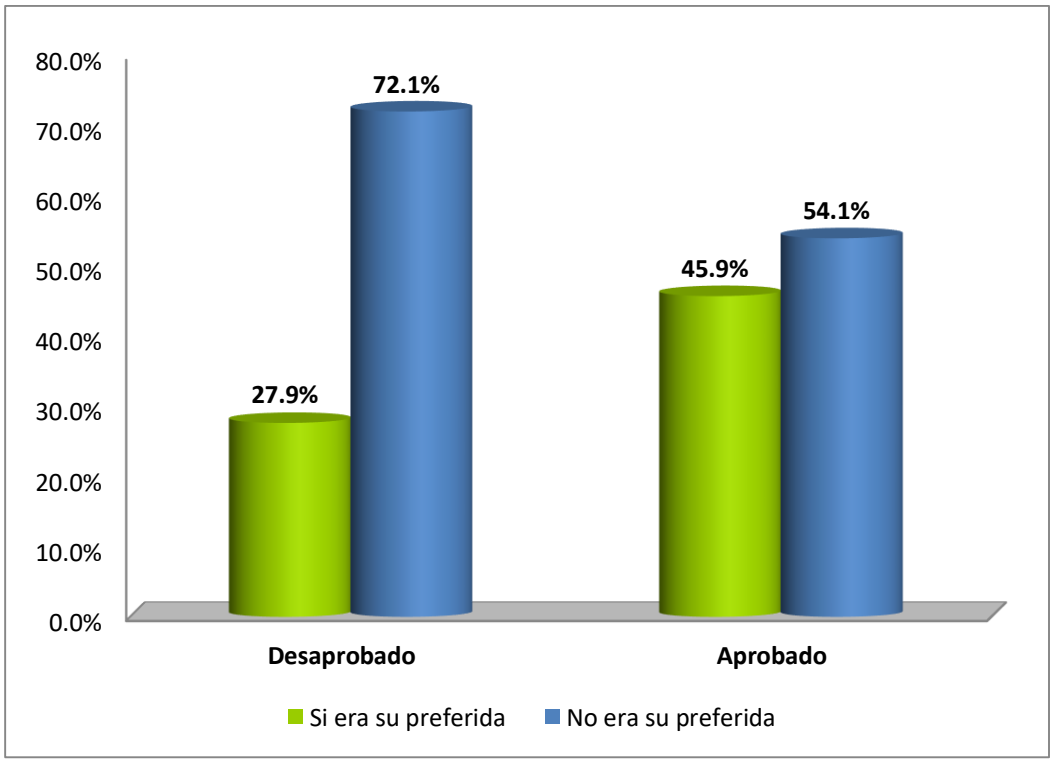
Cuadro 3.3.9

Resultado Final en Cálculo I y Preferencia por la Carrera

			Preferencia por la carrera		Total
			Si era su preferida	No era su preferida	
Resultado final Cálculo I	Desaprobado	Recuento	19	49	68
		% dentro de Resultado final Cálculo I	27,9%	72,1%	100,0%
	Aprobado	Recuento	17	20	37
		% dentro de Resultado final Cálculo I	45,9%	54,1%	100,0%
Total		Recuento	36	69	105
		% dentro de Resultado final Cálculo I	34,3%	65,7%	100,0%

Gráfico 3.3.5

Resultado Final en Cálculo I y Preferencia por la Carrera



En la tabla resumen (3.3.9) se destaca que de los estudiantes que no reprobaron la asignatura de Cálculo I el 72.1% la carrera que estudia en la actualidad no era su preferida, por otro lado, del conjunto de estudiantes que aprobaron el 45.9% carrera que estudia actualmente si era de su preferencia.

Cuadro 3.3.10

Test de Chi cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	**3,448[a]**	**1**	**,063**
Corrección por continuidad[b]	2,695	1	,101
Razón de verosimilitudes	3,396	1	,065
Estadístico exacto de Fisher			
Asociación lineal por lineal	3,415	1	,065
N de casos válidos	105		

El "test" Chi-cuadrado (cuadro 3.3.10) para ver la dependencia de las variables, se observa que el *p_valor* es igual a 0.063, este valor nos indica que las variables "Resultado final del curso de Cálculo I" y "Preferencia por la carrera que estudia" no están relacionadas, con un nivel de significancia del 5%.

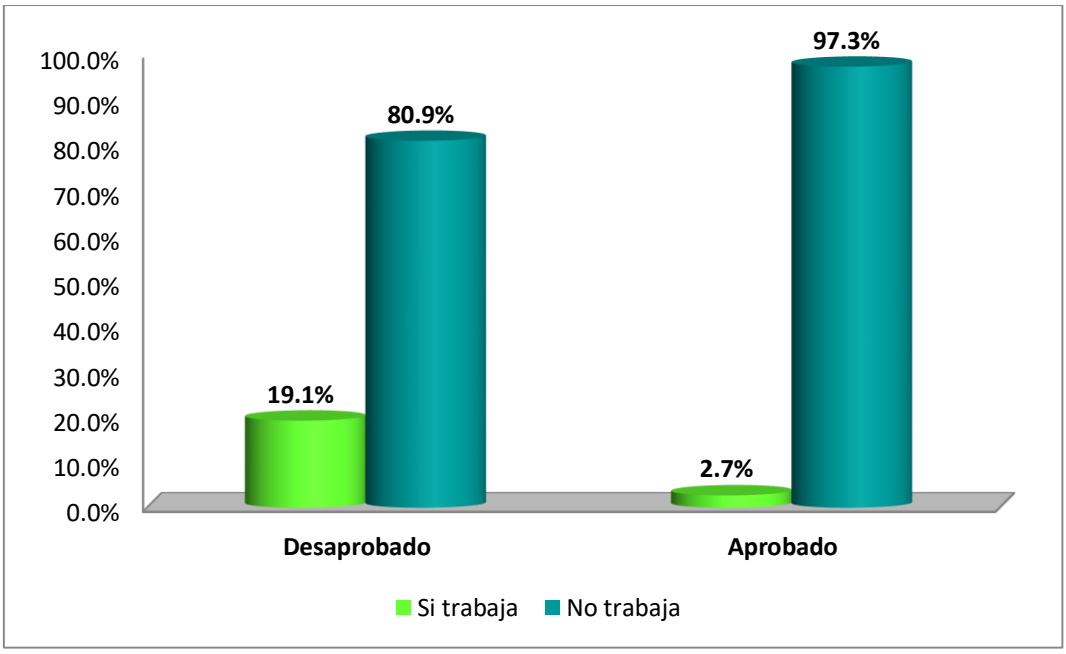
Cuadro 3.3.11

Resultado Final en Cálculo I y Condición Laboral del Alumno

			Condición Laboral		Total
			Si trabaja	No trabaja	
Resultado final Cálculo I	Desaprobado	Recuento	13	55	68
		% dentro de Resultado final Cálculo I	19,1%	80,9%	100,0%
	Aprobado	Recuento	1	36	37
		% dentro de Resultado final Cálculo I	2,7%	97,3%	100,0%
Total		Recuento	14	91	105
		% dentro de Resultado final Cálculo I	13,3%	86,7%	100,0%

Gráfico 3.3.6

Resultado Final en Cálculo I y Condición Laboral del Alumno



En la tabla resumen (3.3.11) se visualiza que del conjunto de alumnos que aprueban la asignatura de Cálculo I el 19.1% no tenían un trabajo paralelo a sus estudios universitarios. Del conjunto de alumnos que reprueban la asignatura de Cálculo I el 2.7% si tenían un trabajo paralelo a sus estudios universitarios.

Cuadro 3.3.12

Test de Chi cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	**5,587[a]**	**1**	**,018**
Corrección por continuidad[b]	4,257	1	,039
Razón de verosimilitudes	6,909	1	,009
Estadístico exacto de Fisher			
Asociación lineal por lineal	5,534	1	,019
N de casos válidos	105		

El “test” Chi-cuadrado (cuadro 3.3.12) para ver la dependencia de las variables, se observa que el *p_valor* es igual a 0.018, este valor nos indica que las variables “Resultado final del curso de Cálculo I” y “Condición Laboral del Alumno” están relacionadas, con un nivel de significancia del 5%.

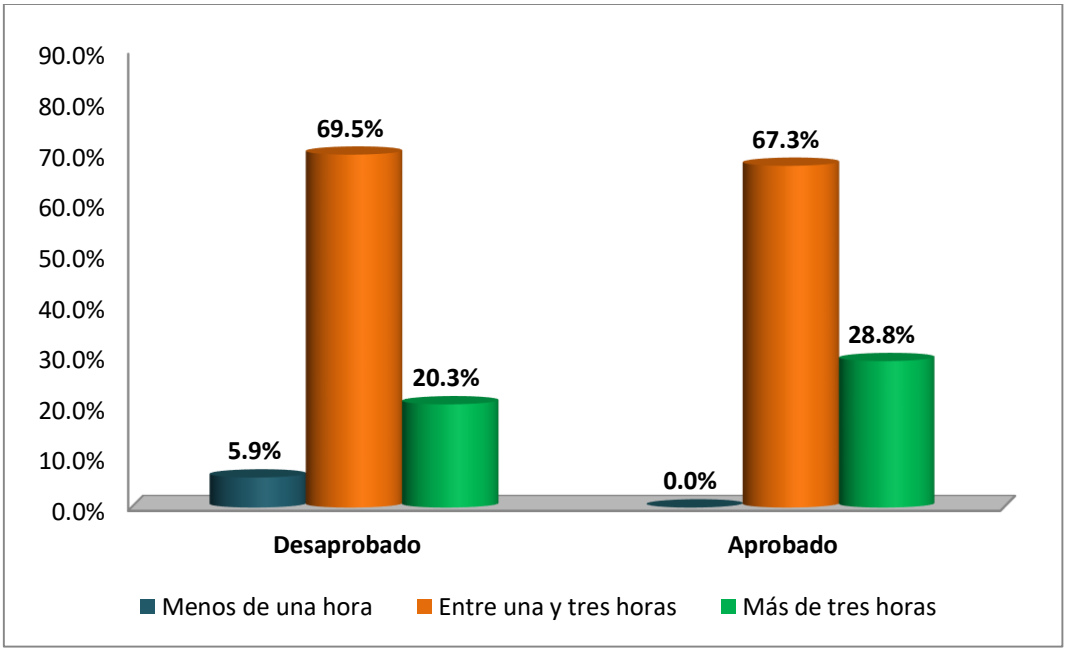
Cuadro 3.3.13

Resultado Final en Cálculo I y Tiempo de estudio

			Tiempo de estudio			Total
			< 1 hora	1 hora - 3 horas	Más de 3 horas	
Resultado final Matemática Básica	Desaprobado	Recuento	4	58	6	68
		% dentro de Resultado final Matemática Básica	5,9%	69,5%	20,3%	100,0%
	Aprobado	Recuento	0	29	8	37
		% dentro de Resultado final Matemática Básica	0,0%	67,3%	28,8%	100,0%
Total		Recuento	4	87	14	105
		% dentro de Resultado Matemática Básica	7,2%	68,5%	24,3%	100,0%

Gráfico 3.3.7

Resultado Final en Cálculo I y Tiempo de estudio



En la tabla resumen (3.3.13) se visualiza que de los estudiantes que reprueban la asignatura de Cálculo I el 20.3% posteriormente a las clases estudia tres horas o más, este porcentaje se incrementa en un 28,8% en el conjunto de los que aprueban la asignatura de Cálculo I.

Cuadro 3.3.14

Test de Chi cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	**5,258[a]**	**2**	**,072**
Razón de verosimilitudes	6,396	2	,041
Asociación lineal por lineal	5,101	1	,024
N de casos válidos	105		

El "test" Chi-cuadrado (cuadro 3.3.14) para ver la dependencia de las variables, se observa que el *p_valor* es igual a 0.072, este valor indica que las variables "Resultado final del curso de Matemática Básica" y "Tiempo de estudio" no están relacionadas, con un nivel de significancia del 5%.

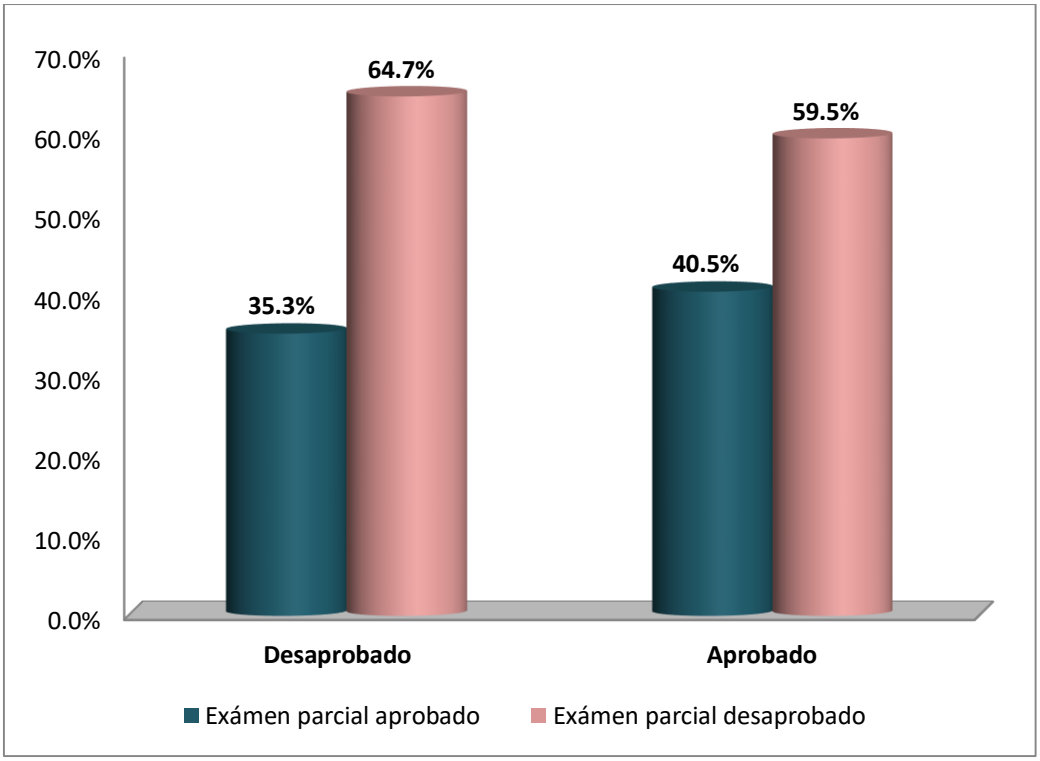
Cuadro 3.3.15

Resultado Final en Cálculo I y Resultado del Examen Parcial de Cálculo I

			Resultado del examen parcial de Cálculo I		Total
			Aprobado	Desaprobado	
Resultado final Cálculo I	Desaprobado	Recuento	24	44	68
		% dentro de Resultado final Cálculo I	35,3%	64,7%	100,0%
	Aprobado	Recuento	15	22	37
		% dentro de Resultado final Cálculo I	40,5%	59,5%	100,0%
Total		Recuento	39	66	105
		% dentro de Resultado final Cálculo I	37,1%	62,9%	100,0%

Gráfico 3.3.8

Resultado Final en Cálculo I y Resultado del Examen Parcial de Cálculo I



En el cuadro (3.3.15) se visualiza que de los estudiantes que reprueban el curso de Cálculo I el 64.7% tuvieron en su examen parcial nota desaprobatoria, esta cantidad se incrementa a un 59.5% para el conjunto de estudiantes que si aprobaron la asignatura de Cálculo I.

Cuadro 3.3.16

Test de Chi cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	**,282[a]**	**1**	**,595**
Corrección por continuidad[b]	,102	1	,749
Razón de verosimilitudes	,281	1	,596
Estadístico exacto de Fisher			
Asociación lineal por lineal	,280	1	,597
N de casos válidos	105		

El "test" Chi-cuadrado (cuadro 3.3.16) para ver la dependencia de las variables, se observa que el *p_valor* es igual a 0.595, este valor nos indica que las variables "Resultado final del

curso de Cálculo I" y "Resultado del Examen Parcial de Cálculo I" no están relacionadas, con un nivel de significancia del 5%.

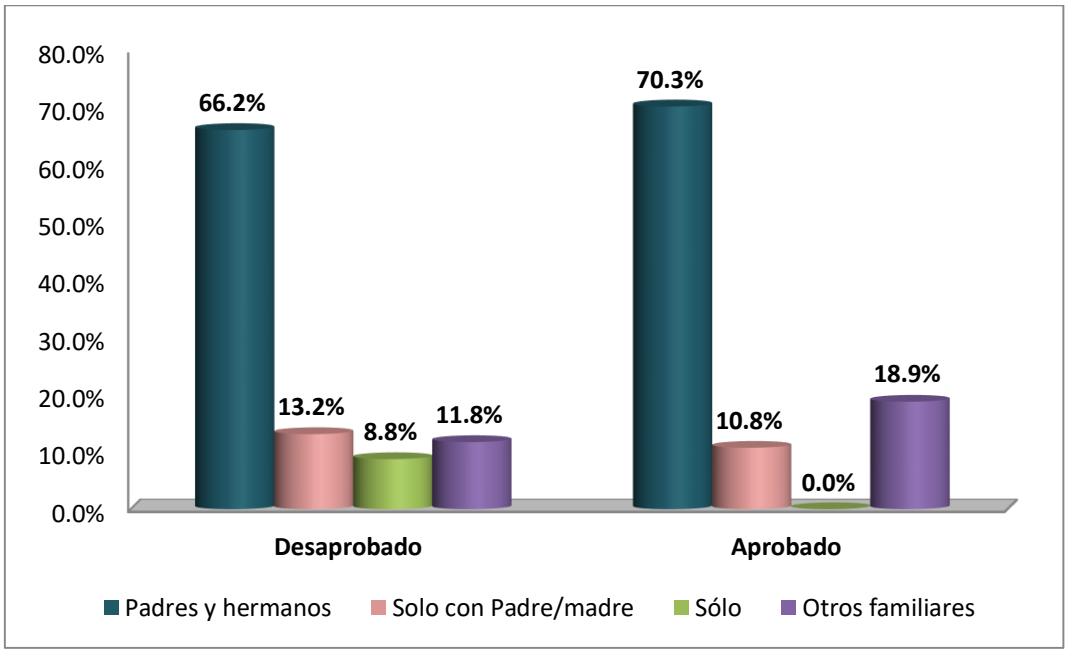
Cuadro 3.3.17

Resultado Final en Cálculo I y Personas con quien vive el alumno

			Personas con quién vive el alumno				Total
			Padres y hermanos	Solo con Padre/madre	Sólo	Otros familiares	
Resultado final Cálculo I	Desaprobado	Recuento	45	9	6	8	68
		% dentro de Resultado final Cálculo I	66,2%	13,2%	8,8%	11,8%	100,0%
	Aprobado	Recuento	26	4	0	7	37
		% dentro de Resultado final Cálculo I	70,3%	10,8%	0,0%	18,9%	100,0%
Total		Recuento	71	13	6	15	105
		% dentro de Resultado final Cálculo I	67,6%	12,4%	5,7%	14,3%	100,0%

Gráfico 3.3.9

Resultado Final en Matemática Básica y Personas con quien vive el alumno



En la tabla resumen (3.3.17) se visualiza que de los estudiantes que aprueban la asignatura de Cálculo I el 70.3% conviven con sus padres y hermanos. De los estudiantes que reprobaron la asignatura el 11.8% convive con otros familiares.

Cuadro 3.3.18

Test de Chi cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	**4,296[a]**	**3**	**,231**
Razón de verosimilitudes	6,215	3	,102
Asociación lineal por lineal	,004	1	,951
N de casos válidos	105		

El "test" Chi-cuadrado (cuadro 3.3.18) para ver la dependencia tiene un *p_valor* de 0.231, quiere decir que las variables "Resultado final del curso de Cálculo I" y "Personas con quien vive el alumno" no están relacionadas, con un nivel de significancia del 5%.

.

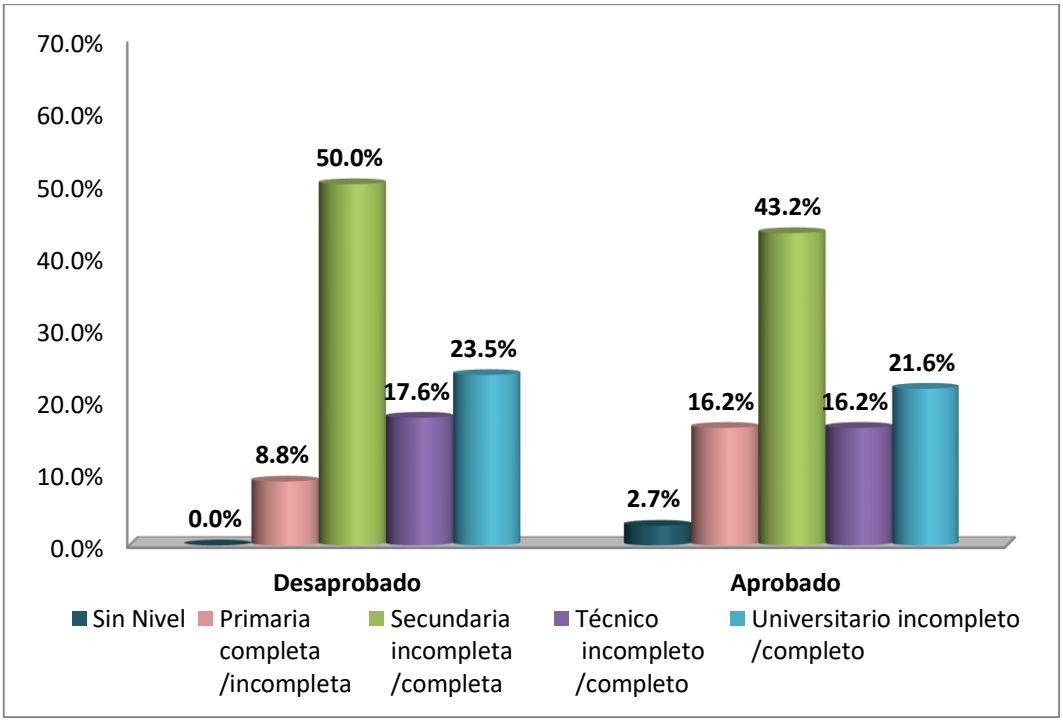
Cuadro 3.3.19

Resultado Final en Cálculo I y Nivel Académico del Padre

			Nivel académico del padre					Total
			Sin Nivel	Primaria completa /incompleta	Secundaria completa /incompleta	Técnico completa /incompleta	Universitario completa /incompleta	
Resultado final Cálculo I	Desaprobado	Recuento	0	6	34	12	16	68
		% dentro de Resultado final Cálculo I	,0%	8,8%	50,0%	17,6%	23,5%	100,0%
	Aprobado	Recuento	1	6	16	6	8	37
		% dentro de Resultado final Cálculo I	2,7%	16,2%	43,2%	16,2%	21,6%	100,0%
Total		Recuento	1	12	50	18	24	105
		% dentro de Resultado final Cálculo I	1,0%	11,4%	47,6%	17,1%	22,9%	100,0%

Gráfico 3.3.10

Resultado Final en Cálculo I y Nivel Académico del Padre



En la tabla resumen (3.3.19) se visualiza que el 21,6% de los estudiantes que aprueban la asignatura de Cálculo I, sus padres estudiaron una carrera universitaria completa /incompleta. El 50,0% del conjunto de estudiantes que reprobaron la asignatura de Cálculo I sus padres estudiaron hasta el grado de Secundaria.

Cuadro 3.3.20

Test de Chi cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	**3,280[a]**	**4**	**,512**
Razón de verosimilitudes	3,481	4	,481
Asociación lineal por lineal	,778	1	,378
N de casos válidos	105		

El "test" Chi-cuadrado (cuadro 3.3.20) para ver la dependencia de las variables, se observa que el *p_valor* es igual a 0.512, este valor nos indica que las variables "Resultado final del

curso de Cálculo I" y "Nivel académico del padre" son independientes, con un nivel de significancia del 5%.

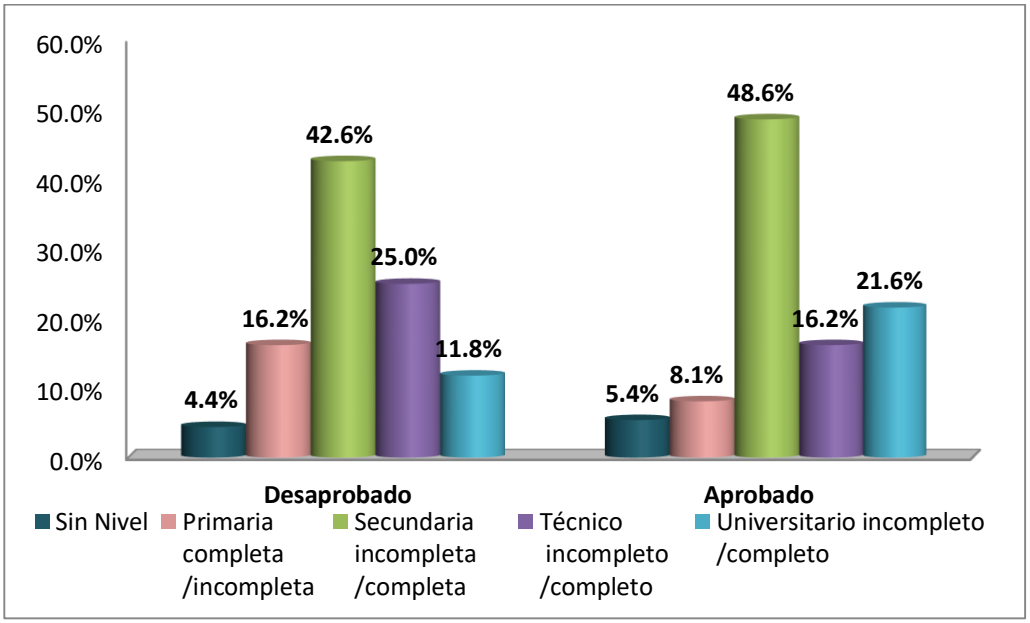
Cuadro 3.3.21

Resultado Final en Cálculo I y nivel académico de la madre

			Nivel académico de la madre					Total
			Sin Nivel	Primaria completa /incompleta	Secundaria completa /incompleta	Técnico completa /incompleta	Universitario completa /incompleta	
Resultado final Cálculo I	Desaprobado	Recuento	3	11	29	17	8	68
		% dentro de Resultado final Cálculo I	4,4%	16,2%	42,6%	25,0%	11,8%	100,0%
	Aprobado	Recuento	2	3	18	6	8	37
		% dentro de Resultado final Cálculo I	5,4%	8,1%	48,6%	16,2%	21,6%	100,0%
Total		Recuento	5	14	47	23	16	105
		% dentro de Resultado final Cálculo I	4,8%	13,3%	44,8%	21,9%	15,2%	100,0%

Gráfico 3.3.11

Resultado Final en Cálculo I y nivel académico de la madre



En la tabla resumen (3.3.21) se visualiza que el 25,0% del conjunto de estudiantes que reprueban la asignatura de Cálculo I, sus madres estudiaron una carrera Técnica Completa/Incompleta, por otro lado, el 48.6% del conjunto de estudiantes que aprueban la asignatura de Cálculo I sus madres alcanzaron el grado de Secundaria.

Cuadro 3.3.22

Test de Chi cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	**3,784[a]**	**4**	**,436**
Razón de verosimilitudes	3,852	4	,426
Asociación lineal por lineal	,645	1	,422
N de casos válidos	105		

El "test" Chi-cuadrado (cuadro 3.3.22) para ver la dependencia de las variables, se observa que el *p_valor* es igual a 0.436, este valor indica que las variables "Resultado final del curso de Cálculo I" y "Nivel académico de la madre" son independientes, con un nivel de significancia del 5%.

4 AJUSTE DEL MODELO DE REGRESIÓN LOGÍSTICA

Los modelos de Regresión Lineal o Logística pueden utilizarse con dos fines: 1) *Predecir* una variable dependiente en función a un grupo de variables independientes y 2) *Estimar*, lo cual consiste en calcular la relación de la variable dependiente con las variables independientes.

Es a causa de estos fines que los modelos de regresión tienen la tarea de crear mecanismos con el fin de hallar el modelo más adecuado que pueda cumplir ambos fines.

Los modelos Predictivos manejan dos objetivos el primero estimar el coeficiente de la variable de interés y el segundo trata de determinar factores que determinan los factores causas de un suceso o una enfermedad.

En la presente investigación se enfoca en el segundo objetivo ya que se busca saber cuáles son los factores que se asocian al bajo rendimiento de los alumnos ingresantes en los cursos de Matemática Básica y Cálculo I.

Por tal motivo nos concentramos en las "pruebas" que estiman los coeficientes y sus significancias ya que mediante estos indicadores podemos observar la relación que existe con la variable independiente.

Para realizar el ajuste del modelo se tiene que evaluar si las variables a considerar son las más apropiadas para el análisis. En el capítulo I se indicó las variables que fueron incluidas en el cuestionario aplicado a los estudiantes, así mismo utilizando el análisis exploratorio de las variables confirmaremos los resultados del presente estudio.

El análisis y resultados de la aplicación del modelo se realizó utilizando el programa es estadístico SPSS versión 21.0.

Como primer paso se ajustará un modelo de Regresión Logística tomando en cuenta todas las variables indicadas en el cuadro (1.4.3) del Capítulo 1.

4.1 ANÁLISIS DE RESULTADOS

4.1.1AJUSTE DEL MODELO MEDIANTE EL MÉTODO DIRECTO

a) Análisis del modelo directo en el resultado del curso de Matemática Básica

Según los resultados obtenidos utilizando el software SPSS 21.0 observamos que el cálculo de *–2 log de la verosimilitud* trabajando únicamente con la constante y con las variables en su totalidad se obtiene los siguientes resultados:

Cuadro 4.1.1

-2log de la verosimilitud para el modelo solo con la constante y con todas las variables

Modelo	-2 log de la verosimilitud
Solo con la constante	144.789
Con todas las variables	71.295

De los valores obtenidos podemos contrastar las siguientes hipótesis:

Ho: Todos los β_0 son equivalentes a cero.

H1: Al menos un β_0 es distinto a cero.

El estadístico G de la ecuación nos muestra el siguiente valor G = 144.789 – 71.295 = 73.494 y el p valor asociado para la prueba es

$P[\ x_{21}^2 > 73.594] = 0$, trabajando con un nivel de significancia equivalente al 5% se recha Ho y se concluye que como mínimo un β_0 es distinto de cero.

Para analizar la bondad del ajuste utilizaremos el Test de Hosmer y Lemeshow con las siguientes hipótesis, donde se plantean las siguientes hipótesis:

Ho: El modelo se ajusta apropiadamente.

H1: El modelo no se ajusta apropiadamente.

Cuadro 4.1.2

Prueba de Hosmer - Lemeshow

Paso	Chi cuadrado	gl	Sig.
1	12,033	8	,150

En la tabla (4.1.2) se observa que el p_*valor* es igual a 0.150, este valor nos indica que el ajuste del modelo es bueno debido a que no existen motivos para no aceptar Ho trabajando con un nivel de significancia equivalente al 5%.

Cuadro 4.1.3

Resumen del modelo

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	71,295[a]	,503	,673

Cuadro 4.1.4

Tabla de Clasificación

Observado			Pronosticado		
			Resultado final Matemática Básica		Porcentaje correcto
			Desaprobado	Aprobado	
Paso 1	Resultado Final Matemática Básica	Desaprobado	49	8	86,0
		Aprobado	8	40	83,3
	Porcentaje global				84,8

El valor del R^2 *de Nagelkerke* es de tan solo 0.673 lo que nos indica que las variables independientes explican adecuadamente a la variable dependiente. Por otro lado, la cantidad de casos clasificados adecuadamente por el modelo en una determinada categoría es bueno. De los casos observados del conjunto de estudiante que aprueban la asignatura de Matemática Básica (Cuadro 4.1.4) 83.3 % fueron clasificados adecuadamente por el modelo, de igual forma la clasificación global también es aceptable (84.8% de casos fueron correctamente clasificados por el modelo). Todos estos resultados fueron obtenidos mediante el método directo (Anexo 2, Resultados del análisis del curso de Matemática Básica**,** sección Bloque 1: Método Introducir).

b) Análisis del modelo directo en el resultado del curso de Cálculo I.

Según los resultados obtenidos utilizando el software SPSS 21.0 vemos que el *–2 log de la verosimilitud* trabajando únicamente con la constante y con las variables en su totalidad se obtiene los siguientes resultados:

Cuadro 4.1.5

-2log de la verosimilitud para el modelo solo con la constante y con todas las variables.

Modelo	-2 log de la verosimilitud
Solo con la constante	136,271
Con todas las variables	67,340

De estos valores obtenidos podemos contrastar las siguientes hipótesis:

Ho: Todos los β_0 son equivalentes a cero.

H1: Al menos un β_0 es distinto a cero.

El estadístico G de la ecuación nos muestra el siguiente valor se obtiene el siguiente valor G =136.271 – 67.340 = 68.698 y el p valor asociado para la prueba es

$P[\, x_{21}^2 > 68.931] = 0$, trabajando con un nivel de significancia equivalente al 5% rechazamos Ho y se concluye que como mínimo un β_0 es distinto de cero.

Para analizar la bondad del ajuste utilizaremos la prueba de Hosmer y Lemeshow donde se plantean las siguientes hipótesis:

Ho: El modelo se ajusta apropiadamente.

H1: El modelo no se ajusta apropiadamente.

Cuadro 4.1.6

Prueba de Hosmer - Lemeshow

Paso	Chi cuadrado	gl	Sig.
1	14,218	8	,076

En la tabla (4.1.6) se observa que el *p_valor* es igual a 0.076, este valor nos indica que el ajuste del modelo es bueno debido a que no existen motivos para no aceptar Ho trabajando con un nivel de significancia equivalente al 5%.

Cuadro 4.1.7

Resumen del modelo

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	67,340[a]	,481	,662

Cuadro 4.1.8

Tabla de Clasificación

Observado			Pronosticado		
			Resultado Final Cálculo I		Porcentaje correcto
			Desaprobado	Aprobado	
Paso 1	Resultado final cálculo	Desaprobado	64	4	94,1
		Aprobado	7	30	81,1
	Porcentaje global				89,5

El valor del R^2 *de Nagelkerke* es de tan solo 0.662 lo que nos indica que las variables independientes explican adecuadamente a la variable dependiente. Por otro lado, la cantidad de casos clasificados adecuadamente por el modelo en una determinada categoría es bueno. De los casos observados del conjunto de estudiante que aprueban la asignatura de Matemática Básica (Cuadro 4.1.8) 94.1 % fueron clasificados adecuadamente por el modelo, de igual forma la clasificación global también es aceptable (89.5% de casos fueron correctamente clasificados por el modelo). Todos estos resultados fueron obtenidos mediante el método directo (Anexo 2, Resultados del análisis del curso de Cálculo I, Bloque 1: Método Introducir)

4.2 AJUSTE DEL MODELO MEDIANTE EL MÉTODO DE WALD HACIA ADELANTE (FORWARD STEPWISE (WALD)) PARA LA SELECCIÓN DE VARIABLES

a) Análisis de los resultados para determinar los factores asociados al rendimiento del curso de Matemática Básica

Esta metodología inicia con el modelo ajustado tomando solo la constante. Las posteriores etapas van incluyendo variables considerando su menor p-valor relacionado a la 'Puntuación eficiente de Rao' y depurando variables según considerando el mayor p-valor relacionado al resultado de la prueba de Wald.

A continuación, se describirá etapa por etapa como va ingresando cada variable al modelo (Los detalles de este proceso se podrá observar en las tablas del Anexo 3 tituladas " Variables en la ecuación " y " Variables que no están en la ecuación " .

Paso Nº 0: Solo se considera la constante.

Paso Nº1: En esta primera etapa de todas las variables que no forman parte del modelo se selecciona la variable que tiene el menor o mayor *p_valor* relacionado al estadístico ''Puntuación" (ver Anexo 3, Cuadro 3.6).

La variable "Tipo de colegio" es la muestra el valor más alto (18.737) con un *p_valor* de 0.000 menor a 0.05 por lo tanto esta variable formara parte del modelo.

Paso Nª2: En esta segunda etapa de todas las variables que no forman parte del modelo se selecciona la variable que tiene el menor o mayor *p_valor* relacionado al estadístico ''Puntuación" (ver Anexo 3, Cuadro 3.12, paso 1),

La variable "Preferencia por la carrera" es la muestra el valor más alto (13.952) con un *p_valor* de 0.000 menor a 0.05 por lo tanto esta variable formara parte del modelo

Paso Nª3: Del conjunto de Variables que aún no forman parte del modelo el que posee el valor más alto del estadístico ''Puntuación'' (ver Anexo 3, Cuadro 3.12, paso 2), es la variable *"Condición laboral del alumno (trabaja en la actualidad)"* (4.768) que posee p_*valor* equivalente a 0.029 el cual es menor a 0.05 por lo tanto esta variable se incluirá en el modelo.

***Paso Nª4**:* Del conjunto de Variables que aún no forman parte del modelo el que posee el valor más alto del estadístico ‘‘Puntuación’’ (ver Anexo 3, Cuadro 3.12, paso 3), es la variable *“Con quien convive el alumno”* (9.109) que posee p_*valor* equivalente a 0.028 el cual es menor a 0.05 por lo tanto esta variable se incluirá en el modelo.

La prueba de ómnibus de cada uno de los pasos es la siguiente:

Cuadro 4.2.1

Prueba ómnibus sobre los coeficientes del modelo

		Chi cuadrado	gl	Sig.
Paso 1	Paso	19,617	1	,000
	Bloque	19,617	1	,000
	Modelo	19,617	1	,000
Paso 2	Paso	14,148	1	,000
	Bloque	33,765	2	,000
	Modelo	33,765	2	,000
Paso 3	Paso	5,482	1	,019
	Bloque	39,248	3	,000
	Modelo	39,248	3	,000
Paso 4	Paso	10,099	3	,018
	Bloque	49,347	6	,000
	Modelo	49,347	6	,000

En cada uno de los pasos se evaluarán las conjeturas siguientes:

En el caso del modelo que considera todas las variables se evalúan las conjeturas siguientes

H_0: Los β_0 son equivalentes a cero.

Para revisar la mejora que se manifiesta al cambiar de un paso a otro se evalúan las siguientes conjeturas:

H_1: La mejora del cambio de un paso a otro no es significativa.

Se observa que los p_valores relacionados al estadístico Chi-cuadrado son inferiores a 0.05 en consecuencia con un nivel de significancia del 0.05 podemos concluir que los β_0 que acompañan a las variables en un paso del proceso no son iguales a cero; esto significa que existe una mejora de un paso a otro y esta es significativa.

Para evaluar las siguientes hipótesis:

H_0: El modelo se ajusta apropiadamente a los datos en cada paso.

Contra

H_1: El modelo no se ajusta apropiadamente a los datos en cada paso.

La prueba de bondad de ajuste de Hosmer-Lemeshow presenta los siguientes resultados:

Cuadro 4.2.2

Prueba de Hosmer – Lemeshow

Paso	Chi cuadrado	gl	Sig.
1	,000	0	.
2	5,111	2	,078
3	8,392	4	,078
4	7,503	6	,277

Observamos que en cada paso el modelo se ajusta adecuadamente.

Cuadro 4.2.3

Resumen del modelo

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	125,171[a]	,170	,228
2	111,023[a]	,275	,368
3	105,541[b]	,312	,417
4	95,442[c]	,375	,501

En el cuarto paso el valor del R^2 de Nagelkerke es de 0.501 el cual comparado con el método "Directo" no se muestra una mejora significativa, esto podría ser causado por la presencia de casos discordantes u "outliers".

Finalmente las variables seleccionadas por el método Wald con pasos hacia delante (Forward Stepwise (Wald)) son los siguientes:

Cuadro 4.2.4

Variables seleccionadas

	B	E.T.	Wald	gl	Sig.	Exp(B)
TipodeColegio(1)	3,246	,836	15,074	1	,000	25,696
Preferenciaporlacarrera(1)	-1,580	,533	8,789	1	,003	,206
Trabajaenlaactualida(1)	2,516	1,148	4,804	1	,028	12,382
Conquienvive			7,773	3	,051	
Conquienvive(1)	-,794	,809	,964	1	,326	,452
Conquienvive(2)	-1,455	1,216	1,432	1	,232	,233
Conquienvive(3)	-2,559	1,008	6,447	1	,011	,077
Constante	-1,692	1,207	1,966	1	,161	,184

Por lo tanto, el modelo logístico será:

$$p_i = \frac{1}{1 + e^z}$$

Dónde:

Z= -1.692+ 3.246 TIPO_DE_COLEGIO (1) - 1.580 PREFERENCIA_ POR_ LA_ CARRERA (1) + 2.516 TRABAJA_EN_LA_ACTUALIDAD (1) - 0.794 CON_QUIEN_ VIVE (1) - 1.455 CON_QUIEN_VIVE(2) - 2.559 CON_QUIEN_VIVE(3)

Del presente modelo se puede visualizar que el signo que acompaña a los coeficientes es positivo esto indica que la variable considerada incrementa la probabilidad de ocurrencia del suceso en estudio, es decir aumentan la probabilidad de aprobar la asignatura de Matemática Básica.

Cuadro 4.2.5

Variables con signo positivo

	B	Exp(B)
Tipodecolegio(1)	3,246	15,074
Trabajaenlaactualida(1)	2,516	4,804

$e^{3.246} = 25.6874,$ este valor nos indica que existen aproximadamente veintiséis veces más posibilidades de aprobar el curso de Matemática Básica si el alumno estudio su etapa escolar en un colegio particular.

$e^{2.516} = 12.3790,$ este valor nos indica que existe aproximadamente doce veces más posibilidades de aprobar el curso de Matemática Básica si el alumno no trabaja.

Observamos también que existe una variable acompañado de un valor negativo, esto nos indica que la variable que está acompañando al coeficiente reduce la probabilidad de ocurrencia del evento que se está estudiando, por lo tanto, para este estudio indica que disminuye la probabilidad de aprobar el curso de Matemática Básica. Se tiene el resumen siguiente:

Cuadro 4.2.6

Variables con signo negativo

	B	Exp(B)
Preferenciaporlacarrera(1)	-1,580	0,533
Conquienvive(1)	-,794	,964
Conquienvive(2)	-1,455	1,432
Conquienvive(3)	-2,559	6,447
Constante	-1,692	,964

La variable que no favorece a aprobar el curso de Matemática Básica son: “preferencia por la carrera” (en el caso de que la carrera que estudia no era su favorita) y “con quien convive el alumno” (en el caso de que el alumno convive solo con padre o madre, con otros familiares o vive solo)

Cuadro 4.2.7

Tabla de Clasificación

Observado		Pronosticado		
		Resultado final Matemática Básica		Porcentaje correcto
		Desaprobado	Aprobado	
Resultado final Matemática Básica	Desaprobado	47	10	82,5
	Aprobado	15	33	68,8
Porcentaje global				76,2

Algo que se obtiene de los resultados es que los alumnos que aprueban y desaprueban la asignatura de Matemática Básica son clasificados adecuadamente por el modelo, observando la tabla de clasificación nos indica que el 82.5% de los alumnos desaprobados están correctamente clasificados, mientras que el 68.8% de los aprobados están correctamente clasificados.

En general nos indica que el 76.2% de los alumnos están correctamente clasificados lo que nos determina que el modelo calculado clasifica adecuadamente a los individuos.

b) Análisis de los resultados para determinar los factores asociados al rendimiento del curso de Cálculo I

Al igual que en el análisis de los resultados del curso de Matemática Básica, se realizará el análisis de resultados para determinar los factores asociados al rendimiento del curso de Cálculo I.

En este caso el método de Wald pasos hacia adelante considera tres pasos donde se irán introduciendo las variables que mejor se ajustan al modelo.

Paso Nº 0: Solo se considera la constante.

Paso Nº1: En esta primera etapa de todas las variables que no forman parte del modelo se selecciona la variable que tiene el menor o mayor *p_valor* relacionado al estadístico ‘‘Puntuación” (ver Anexo 3, Cuadro 3.18).

La variable “Tipo de colegio” es la muestra el valor más alto (24.021) con un *p_valor* de 0.000 menor a 0.05 por lo tanto esta variable formara parte del modelo.

Paso Nª2: En esta segunda etapa de todas las variables que no forman parte del modelo se selecciona la variable que tiene el menor o mayor *p_valor* relacionado al estadístico ‘‘Puntuación” (ver Anexo 3, Cuadro 3.25, paso 1),

La variable “Condición laboral del alumno” (trabaja en la actualidad)” es la muestra el valor más alto (6.173) con un *p_valor* de 0.000 menor a 0.05 por lo tanto esta variable formara parte del modelo

Paso Nª3: Del conjunto de Variables que aún no forman parte del modelo el que posee el valor más alto del estadístico ‘‘Puntuación’’ (ver Anexo 3, Cuadro 3.25, paso 2), es la variable *“Lugar de preparación”* (9.090) que posee p_*valor* equivalente a 0.028 el cual es menor a 0.05 por lo tanto esta variable se incluirá en el modelo.

La prueba de ómnibus de cada uno de los pasos es la siguiente:

Cuadro 4.2.8

Pruebas Omnibus sobre los coeficientes del modelo

		Chi cuadrado	gl	Sig.
Paso 1	Paso	23,577	1	,000
	Bloque	23,577	1	,000
	Modelo	23,577	1	,000
Paso 2	Paso	7,696	1	,006
	Bloque	31,273	2	,000
	Modelo	31,273	2	,000
Paso 3	Paso	10,450	3	,015
	Bloque	41,723	5	,000
	Modelo	41,723	5	,000

En cada uno de los pasos se evaluarán las conjeturas siguientes:

En el caso del modelo que considera todas las variables se evalúan las conjeturas siguientes

H_0: Los β_0 son equivalentes a cero.

Para revisar la mejora que se manifiesta al cambiar de un paso a otro se evalúan las siguientes conjeturas:

H_1: La mejora del cambio de un paso a otro no es significativa.

Se observa que los p_valores relacionados al estadístico Chi-cuadrado son inferiores a 0.05 en consecuencia con un nivel de significancia del 0.05 podemos concluir que los β_0 que acompañan a las variables en un paso del proceso no son iguales a cero; esto significa que existe una mejora de un paso a otro y esta es significativa.

Para contrastar la siguiente hipótesis:

H_0: El modelo se ajusta apropiadamente a los datos en cada paso.

Contra

H_1: El modelo no se ajusta apropiadamente a los datos en cada paso.

La prueba de bondad de ajuste de Hosmer-Lemeshow presenta los siguientes resultados:

Cuadro 4.2.9

Prueba de Hosmer - Lemeshow

Paso	Chi cuadrado	gl	Sig.
1	,000	0	.
2	,451	2	,798
3	1,990	5	,851

Con lo que concluimos que los modelos en cada paso si se ajustan adecuadamente, p-valor > 0.05.

.Cuadro 4.2.10

Resumen del modelo

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	112,694[a]	,201	,277
2	104,998[b]	,258	,354
3	94,548[c]	,328	,451

En el cuarto paso el valor del R^2 de Nagelkerke es de 0.451 el cual comparado con el método "Directo" no ha tenido mejora alguna, esto podría deberse a la influencia de los casos discordantes u "outliers".

Finalmente, las variables seleccionadas por el método Wald con pasos hacia delante (Forward Stepwise (Wald)) son los siguientes:

Cuadro 4.2.11

	B	E.T.	Wald	gl	Sig.	Exp(B)
Tipodecolegio(1)	2,713	,630	18,525	1	,000	15,067
Lugardepreparación			7,428	3	,059	
Lugardepreparación(1)	1,055	,628	2,824	1	,093	2,871
Lugardepreparación(2)	-2,328	1,187	3,846	1	,050	,098
Lugardepreparación(3)	,013	1,340	,000	1	,992	1,013
Trabajaenlaactualida(1)	2,706	1,322	4,188	1	,041	14,971
Constante	-3,849	1,357	8,047	1	,005	,021

Entonces el modelo logístico será

$$p_i = \frac{1}{1 + e^z}$$

Dónde:

Z= -3.849 + 2.713TIPO_COLEGIO (1) + 1.055LUGAR_DE_PREPARACIÓN (1) – 2.328LUGAR_DE_PREPARACIÓN (2) + 0.013LUGAR_DE_PREPARACIÓN (3) + 2.706TRABAJ_EN_LA_ACTUALIDAD (1)

Del presente modelo se puede visualizar que el signo que acompaña a los coeficientes es positivo esto indica que la variable considerada incrementa la probabilidad de ocurrencia del suceso en estudio, es decir aumentan la probabilidad de aprobar la asignatura de Cálculo I.

Cuadro 4.2.12

	B	Exp(B)
Tipodecolegio(1)	2,713	15,067
Lugardepreparación(1)	1,055	2,871
Lugardepreparación(3)	,013	1,013
Trabajaenlaactualida(1)	2,706	14,971

$e^{2.713} = 15.067$, Este valor nos indica que existen aproximadamente quince veces más posibilidades de aprobar el curso si el alumno estudio en un colegio particular.

$e^{1.055} = 2.871$ Este valor nos indica que existen aproximadamente tres veces más posibilidades de aprobar el curso si el alumno se preparó para ingresar a la universidad en el centro pre universitario de la UNMSM.

$e^{0.13} = 1.013$, Este valor nos indica que existe aproximadamente una posibilidad más de aprobar el curso si el alumno se preparó para ingresar a la universidad en un grupo de estudio.

$e^{2.706} = 14.971$, Este valor nos indica que existen aproximadamente catorce veces más posibilidades de aprobar el curso si el alumno no trabaja.

Observamos también que existe una variable acompañado de un valor negativo, esto nos indica que la variable que está acompañando al coeficiente reduce la probabilidad de ocurrencia del evento que se está estudiando, por lo tanto, para este estudio indica que disminuye la probabilidad de aprobar el curso de Cálculo I. Se tiene el resumen siguiente:

Cuadro 4.2.13

	B	Exp(B)
Lugardepreparación(2)	-2,328	,098

La variable que no favorece a aprobar el curso de Cálculo I es: Lugar de preparación (sea el caso en que el alumno se preparó para ingresar a la universidad en su hogar).

Cuadro 4.2.14

Observado		Pronosticado		
		Resultado final cálculo I		Porcentaje correcto
		Desaprobado	Aprobado	
Resultado final cálculo	Desaprobado	64	4	94,1
	Aprobado	19	18	48,6
Porcentaje global				78,1

Algo que se obtiene de los resultados es que los alumnos que aprueban y desaprueban la asignatura de Matemática Básica son clasificados adecuadamente por el modelo, observando la tabla de clasificación nos indica que el 94.1% de los alumnos desaprobados están correctamente clasificados, mientras que el 48.6% de los aprobados están correctamente clasificados.

En general nos indica que el 78.1% de los alumnos están correctamente clasificados lo que nos determina que el modelo calculado clasifica adecuadamente a los individuos

CAPÍTULO IV
CONCLUSIONES Y RECOMENDACIONES

CONCLUSIONES

Con los resultados obtenidos mediante el análisis de regresión logística podemos obtener las siguientes conclusiones:

1.- El objetivo de la presenta investigación es determinar qué factores influyen en el rendimiento académico de los alumnos ingresantes en los cursos de Matemática Básica y Cálculo I, según los resultados calculados aplicando el análisis de Regresión Logística, el valor del estadístico de Wald en el caso del Curso de Matemática Básica nos indica que el grupo de variables que más influyen son: "Tipo de Colegio", "Preferencia por la carrera", "Condición laboral del alumno" y "Con quien convive el alumno". Todas estas variables poseen un p-valor del estadístico de Wald inferior al 5%. Por otro lado, en el caso de los factores que influyen en el rendimiento académico del curso de Cálculo I las variables seleccionadas son: "Tipo de Colegio" y "condición laboral del alumno".

2.- En el caso del curso de Matemática Básica observando los coeficientes del modelo concluimos que en primer lugar la variable más importante es "Tipo de colegio" y en segundo lugar "Condición laboral del alumno".

3.- El grupo de variables que no fueron consideradas como predominantes en el rendimiento del curso de Matemática Básica son: sexo, edad, lugar de preparación, tiempo de estudio del curso, resultado del examen parcial de Matemática Básica, nivel académico alcanzado por el padre y nivel académico alcanzado por parte de la madre.

4.-En el modelo logístico del curso de Matemática Básica la variable "tipo de colegio" tiene el coeficiente con signo positivo, mediante este valor podemos concluir que si el alumno estudio en un colegio Particular posee mayores posibilidades de aprobar el curso que los alumnos que provienen de un colegio Nacional.

5.-La variable "Preferencia por la carrera" tiene coeficiente negativo en el modelo logístico del curso de Matemática Básica, lo que indica que si el alumno no tuvo como preferida la carrera que estudia en la actualidad, disminuye sus posibilidades de aprobar el curso de Matemática Básica. Este resultado muestra claramente el mejor performance que tiene el alumno en la universidad si tenía como vocación y preferencia la carrera que estudia.

6.-La variable “Con quien convive el alumno” tiene coeficientes negativos en el modelo logístico del curso de Matemática Básica, lo que indica que si el alumno convive únicamente con uno de sus padres, vive solo o el alumno vive con otros familiares tiene menos posibilidades de aprobar el curso de Matemática Básica que los alumnos que conviven con todos los miembros de su familia (Padres y hermanos). De este resultado podemos concluir que si el alumno convive en un hogar bien constituido con todo sus miembros de su familia influye positivamente en su performance en la universidad.

7.-La variable “Condición laboral del alumno” en el modelo logístico del curso de Matemática Básica, tiene coeficiente positivo lo que determina que si el alumno no trabaja tiene más posibilidades de aprobar que los alumnos que si trabajan.

8.- El conjunto de variable que influyen directamente en aprobar del curso de Cálculo I en primer lugar está el “tipo de colegio” esto por tener el mayor coeficiente, en segundo lugar “condición laboral del alumno “y en tercer lugar el “lugar de preparación”.

9.- El grupo de variables que no se consideran como predominante en el rendimiento del curso de Cálculo I son: sexo, edad, preferencia por la carrera, tiempo de estudio del curso, resultado del examen parcial de Cálculo I, nivel académico alcanzado tanto por parte del padre como la madre

10.-La variable “Tipo de colegio” está asociada a un coeficiente positivo dentro del modelo de regresión logística del curso de Cálculo I, mediante este valor podemos concluir que el alumno incrementa sus posibilidades de aprobar el curso de Cálculo I si estuvo previo a su ingreso a la Universidad en un colegio Particular.

11.- La variable “Lugar de preparación” en el modelo logístico del curso de Cálculo I tiene signos positivos y negativos, de este resultado podemos indicar que el alumno que se preparó para ingresar a la universidad en el centro Pre de la UNMSM y en un grupo de estudios, tiene mayor posibilidad de aprobar que si se preparó en una academia y en su hogar.

12.- La variable “Condición laboral del alumno” en el modelo logístico del curso de Cálculo I tiene coeficientes positivos lo que determina que el alumno tiene mayores posibilidades de aprobar el curso de Cálculo I si no trabaja.

RECOMENDACIONES

1.-La presenta investigación no incide en factores asociados a la enseñanza del profesor que pueden influir en el rendimiento académico del alumno, se sugiere que en estudios futuros se estructure preguntas que incidan en este tema.

2.-Una de las variables que se puede tomar en cuenta en investigaciones posteriores son las asociadas a las metodologías de enseñanza que se utilizan dentro de la Facultad de Ciencia Matemáticas para que los alumnos puedan comprender de manera adecuada los cursos

3.-En el presente trabajo se consideró como punto de corte para determinar la proporción de grupos de aprobados y desaprobados igual a 0.5, en futuros estudios so podría considerar un punto de corte diferente.

4.-Del conjunto de variables consideradas se destaca que una de las variables que más influye en tener un buen performance de los alumnos en la universidad es si la carrera que están estudiando los alumnos eran de sus preferidas previo a su ingreso, por lo tanto, se podría incentivar y motivar a los estudiantes sobre los campos laborales y de investigación que maneja cada una de las carreras de la Facultad de Ciencia Matemáticas.

BIBLIOGRAFÍA

ÁLVAREZ, J. (2011). Causas endógenas y exógenas del rendimiento académico de los estudiantes de matemática de la facultad de ciencias de la educación de la UNJBG de Tacna. Ciencia y Desarrollo. Perú – Tacna.

BARAHONA, P. (2012) .Factores determinantes del rendimiento académico de los estudiantes de la Universidad de Atacama. (Tesis de Maestría). Universidad de Atacama. Chile.

HERNÁNDEZ, D. (2011). Impacto de la didáctica en el rendimiento académico universitario. (Tesis de Licenciatura). Universidad Nacional Federico Villarreal. Perú.

HOSMER. W. D. y LEMESHOW, S. Applied Logistic Regression. Second Edition. A Wiley-Interscience Publication. John Wiley & Sons, INC. 2000.

IZAR, J. (2011) .Factores que afectan el desempeño académico de los estudiantes del nivel superior en Rioverde, San Luis de Potosí, México. Universidad de Veracruz. México.

LAWRENCE C. HAMILTON. Regression with Graphics. An Second Course inApplied Statistics. Wadsworth Publishing Company Belmont, California. 1992.

MARTÍNEZ, H.; RAMÍREZ, G. & ZALAZAR, L. (2011). Análisis del bajo rendimiento en matemática de los ingresantes de la facultad de ciencias económicas. (Tesis Maestría) Universidad Nacional del Nordeste. Argentina.

OCAÑA, Y. (2012). Artículo: Variables académicas que influyen en el rendimiento académico de los estudiantes universitarios. Investigación Educativa. Universidad Nacional Mayor de San Marcos. Perú.

PLASENCIA, S. (2010). Factores relacionados con el rendimiento académico en matemática en los estudiantes de la Universidad Nacional de Educación “Enrique Guzmán y Valle” en el año 2010. (Tesis de Maestría). Universidad Nacional de Educación “Enrique Guzmán y Valle”. Perú.

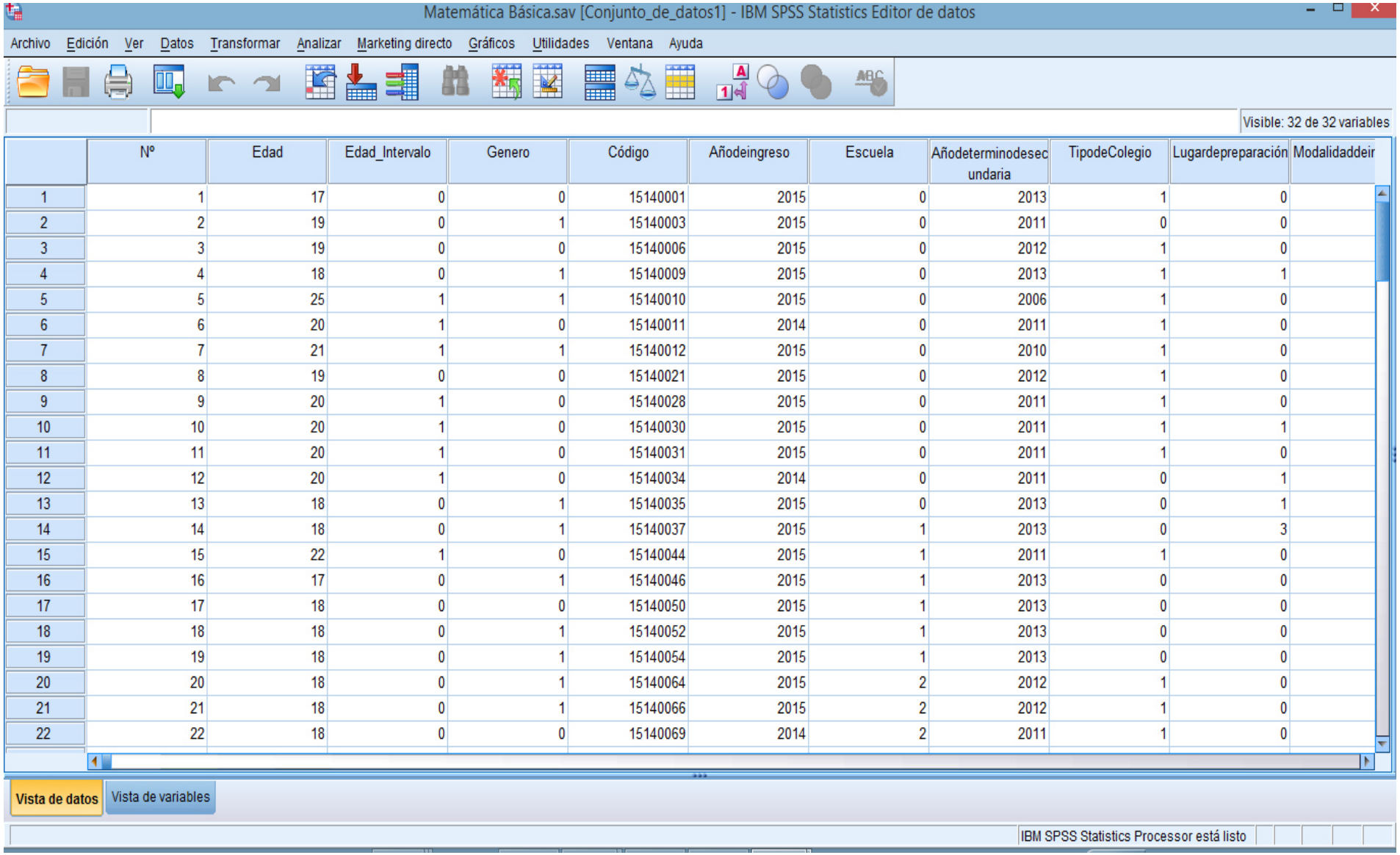
ROQUE, M. (200). Factores que influyen en el rendimiento de la matemática en el estudiante del ciclo básico, del Instituto Oficial Mixto Básico Leonídas Méncos Ávila. Tiquisate – Escuintla. (Tesis de Licenciatura) Universidad de San Carlos de Guatemala. Guatemala.

TEJEDOR, J. Y VALCÁRCEL, A. (2011). Causas del bajo rendimiento del estudiante universitario (en la opinión de profesores y alumnos) Propuestas de mejora en el marco de la EEES (Espacio Europeo de Educación Superior). (Tesis de Grado) Universidad de Salamanca. España.

ANEXOS

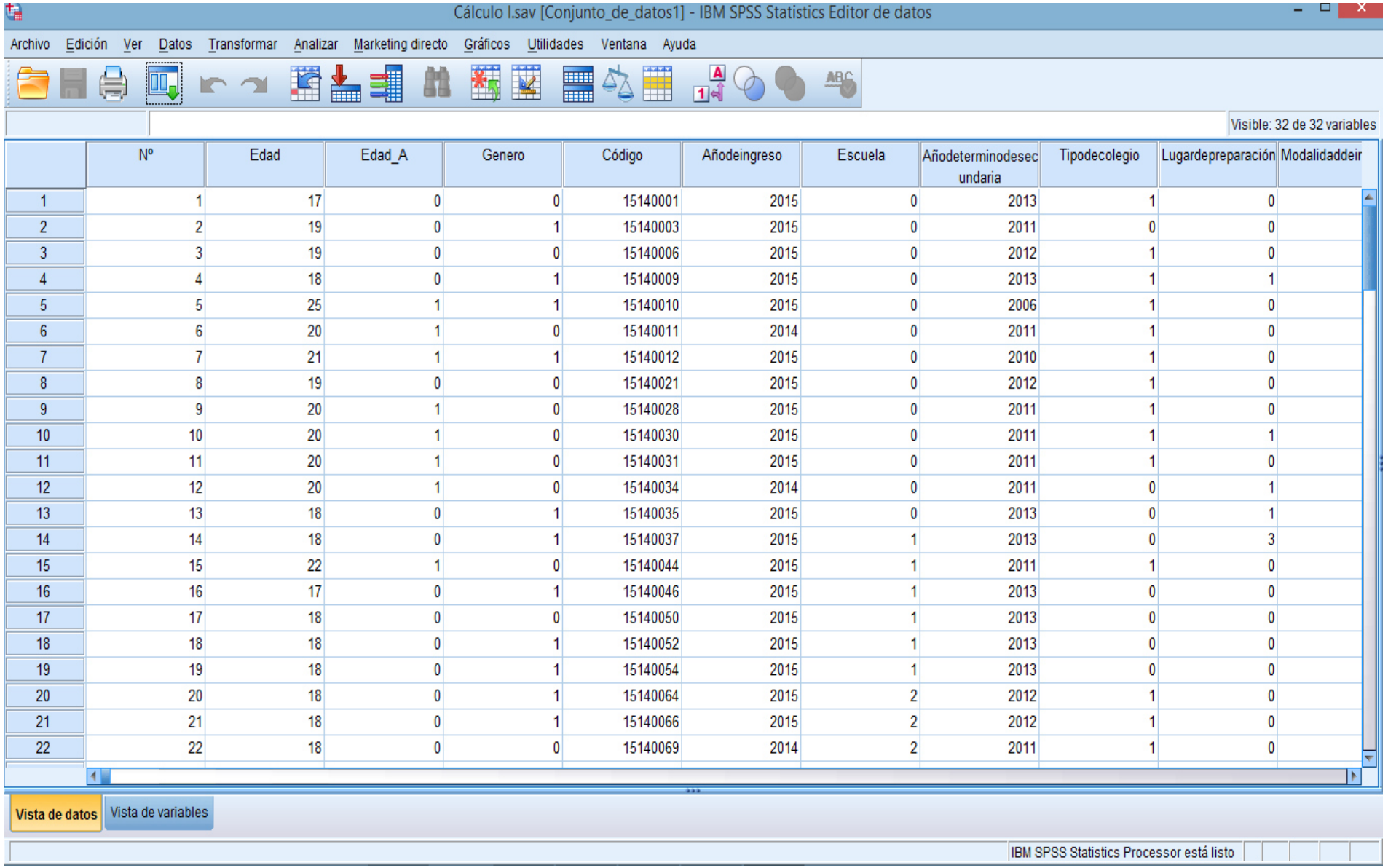
ANEXO 1

Parte de la Base de datos curso Matemática Básica



	Nº	Edad	Edad_Intervalo	Genero	Código	Añodeingreso	Escuela	Añodeterminodesecundaria	TipodeColegio	Lugardepreparación	Modalidaddeir
1	1	17	0	0	15140001	2015	0	2013	1	0	
2	2	19	0	1	15140003	2015	0	2011	0	0	
3	3	19	0	0	15140006	2015	0	2012	1	0	
4	4	18	0	1	15140009	2015	0	2013	1	1	
5	5	25	1	1	15140010	2015	0	2006	1	0	
6	6	20	1	0	15140011	2014	0	2011	1	0	
7	7	21	1	1	15140012	2015	0	2010	1	0	
8	8	19	0	0	15140021	2015	0	2012	1	0	
9	9	20	1	0	15140028	2015	0	2011	1	0	
10	10	20	1	0	15140030	2015	0	2011	1	1	
11	11	20	1	0	15140031	2015	0	2011	1	0	
12	12	20	1	0	15140034	2014	0	2011	0	1	
13	13	18	0	1	15140035	2015	0	2013	0	1	
14	14	18	0	1	15140037	2015	1	2013	0	3	
15	15	22	1	0	15140044	2015	1	2011	1	0	
16	16	17	0	1	15140046	2015	1	2013	0	0	
17	17	18	0	0	15140050	2015	1	2013	0	0	
18	18	18	0	1	15140052	2015	1	2013	0	0	
19	19	18	0	1	15140054	2015	1	2013	0	0	
20	20	18	0	1	15140064	2015	2	2012	1	0	
21	21	18	0	1	15140066	2015	2	2012	1	0	
22	22	18	0	0	15140069	2014	2	2011	1	0	

Parte de la Base de datos curso Cálculo I



	Nº	Edad	Edad_A	Genero	Código	Añodeingreso	Escuela	Añodeterminodesecundaria	Tipodecolegio	Lugardepreparación	Modalidaddeir
1	1	17	0	0	15140001	2015	0	2013	1	0	
2	2	19	0	1	15140003	2015	0	2011	0	0	
3	3	19	0	0	15140006	2015	0	2012	1	0	
4	4	18	0	1	15140009	2015	0	2013	1	1	
5	5	25	1	1	15140010	2015	0	2006	1	0	
6	6	20	1	0	15140011	2014	0	2011	1	0	
7	7	21	1	1	15140012	2015	0	2010	1	0	
8	8	19	0	0	15140021	2015	0	2012	1	0	
9	9	20	1	0	15140028	2015	0	2011	1	0	
10	10	20	1	0	15140030	2015	0	2011	1	1	
11	11	20	1	0	15140031	2015	0	2011	1	0	
12	12	20	1	0	15140034	2014	0	2011	0	1	
13	13	18	0	1	15140035	2015	0	2013	0	1	
14	14	18	0	1	15140037	2015	1	2013	0	3	
15	15	22	1	0	15140044	2015	1	2011	1	0	
16	16	17	0	1	15140046	2015	1	2013	0	0	
17	17	18	0	0	15140050	2015	1	2013	0	0	
18	18	18	0	1	15140052	2015	1	2013	0	0	
19	19	18	0	1	15140054	2015	1	2013	0	0	
20	20	18	0	1	15140064	2015	2	2012	1	0	
21	21	18	0	1	15140066	2015	2	2012	1	0	
22	22	18	0	0	15140069	2014	2	2011	1	0	

ANEXO 2

Modelo obtenido utilizando el método directo (ENTER)

Resultados del análisis del curso de Matemática Básica

Cuadro 2.1

Resumen del procesamiento de los casos

Casos no ponderados[a]		N	Porcentaje
Casos seleccionados	Incluidos en el análisis	105	100,0
	Casos perdidos	0	,0
	Total	105	100,0
Casos no seleccionados		0	,0
Total		105	100,0

a. Si está activada la ponderación, consulte la tabla de clasificación para ver el número total de casos.

Bloque 0: Bloque inicial

Cuadro 2.2

Historial de iteraciones[a,b,c]

Iteración		-2 log de la verosimilitud	Coeficientes Constant
Paso 0	1	144,789	-,171
	2	144,789	-,172

Cuadro 2.2

Tabla de clasificación[a,b]

Observado			Pronosticado		
			Resultado final Matemática Básica		Porcentaje correcto
			Desaprobado	Aprobado	
Paso 0	Resultado final Matemática Básica	Desaprobado	57	0	100,0
		Aprobado	48	0	,0
	Porcentaje global				54,3

Cuadro 2.3

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 0	Constante	-,172	,196	,770	1	,380	,842

Bloque 1: Método = Introducir

Cuadro 2.4

Pruebas omnibus sobre los coeficientes del modelo

		Chi cuadrado	gl	Sig.
Paso 1	Paso	73,494	22	,000
	Bloque	73,494	22	,000
	Modelo	73,494	22	,000

Cuadro 2.5

Resumen del modelo

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	71,295[a]	,503	,673

Cuadro 2.6

Prueba de Hosmer y Lemeshow

Paso	Chi cuadrado	gl	Sig.
1	12,033	8	,150

Cuadro 2.7

Tabla de clasificación[a]

Observado			Pronosticado		
			Resultado final Matemática Básica		Porcentaje correcto
			Desaprobado	Aprobado	
Paso 1	Resultado final Matemática Básica	Desaprobado	49	8	86,0
		Aprobado	8	40	83,3
	Porcentaje global				84,8

Cuadro 2.8

Variables en la ecuación

		B	E.T.	Wald	gl
Paso 1[a]	Edad_Intervalo(1)	-,104	1,016	,011	1
	Genero(1)	,161	,761	,045	1
	TipodeColegio(1)	3,371	1,068	9,959	1
	Lugardepreparación			4,984	3
	Lugardepreparación(1)	-1,016	,952	1,139	1
	Lugardepreparación(2)	-2,572	1,185	4,707	1
	Lugardepreparación(3)	-1,038	1,588	,428	1
	Preferenciaporlacarrera(1)	-2,492	,767	10,553	1
	Trabajaenlaactualida(1)	2,794	1,387	4,060	1
	Tiempodeestudio			,118	2
	Tiempodeestudio(1)	22,279	16739,840	,000	1
	Tiempodeestudio(2)	21,939	16739,840	,000	1
	ResultadodelexamendeMateBasica(1)	1,020	,880	1,346	1
	Conquienvive			7,062	3
	Conquienvive(1)	-1,157	1,103	1,099	1
	Conquienvive(2)	-1,706	1,553	1,207	1
	Conquienvive(3)	-3,630	1,529	5,640	1
	Nivelacademicodelpadre			4,054	4
	Nivelacademicodelpadre(1)	-20,153	40193,091	,000	1
	Nivelacademicodelpadre(2)	-24,453	40193,091	,000	1
	Nivelacademicodelpadre(3)	-23,670	40193,091	,000	1
	Nivelacademicodelpadre(4)	-24,314	40193,091	,000	1
	Nivelacademicodelamadre			3,708	4
	Nivelacademicodelamadre(1)	-,454	2,034	,050	1
	Nivelacademicodelamadre(2)	2,945	2,311	1,624	1
	Nivelacademicodelamadre(3)	2,424	2,355	1,060	1
	Nivelacademicodelamadre(4)	2,374	2,556	,863	1
	Constante	-2,035	43539,736	,000	1

Resultados del análisis del curso de Cálculo I

Cuadro 2.9

Resumen del procesamiento de los casos

Casos no ponderados[a]		N	Porcentaje
Casos seleccionados	Incluidos en el análisis	105	100,0
	Casos perdidos	0	,0
	Total	105	100,0
Casos no seleccionados		0	,0
Total		105	100,0

Bloque 0: Bloque inicial

Cuadro 2.10

Historial de iteraciones[a,b,c]

Iteración		-2 log de la verosimilitud	Coeficientes Constant
Paso 0	1	136,279	-,590
	2	136,271	-,609
	3	136,271	-,609

Cuadro 2.11

Tabla de clasificación[a,b]

Observado			Pronosticado		
			Resultado final cálculo		Porcentaje correcto
			Desaprobado	Aprobado	
Paso 0	Resultado final cálculo	Desaprobado	68	0	100,0
		Aprobado	37	0	,0
	Porcentaje global				64,8

Cuadro 2.12

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 0	Constante	-,609	,204	8,875	1	,003	,544

Bloque 1: Método = Introducir

Cuadro 2.13

Pruebas omnibus sobre los coeficientes del modelo

		Chi cuadrado	gl	Sig.
Paso 1	Paso	68,930	22	,000
	Bloque	68,930	22	,000
	Modelo	68,930	22	,000

Cuadro 2.14

Resumen del modelo

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	67,340[a]	,481	,662

Cuadro 2.15

Prueba de Hosmer y Lemeshow

Paso	Chi cuadrado	gl	Sig.
1	14,218	8	,076

Cuadro 2.16

Tabla de clasificación[a]

Observado			Pronosticado		
			Resultado final cálculo		Porcentaje correcto
			Desaprobado	Aprobado	
Paso 1	Resultado final cálculo	Desaprobado	64	4	94,1
		Aprobado	7	30	81,1
	Porcentaje global				89,5

a. El valor de corte es ,500

Cuadro 2.17

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1a	Genero(1)	,595	,691	,743	1	,389	1,813
	Edad_A(1)	-,660	1,150	,329	1	,566	,517
	Tipodecolegio(1)	3,586	1,029	12,135	1	,000	36,078
	Lugardepreparación			3,948	3	,267	
	Lugardepreparación(1)	1,104	,879	1,577	1	,209	3,015
	Lugardepreparación(2)	-2,689	2,161	1,549	1	,213	,068
	Lugardepreparación(3)	-1,073	1,487	,521	1	,470	,342
	Preferenciaporlacarrera(1)	-1,343	,691	3,775	1	,052	,261
	Resultadodelexamendecálculo(1)	-,356	,745	,229	1	,633	,700
	Trabajaenlaactualida(1)	1,792	1,483	1,462	1	,227	6,004
	Tiempodeestudio			2,968	2	,227	
	Tiempodeestudio(1)	51,947	17214,390	,000	1	,998	
	Tiempodeestudio(2)	53,709	17214,390	,000	1	,998	
	Conquienvive			1,713	3	,634	
	Conquienvive(1)	-1,373	1,083	1,609	1	,205	,253
	Conquienvive(2)	-35,299	13863,075	,000	1	,998	,000
	Conquienvive(3)	-,542	,966	,316	1	,574	,581
	Nivelacademicodelpadre			1,935	4	,748	
	Nivelacademicodelpadre(1)	-2,695	40727,179	,000	1	1,000	,068
	Nivelacademicodelpadre(2)	-37,632	40872,809	,000	1	,999	,000
	Nivelacademicodelpadre(3)	-37,811	40872,809	,000	1	,999	,000
	Nivelacademicodelpadre(4)	-39,091	40872,809	,000	1	,999	,000
	Nivelacademicodelamadre			2,679	4	,613	
	Nivelacademicodelamadre(1)	-17,598	6575,005	,000	1	,998	,000
	Nivelacademicodelamadre(2)	16,667	7423,915	,000	1	,998	
	Nivelacademicodelamadre(3)	16,566	7423,915	,000	1	,998	
	Nivelacademicodelamadre(4)	18,452	7423,915	,000	1	,998	
	Constante	-33,222	43724,236	,000	1	,999	,000

ANEXO 3

Resultados utilizando el método Wald pasos hacia adelante-Matemática Básica

Cuadro 3.1

Resumen del procesamiento de los casos

Casos no ponderados[a]		N	Porcentaje
Casos seleccionados	Incluidos en el análisis	105	100,0
	Casos perdidos	0	,0
	Total	105	100,0
Casos no seleccionados		0	,0
Total		105	100,0

Cuadro 3.2

Codificación de la variable dependiente

Valor original	Valor interno
Desaprobado	0
Aprobado	1

Bloque 0: Bloque inicial

Cuadro 3.3

Historial de iteraciones[a,b,c]

Iteración		-2 log de la verosimilitud	Coeficientes Constant
Paso 0	1	144,789	-,171
	2	144,789	-,172

Cuadro 3.4

Tabla de clasificación[a,b]

Observado			Pronosticado		
			Resultado final Matemática Básica		Porcentaje correcto
			Desaprobado	Aprobado	
Paso 0	Resultado final Matemática Básica	Desaprobado	57	0	100,0
		Aprobado	48	0	,0
	Porcentaje global				54,3

Cuadro 3.5

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 0	Constante	-,172	,196	,770	1	,380	,842

Cuadro 3.6

Variables que no están en la ecuación

			Puntuación	gl	Sig.
Paso 0	Variables	Edad_Intervalo(1)	,053	1	,818
		Genero(1)	1,036	1	,309
		TipodeColegio(1)	18,737	1	,000
		Lugardepreparación	4,030	3	,258
		Lugardepreparación(1)	,161	1	,688
		Lugardepreparación(2)	3,839	1	,050
		Lugardepreparación(3)	,069	1	,793
		Preferenciaporlacarrera(1)	12,431	1	,000
		Trabajaenlaactualida(1)	6,430	1	,011
		Tiempodeestudio	4,108	2	,128
		Tiempodeestudio(1)	,014	1	,905
		Tiempodeestudio(2)	,850	1	,356
		ResultadodelexamendeMateBasica(1)	2,594	1	,107
		Conquienvive	6,252	3	,100
		Conquienvive(1)	,314	1	,575
		Conquienvive(2)	2,164	1	,141
		Conquienvive(3)	2,558	1	,110
		Nivelacademicodelpadre	3,467	4	,483
		Nivelacademicodelpadre(1)	,869	1	,351
		Nivelacademicodelpadre(2)	2,289	1	,130
		Nivelacademicodelpadre(3)	,161	1	,688
		Nivelacademicodelpadre(4)	,230	1	,631
		Nivelacademicodelamadre	,893	4	,926
		Nivelacademicodelamadre(1)	,651	1	,420
		Nivelacademicodelamadre(2)	,356	1	,551
		Nivelacademicodelamadre(3)	,053	1	,818
		Nivelacademicodelamadre(4)	,029	1	,864
	Estadísticos globales		52,524	22	,000

Bloque 1: Método = Por pasos hacia adelante (Wald)

Cuadro 3.7

Pruebas omnibus sobre los coeficientes del modelo

		Chi cuadrado	gl	Sig.
Paso 1	Paso	19,617	1	,000
	Bloque	19,617	1	,000
	Modelo	19,617	1	,000
Paso 2	Paso	14,148	1	,000
	Bloque	33,765	2	,000
	Modelo	33,765	2	,000
Paso 3	Paso	5,482	1	,019
	Bloque	39,248	3	,000
	Modelo	39,248	3	,000
Paso 4	Paso	10,099	3	,018
	Bloque	49,347	6	,000
	Modelo	49,347	6	,000

Cuadro 3.8

Resumen del modelo

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	125,171[a]	,170	,228
2	111,023[a]	,275	,368
3	105,541[b]	,312	,417
4	95,442[c]	,375	,501

Cuadro 3.9

Prueba de Hosmer y Lemeshow

Paso	Chi cuadrado	gl	Sig.
1	,000	0	.
2	5,111	2	,078
3	8,392	4	,078
4	7,503	6	,277

Cuadro 3.10

Tabla de contingencias para la prueba de Hosmer y Lemeshow

		Resultado final Matemática Básica = Desaprobado		Resultado final Matemática Básica = Aprobado		
		Observado	Esperado	Observado	Esperado	Total
Paso 1	1	52	52,000	26	26,000	78
	2	5	5,000	22	22,000	27
Paso 2	1	43	41,562	9	10,438	52
	2	9	10,438	17	15,562	26
	3	3	4,438	14	12,562	17
	4	2	,562	8	9,438	10
Paso 3	1	11	10,427	0	,573	11
	2	33	32,070	9	9,930	42
	3	1	1,309	1	,691	2
	4	8	9,503	17	15,497	25
	5	2	3,281	14	12,719	16
	6	2	,410	7	8,590	9
Paso 4	1	11	10,677	0	,323	11
	2	11	10,371	0	,629	11
	3	4	4,019	1	,981	5
	4	19	19,054	9	8,946	28
	5	3	5,358	7	4,642	10
	6	5	5,791	14	13,209	19
	7	4	1,613	10	12,387	14
	8	0	,117	7	6,883	7

Cuadro 3.11

Tabla de clasificación[a]

Observado			Pronosticado		
			Resultado final Matemática Básica		Porcentaje correcto
			Desaprobado	Aprobado	
Paso 1	Resultado final Matemática Básica	Desaprobado	52	5	91,2
		Aprobado	26	22	45,8
	Porcentaje global				70,5
Paso 2	Resultado final Matemática Básica	Desaprobado	43	14	75,4
		Aprobado	9	39	81,3
	Porcentaje global				78,1
Paso 3	Resultado final Matemática Básica	Desaprobado	45	12	78,9
		Aprobado	10	38	79,2
	Porcentaje global				79,0
Paso 4	Resultado final Matemática Básica	Desaprobado	47	10	82,5
		Aprobado	15	33	68,8
	Porcentaje global				76,2

a. El valor de corte es ,500

Cuadro 3.12

Variables que no están en la ecuación

			Puntuación	gl	Sig.
gPaso 1	Variables	Edad_Intervalo(1)	,654	1	,419
		Genero(1)	,629	1	,428
		Lugardepreparación	3,405	3	,333
		Lugardepreparación(1)	,376	1	,540
		Lugardepreparación(2)	2,727	1	,099
		Lugardepreparación(3)	,431	1	,512
		Preferenciaporlacarrera(1)	13,952	1	,000
		Trabajaenlaactualida(1)	6,755	1	,009
		Tiempodeestudio	2,726	2	,256
		Tiempodeestudio(1)	,002	1	,966
		Tiempodeestudio(2)	,808	1	,369
		ResultadodelexamendeMateBasica(1)	1,320	1	,251
		Conquienvive	9,215	3	,027
		Conquienvive(1)	,694	1	,405
		Conquienvive(2)	,813	1	,367
		Conquienvive(3)	6,114	1	,013
		Nivelacademicodelpadre	1,734	4	,784
		Nivelacademicodelpadre(1)	1,114	1	,291
		Nivelacademicodelpadre(2)	,765	1	,382
		Nivelacademicodelpadre(3)	,116	1	,733
		Nivelacademicodelpadre(4)	,031	1	,860
		Nivelacademicodelamadre	2,294	4	,682
		Nivelacademicodelamadre(1)	,493	1	,482
		Nivelacademicodelamadre(2)	,219	1	,640
		Nivelacademicodelamadre(3)	,962	1	,327
		Nivelacademicodelamadre(4)	1,138	1	,286
	Estadísticos globales		40,242	21	,007
Paso 2	Variables	Edad_Intervalo(1)	,498	1	,481
		Genero(1)	,719	1	,397
		Lugardepreparación	5,971	3	,113
		Lugardepreparación(1)	,049	1	,825
		Lugardepreparación(2)	4,370	1	,037
		Lugardepreparación(3)	,606	1	,436
		Trabajaenlaactualida(1)	4,768	1	,029
		Tiempodeestudio	3,702	2	,157
		Tiempodeestudio(1)	,121	1	,728
		Tiempodeestudio(2)	,513	1	,474
		ResultadodelexamendeMateBasica(1)	,630	1	,427
		Conquienvive	8,070	3	,045
		Conquienvive(1)	,847	1	,357
		Conquienvive(2)	,965	1	,326
		Conquienvive(3)	4,657	1	,031
		Nivelacademicodelpadre	2,955	4	,565
		Nivelacademicodelpadre(1)	2,164	1	,141
		Nivelacademicodelpadre(2)	1,158	1	,282
		Nivelacademicodelpadre(3)	,034	1	,854
		Nivelacademicodelpadre(4)	,022	1	,881
		Nivelacademicodelamadre	1,579	4	,813
		Nivelacademicodelamadre(1)	,528	1	,468
		Nivelacademicodelamadre(2)	,007	1	,932
		Nivelacademicodelamadre(3)	,638	1	,424
		Nivelacademicodelamadre(4)	,431	1	,512
	Estadísticos globales		30,876	20	,057

Variables que no están en la ecuación

			Puntuación	gl	Sig.
Paso 3	Variables	Edad_Intervalo(1)	,131	1	,717
		Genero(1)	,989	1	,320
		Lugardepreparación	6,396	3	,094
		Lugardepreparación(1)	,076	1	,782
		Lugardepreparación(2)	5,626	1	,018
		Lugardepreparación(3)	,001	1	,978
		Tiempodeestudio	3,872	2	,144
		Tiempodeestudio(1)	,092	1	,761
		Tiempodeestudio(2)	,530	1	,466
		ResultadodelexamendeMateBasica(1)	1,840	1	,175
		Conquienvive	9,109	3	,028
		Conquienvive(1)	,178	1	,673
		Conquienvive(2)	1,006	1	,316
		Conquienvive(3)	6,724	1	,010
		Nivelacademicodelpadre	2,245	4	,691
		Nivelacademicodelpadre(1)	1,335	1	,248
		Nivelacademicodelpadre(2)	1,411	1	,235
		Nivelacademicodelpadre(3)	,038	1	,846
		Nivelacademicodelpadre(4)	,053	1	,818
		Nivelacademicodelamadre	1,381	4	,848
		Nivelacademicodelamadre(1)	,615	1	,433
		Nivelacademicodelamadre(2)	,020	1	,888
		Nivelacademicodelamadre(3)	,624	1	,430
		Nivelacademicodelamadre(4)	,314	1	,575
	Estadísticos globales		26,640	19	,113
Paso 4	Variables	Edad_Intervalo(1)	,094	1	,759
		Genero(1)	,680	1	,410
		Lugardepreparación	6,992	3	,072
		Lugardepreparación(1)	,008	1	,930
		Lugardepreparación(2)	6,131	1	,013
		Lugardepreparación(3)	,224	1	,636
		Tiempodeestudio	5,025	2	,081
		Tiempodeestudio(1)	,207	1	,649
		Tiempodeestudio(2)	,407	1	,523
		ResultadodelexamendeMateBasica(1)	,735	1	,391
		Nivelacademicodelpadre	2,974	4	,562
		Nivelacademicodelpadre(1)	1,141	1	,285
		Nivelacademicodelpadre(2)	1,178	1	,278
		Nivelacademicodelpadre(3)	,117	1	,732
		Nivelacademicodelpadre(4)	,022	1	,883
		Nivelacademicodelamadre	2,007	4	,734
		Nivelacademicodelamadre(1)	1,486	1	,223
		Nivelacademicodelamadre(2)	,208	1	,648
		Nivelacademicodelamadre(3)	,278	1	,598
		Nivelacademicodelamadre(4)	,262	1	,609
	Estadísticos globales		20,248	16	,209

Resultados utilizando el método Wald pasos hacia adelante-Cálculo I

Cuadro 3.13

Casos no ponderados[a]		N	Porcentaje
Casos seleccionados	Incluidos en el análisis	105	100,0
	Casos perdidos	0	,0
	Total	105	100,0
Casos no seleccionados		0	,0
Total		105	100,0

Cuadro 3.14

Codificación de la variable dependiente

Valor original	Valor interno
Desaprobado	0
Aprobado	1

Bloque 0: Bloque inicial

Cuadro 3.15

Historial de iteraciones[a,b,c]

Iteración		-2 log de la verosimilitud	Coeficientes
			Constant
Paso 0	1	136,279	-,590
	2	136,271	-,609
	3	136,271	-,609

Cuadro 3.16

Observado			Pronosticado		
			Resultado final cálculo		Porcentaje correcto
			Desaprobado	Aprobado	
Paso 0	Resultado final cálculo	Desaprobado	68	0	100,0
		Aprobado	37	0	,0
	Porcentaje global				64,8

Cuadro 3.17

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 0	Constante	-,609	,204	8,875	1	,003	,544

Cuadro 3.18

Variables que no están en la ecuación

			Puntuación	gl	Sig.
Paso 0	Variables	Genero(1)	,003	1	,953
		Edad_A(1)	,876	1	,349
		Tipodecolegio(1)	24,021	1	,000
		Lugardepreparación	8,677	3	,034
		Lugardepreparación(1)	3,930	1	,047
		Lugardepreparación(2)	5,587	1	,018
		Lugardepreparación(3)	,534	1	,465
		Preferenciaporlacarrera(1)	3,448	1	,063
		Resultadodelexamendecálculo	,282	1	,595
		Trabajaenlaactualida(1)	5,587	1	,018
		Tiempodeestudio	5,258	2	,072
		Tiempodeestudio(1)	,807	1	,369
		Tiempodeestudio(2)	3,396	1	,065
		Conquienvive	4,296	3	,231
		Conquienvive(1)	,130	1	,719
		Conquienvive(2)	3,463	1	,063
		Conquienvive(3)	1,002	1	,317
		Nivelacademicodelpadre	3,280	4	,512
		Nivelacademicodelpadre(1)	1,294	1	,255
		Nivelacademicodelpadre(2)	,439	1	,508
		Nivelacademicodelpadre(3)	,035	1	,853
		Nivelacademicodelpadre(4)	,049	1	,824
		Nivelacademicodelamadre	3,784	4	,436
		Nivelacademicodelamadre(1)	1,350	1	,245
		Nivelacademicodelamadre(2)	,349	1	,555
		Nivelacademicodelamadre(3)	1,081	1	,299
		Nivelacademicodelamadre(4)	1,802	1	,179
	Estadísticos globales		49,615	22	,001

Bloque 1: Método = Por pasos hacia adelante (Wald)

Cuadro 3.19

Pruebas omnibus sobre los coeficientes del modelo

		Chi cuadrado	gl	Sig.
Paso 1	Paso	23,577	1	,000
	Bloque	23,577	1	,000
	Modelo	23,577	1	,000
Paso 2	Paso	7,696	1	,006
	Bloque	31,273	2	,000
	Modelo	31,273	2	,000
Paso 3	Paso	10,450	3	,015
	Bloque	41,723	5	,000
	Modelo	41,723	5	,000

Cuadro 3.20

Resumen del modelo

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	112,694[a]	,201	,277
2	104,998[b]	,258	,354
3	94,548[c]	,328	,451

Cuadro 3.22

Prueba de Hosmer y Lemeshow

Paso	Chi cuadrado	gl	Sig.
1	,000	0	.
2	,451	2	,798
3	1,990	5	,851

Cuadro 3.23

Tabla de contingencias para la prueba de Hosmer y Lemeshow

		Resultado final cálculo = Desaprobado		Resultado final cálculo = Aprobado		
		Observado	Esperado	Observado	Esperado	Total
Paso 1	1	61	61,000	17	17,000	78
	2	7	7,000	20	20,000	27
Paso 2	1	11	10,723	0	,277	11
	2	2	2,277	1	,723	3
	3	50	50,277	17	16,723	67
	4	5	4,723	19	19,277	24
Paso 3	1	10	9,791	0	,209	10
	2	12	11,638	0	,362	12
	3	1	,942	0	,058	1
	4	31	30,329	9	9,671	40
	5	4	5,140	3	1,860	7
	6	6	6,787	7	6,213	13
	7	4	3,372	18	18,628	22

Cuadro 3.24

Tabla de clasificación[a]

Observado			Pronosticado		
			Resultado finalcálculo		Porcentaje correcto
			Desaprobado	Aprobado	
Paso 1	Resultado final calculo	Desaprobado	61	7	89,7
		Aprobado	17	20	54,1
	Porcentaje global				77,1
Paso 2	Resultado final cálculo	Desaprobado	63	5	92,6
		Aprobado	18	19	51,4
	Porcentaje global				78,1
Paso 3	Resultado final cálculo	Desaprobado	64	4	94,1
		Aprobado	19	18	48,6
	Porcentaje global				78,1

Cuadro 3.25

Variables que no están en la ecuación

			Puntuación	gl	Sig.
Paso 1	Variables	Genero(1)	,207	1	,649
		Edad_A(1)	,019	1	,890
		Lugardepreparación	10,430	3	,015
		Lugardepreparación(1)	6,081	1	,014
		Lugardepreparación(2)	4,597	1	,032
		Lugardepreparación(3)	1,521	1	,217
		Preferenciaporlacarrera(1)	3,702	1	,054
		Resultadodelexámendecálculol	,438	1	,508
		Trabajaenlaactualida(1)	6,173	1	,013
		Tiempodeestudio	4,702	2	,095
		Tiempodeestudio(1)	1,509	1	,219
		Tiempodeestudio(2)	3,815	1	,051
		Conquienvive	2,429	3	,488
		Conquienvive(1)	,425	1	,514
		Conquienvive(2)	1,811	1	,178
		Conquienvive(3)	,157	1	,692
		Nivelacademicodelpadre	2,900	4	,575
		Nivelacademicodelpadre(1)	1,765	1	,184
		Nivelacademicodelpadre(2)	,035	1	,851
		Nivelacademicodelpadre(3)	,110	1	,740
		Nivelacademicodelpadre(4)	1,165	1	,280
		Nivelacademicodelamadre	1,718	4	,787
		Nivelacademicodelamadre(1)	1,236	1	,266
		Nivelacademicodelamadre(2)	,202	1	,653
		Nivelacademicodelamadre(3)	,111	1	,739
		Nivelacademicodelamadre(4)	,309	1	,578
	Estadísticos globales		31,460	20	,049

Variables que no están en la ecuación

			Puntuación	gl	Sig.
Paso 2	Variables	Genero(1)	,253	1	,615
		Edad_A(1)	,013	1	,909
		Lugardepreparación	9,090	3	,028
		Lugardepreparación(1)	4,924	1	,026
		Lugardepreparación(2)	6,072	1	,014
		Lugardepreparación(3)	,000	1	,993
		Preferenciaporlacarrera(1)	2,680	1	,102
		Resultadodelexámendecálculol	1,067	1	,302
		Tiempodeestudio	3,973	2	,137
		Tiempodeestudio(1)	1,255	1	,263
		Tiempodeestudio(2)	3,147	1	,076
		Conquienvive	1,946	3	,584
		Conquienvive(1)	,031	1	,860
		Conquienvive(2)	1,823	1	,177
		Conquienvive(3)	,002	1	,964
		Nivelacademicodelpadre	1,605	4	,808
		Nivelacademicodelpadre(1)	,964	1	,326
		Nivelacademicodelpadre(2)	,012	1	,913
		Nivelacademicodelpadre(3)	,151	1	,698
		Nivelacademicodelpadre(4)	,490	1	,484
		Nivelacademicodelamadre	2,239	4	,692
		Nivelacademicodelamadre(1)	1,517	1	,218
		Nivelacademicodelamadre(2)	,282	1	,595
		Nivelacademicodelamadre(3)	,186	1	,666
		Nivelacademicodelamadre(4)	,637	1	,425
	Estadísticos globales		26,948	19	,106
Paso 3	Variables	Genero(1)	,075	1	,784
		Edad_A(1)	,001	1	,979
		Preferenciaporlacarrera(1)	2,486	1	,115
		Resultadodelexámendecálculol	1,583	1	,208
		Tiempodeestudio	5,106	2	,078
		Tiempodeestudio(1)	2,793	1	,095
		Tiempodeestudio(2)	4,679	1	,031
		Conquienvive	1,806	3	,614
		Conquienvive(1)	,030	1	,862
		Conquienvive(2)	1,464	1	,226
		Conquienvive(3)	,151	1	,697
		Nivelacademicodelpadre	1,141	4	,888
		Nivelacademicodelpadre(1)	,774	1	,379
		Nivelacademicodelpadre(2)	,079	1	,779
		Nivelacademicodelpadre(3)	,005	1	,944
		Nivelacademicodelpadre(4)	,267	1	,605
		Nivelacademicodelamadre	2,675	4	,614
		Nivelacademicodelamadre(1)	1,527	1	,217
		Nivelacademicodelamadre(2)	,057	1	,811
		Nivelacademicodelamadre(3)	,199	1	,655
		Nivelacademicodelamadre(4)	1,138	1	,286
	Estadísticos globales		19,846	16	,227

ANEXO 4

UNMSM -Facultad de Ciencias Matemáticas

CUESTIONARIO SOBRE CARACTERÍSTICAS EDUCATIVAS DE LOS ALUMNOS

Estimado alumno, usted ha sido elegido al azar para participar en una investigación sobre algunas características socioeducativas de los estudiantes de la FCM. Agradecemos de antemano su valiosa colaboración.

DATOS

1. Edad: _____ 2.Genero: 1.Masculino 2 Femenino

3. Código de matrícula: _____4. Escuela_____

5. Año de ingreso a la universidad: _____

6. Nombre y Apellido:

_____

CARACTERÍSTICAS EDUCATIVAS

1. ¿Qué año termino el colegio? : _____

2. ¿En qué tipo de colegio estudiaste?

Publico/Estatal	1
Particular	2
Preuniversitario	3
Militar o Religioso	4

3. ¿Dónde te preparaste para ingresar a la universidad?

Academia	1
Pre San Marcos	2
Casa	3
Grupo de Estudio	4

4. ¿Mediante qué tipo de modalidad de examen de admisión ingresaste?

Examen general	1
Primeros puestos	2
Pre San Marcos	3
Traslado externo	4
Deportista calificado	5

5. ¿La carrera que estudias en la actualidad era de tu preferencia para estudiar?

1. SI 2. NO

6. ¿Postulaste a otras carreras antes de ingresar a la que estudias en la actualidad?

1. SI **2. NO**

7. ¿cuantas veces postulaste a otras carreras antes de ingresar a la que estudias actualmente?

N° de veces:	_____

12

8. ¿Estudias alguna otra carrera universitaria simultáneamente en la actualidad?

1. SI **2. NO**

9. ¿Estudias alguna otra carrera universitaria no universitaria (Instituto, curso, etc.) simultáneamente en la actualidad?

1. SI **2. NO**

10. ¿Trabajas en la actualidad?

1. SI **2. NO**

11. ¿Cuánto tiempo te toma llegar a la universidad?

Menos de 30 min	1
Entre 30 min y 1 hora	2
Entre 1 hora y 1 hora 30 min	3
Entre 1 hora 30 min y 2 horas	4
Entre 2 horas y 2 horas 30 min	5
Más de 2 horas 30 min.	6

12. ¿En qué lugar ingieres tus alimentos?

Comedor universitario.	1
Casa.	2
Restaurante.	3
Comida propia que trae de su casa	4

13. ¿En qué lugar estudias después de clases?

1. Casa 2.Biblioteca

14. ¿Cuánto tiempo estudias un curso después de clases?

Menos de una hora al día	1
Entre una hora y tres horas al día.	2
Más de tres horas	3

15. ¿Utilizas bibliografía adicional para estudiar un curso?

1. SI **2. NO**

16. ¿Cómo le fue en el último examen parcial de Matemática Básica?

1. Aprobado **2. Desaprobado**

17. ¿Cómo le fue en el último examen parcial de Cálculo 1? (si está llevando el curso)

1. Aprobado **2. Desaprobado**

CARACTERÍSTICAS EN EL HOGAR

18. La vivienda donde vives es:

1. Propia **2.Alquilada**

19. Donde se ubica su vivienda.

Urbanización.	1
Condominio.	2
Quinta.	3
AA.HH.	4

20. ¿Cuenta con computadora en casa?

1. Si **2.No**

21. ¿Cuenta con servicio de internet en su vivienda?

1. Si **2.No**

22. ¿Con quién vive usted?

Con padres y hermanos.	1
Solo con Padre./madre	2
Solo.	7
Con otros familiares.	8

23. ¿Actualmente trabaja su padre?

1. Si **2.No**

24. ¿Actualmente trabaja su madre?

1. Si **2.No**

25. ¿Cuál es el nivel académico de su padre?

Sin Nivel.	1
Primaria Completa/Primaria incompleta.	2
Secundaria incompleta.	3
Secundaria completa.	4
Técnico incompleto.	5
Técnico Completo.	6
Universitario incompleto.	7
Universitario completo.	8

26. ¿Cuál es el nivel académico de su madre?

Sin Nivel.	1
Primaria Completa/Primaria incompleta.	2
Secundaria incompleta.	3
Secundaria completa.	4
Técnico incompleto.	5
Técnico Completo.	6
Universitario incompleto.	7
Universitario completo.	8

Gracias