

**ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL**



**FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICAS  
DEPARTAMENTO DE POSTGRADO**

**PROYECTO DE TITULACIÓN**

PREVIO A LA OBTENCIÓN DEL TÍTULO DE:

**“MAGÍSTER EN ESTADÍSTICA CON MENCIÓN EN  
GESTIÓN DE LA CALIDAD Y PRODUCTIVIDAD”**

**TEMA:**

DISEÑO DE UN MODELO DE PREDICCIÓN DEL RENDIMIENTO  
ACADÉMICO DE LOS ESTUDIANTES DE PRIMER NIVEL DE LAS  
CARRERAS DE INGENIERÍA DE LA ESPOL.

**AUTOR:**

JOHANNA DEL ROCÍO BARZOLA SARMIENTO

Guayaquil - Ecuador

2019

## **DEDICATORIA**

A mis padres Jaime y Anita, por su amor, comprensión y apoyo incondicional en cada una de mis metas propuestas, todo es por ellos y para ellos.

A mis hermanos y sus esposas por estar siempre presentes, acompañándome en estos valiosos momentos.

A mi sobrina Amelia por esos hermosos momentos de felicidad que me brinda y me inspira a mejorar el mundo que se enfrentará.

## **AGRADECIMIENTO**

A Dios por darme vida y permitirme alcanzar esta nueva meta.

A mis padres por siempre apoyarme y animarme a mejor como persona.

A la ESPOL por las facilidades brindadas para realizar esta investigación y sobre todo proporcionarme los conocimientos necesarios.

A mi tutor el PhD. Omar Ruíz por aceptar ser una vez más mi guía en esta nueva etapa académica, por su paciencia y conocimientos.

## DECLARACIÓN EXPRESA

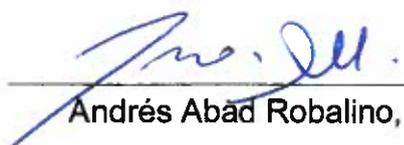
La responsabilidad por los hechos y doctrinas expuestas en este Proyecto de Graduación, me corresponde exclusivamente; el patrimonio intelectual del mismo, corresponde exclusivamente a la **Facultad de Ciencias Naturales y Matemáticas, Departamento de Postgrado** de la Escuela Superior Politécnica del Litoral.



---

Ing. Johanna del Rocío Barzola Sarmiento

## TRIBUNAL DE GRADUACIÓN

  
\_\_\_\_\_  
Andrés Abad Robalino, Ph.D.  
Presidente

  
\_\_\_\_\_  
Omar Ruíz Barzola, Ph.D.  
Director

  
\_\_\_\_\_  
Sandra García Bustos, Ph.D.  
Vocal del Tribunal 1

  
\_\_\_\_\_  
Mgtr. Mario Solórzano Carvajal  
Vocal del Tribunal 2

## **AUTOR DEL PROYECTO**

*Johanna Barzola S.*

---

Ing. Johanna del Rocío Barzola Sarmiento

# ÍNDICE GENERAL

ÍNDICE DE TABLAS.....	vii
ÍNDICE DE GRÁFICOS.....	viii
ABREVIATURAS O SIGLAS.....	xi
PRESENTACIÓN.....	xii
<b>CAPÍTULO 1</b> .....	<b>1</b>
<b>INTRODUCCIÓN</b> .....	<b>1</b>
1.1. Antecedentes.....	1
1.2. Planteamiento del problema .....	2
1.3. Justificación.....	3
1.4. Alcance .....	4
1.5. Objetivos del Proyecto.....	4
1.5.1. Objetivo General.....	4
1.5.2. Objetivos Específicos .....	5
<b>CAPÍTULO 2</b> .....	<b>6</b>
<b>MARCO TEÓRICO Y METODOLÓGICO</b> .....	<b>6</b>
2.1. Marco Teórico.....	6
2.1.1. Consejerías Académicas .....	6
2.1.2. Rendimiento académico .....	8
2.1.3. Estado del Arte .....	11
2.2. Marco Metodológico .....	14
2.2.1. Datos .....	14
2.2.2. Variables .....	14
2.2.3. Regresión lineal múltiple .....	16
2.2.4. Árboles de clasificación CHAID.....	17
<b>CAPÍTULO 3</b> .....	<b>19</b>
<b>RESULTADOS</b> .....	<b>19</b>
3.1. Análisis descriptivo.....	19
3.1.1. Características Sociodemográficas .....	19

3.1.6. Materias del proceso de Admisiones .....	40
3.1.7. Materias registradas en el término académico .....	43
3.2. Índice del grado de dificultad del término académico. ....	46
3.3. Aplicación de Algoritmo CHAID. ....	52
3.4. Matriz de correlación .....	54
3.5. Aplicación de Modelos de Regresión Multivariante. ....	58
3.5.1. Modelo original .....	59
3.5.2. Modelo Final.....	65
3.5.3. Resumen de los modelos .....	71
<b>CAPÍTULO 4</b> .....	<b>75</b>
<b>CONCLUSIONES Y RECOMENDACIONES</b> .....	<b>75</b>
4.1. Conclusiones .....	75
4.2. Recomendaciones .....	76
<b>REFERENCIAS BIBIOGRÁFICAS</b> .....	<b>77</b>
<b>ANEXOS</b> .....	<b>81</b>
i. Código SPSS: Algoritmo de CHAID. ....	81
ii. Código R: Estadística Univariada .....	82
iii. Código R: Estadística Bivariada .....	88
iv. Código R: Matriz de Correlación.....	90
v. Código R: Modelo Regresión Lineal Multivariada .....	90

## ÍNDICE DE TABLAS

Tabla 1. Descripción de las variables de análisis .....	14
Tabla 2. Estadística descriptiva: Edad .....	21
Tabla 3. Estadística descriptiva: Ingresos en el hogar .....	27
Tabla 4. Estadística descriptiva: Número de materias registradas.....	43
Tabla 5. Estadística descriptiva: Número de materias aprobadas .....	44
Tabla 6. Estadística descriptiva: Porcentaje de materias aprobadas .....	45
Tabla 7. Tasa de reprobación de las materias por término académico .....	48
Tabla 8. Estadística descriptiva: Porcentaje de materias aprobadas .....	51

Tabla 9. CHAID: Tabla de Clasificación .....	53
Tabla 10. Matriz de Correlación y valor p .....	56
Tabla 11. Variables a Incluir en el modelo de regresión multivariada .....	58
Tabla 12. Método de mínimos cuadrados: Modelo Original .....	60
Tabla 13. Análisis de varianza del Modelo Original .....	61
Tabla 14. Modelo Original: Factor de Inflación de Varianza (VIF).....	64
Tabla 15. Método de mínimos cuadrados: Modelo Final.....	66
Tabla 16. Análisis de varianza del Modelo Final .....	67
Tabla 17. Factor de Inflación de Varianza (VIF).....	70
Tabla 18. Comparación de los modelos .....	72

## ÍNDICE DE GRÁFICOS

Gráfico 1. Diagrama de barra: Género .....	19
Gráfico 2. Boxplot: Materias aprobadas vs. Género .....	20
Gráfico 3. Histograma, Boxplot, Normal Q-Q Plot: Edad.....	20
Gráfico 4. Diagrama de barra: Estado Civil .....	21
Gráfico 5. Diagrama de barra: Autoidentificación étnica .....	22
Gráfico 6. Diagrama de barra: Estado civil familiar.....	22
Gráfico 7. Boxplot: Materias aprobadas vs. Estado Civil familiar .....	23
Gráfico 8. Diagrama de barra: Grupo que reside .....	24
Gráfico 9. Boxplot: Materias aprobadas vs. Reside con la familia .....	24
Gráfico 10. Diagrama de barra: Parentesco con el jefe del hogar .....	25
Gráfico 11. Boxplot: Materias aprobadas vs. El padre es el jefe del hogar .....	25
Gráfico 12. Diagrama de barra: Nivel de instrucción del jefe del hogar .....	26
Gráfico 13. Boxplot: Mat. aprobadas vs. Jefe del hogar tiene tercer nivel de instrucción .....	26
Gráfico 14. Histograma, Boxplot, Normal Q-Q Plot: Ingresos en el hogar .....	28
Gráfico 15. Diagrama de barra: Estrato Socioeconómico .....	28
Gráfico 16. Boxplot: Materias aprobadas vs. Estrato Socioeconómico .....	29

Gráfico 17. Diagrama de barra: Número de Celulares en el Hogar.....	29
Gráfico 18. Boxplot: Materias aprobadas vs. Número de Celulares en el Hogar .....	30
Gráfico 19. Diagrama de barra: Número de televisores en el hogar .....	30
Gráfico 20. Boxplot: Materias aprobadas vs. Número de televisores en el hogar .....	31
Gráfico 21. Diagrama de barra: Número de vehículos en el hogar .....	31
Gráfico 22. Boxplot: Materias aprobadas vs. Tiene vehículo en el hogar .....	32
Gráfico 23. Diagrama de barra: Tipo de vivienda .....	32
Gráfico 24. Boxplot: Materias aprobadas vs. Tipo de vivienda .....	33
Gráfico 25. Diagrama de barra: Forma de pago de la vivienda .....	33
Gráfico 26. Boxplot: Materias aprobadas vs. Forma de pago de la vivienda ...	34
Gráfico 27. Diagrama de barra: Tecnologías de Información y Comunicación	35
Gráfico 28. Boxplot: Materias aprobadas vs. Tecnologías de Información y Comunicación.....	36
Gráfico 29. Diagrama de barra: Tipo de Colegio .....	37
Gráfico 30. Boxplot: Materias aprobadas vs. Tipo de Colegio .....	37
Gráfico 31. Diagrama de barra: Colegio ubicado en Guayaquil.....	38
Gráfico 32. Boxplot: Materias aprobadas vs. Colegio ubicado en Guayaquil ...	38
Gráfico 33. Diagrama de barra: Familiares afiliados al IESS.....	39
Gráfico 34. Boxplot: Materias aprobadas vs. Familiares afiliados al IESS.....	39
Gráfico 35. Familiares afiliados a un seguro privado .....	40
Gráfico 36. Boxplot: Materias aprobadas vs. Familiares afiliados a un seguro privado .....	40
Gráfico 37. Física en el Proceso de Admisiones .....	41
Gráfico 38. Matemáticas en el Proceso de Admisiones .....	42
Gráfico 39. Calificación de la materia de Química en el Proceso de Admisiones .....	42
Gráfico 40. Número de materias registradas .....	43
Gráfico 41. Número de materias aprobadas .....	45
Gráfico 42. Porcentaje de materias aprobadas .....	46

Gráfico 43. Porcentaje de registrados en materias sociales por término académico .....	47
Gráfico 44. Porcentaje de registrados en materias STEM por término académico .....	47
Gráfico 45. Histograma, Diagrama de Caja y QQ-Normal: Grado de dificultad M. STEM .....	51
Gráfico 46. Histograma, Diagrama de Caja y QQ-Normal: Grado de dificultad M. Sociales .....	52
Gráfico 47. Diagrama de árbol: CHAID para modelar el éxito académico en el primer nivel de las carreras de ingeniería de la ESPOL.....	54
Gráfico 48. Corplot de las variables del modelo .....	55
Gráfico 49. Matriz de Dispersión, Histograma y Correlación .....	57
Gráfico 50. Normal Q-Q de los residuos del Modelo Original .....	61
Gráfico 51. Modelo Original del número de materias aprobadas .....	62
Gráfico 52. Correlograma de residuos del Modelo original .....	63
Gráfico 53. Influenceplot del Modelo original .....	64
Gráfico 54. Validación del Modelo original.....	65
Gráfico 55. Normal Q-Q de los residuos del Modelo Final .....	68
Gráfico 56. Modelo Final del número de materias aprobadas .....	68
Gráfico 57. Correlograma de residuos del Modelo Final .....	69
Gráfico 58. Influenceplot del Modelo Final.....	70
Gráfico 59. Validación del Modelo Final .....	71

## **ABREVIATURAS O SIGLAS**

ESPOL Escuela Superior Politécnica del Litoral

VRA Vicerrectorado Académico ESPOL

CEPROEM Centro de Promoción y Empleo ESPOL

STA Secretaría Técnica Académica ESPOL

INSOEC Encuesta de Inclusión Socio Económico

TICs Tecnologías de la Información y la Comunicación

STEM Science, Technology, Engineering and Mathematics

## PRESENTACIÓN

El rendimiento académico es un indicador de gran importancia en el ámbito educativo, dado que permite medir el nivel de aprendizaje que el estudiante alcanza durante un periodo académico.

Hay estudios que señalan que los estudiantes que inician con un bajo rendimiento generalmente continúan con un bajo rendimiento en los próximos semestres, no culminan sus estudios en el tiempo planificado y otros terminan desertando de sus carreras.

Dentro de ESPOL es importante analizar el rendimiento académico de los estudiantes de ingeniería, debido a que representan la mayor proporción de los estudiantes y por la naturaleza de la institución.

En ESPOL, los estudiantes cuentan con un consejero académico a lo largo de su carrera, el cual los orienta en las actividades que debería cumplir dependiendo del nivel de avance de la carrera.

En el caso de los estudiantes novatos, el consejero no cuenta con información histórica sobre el desempeño académico del estudiante para basar su criterio y guiar de manera correcta en la elección de sus materias.

En la actualidad, el proceso de consejerías en la ESPOL no cuenta con una aplicación informática o sistema que le permita al consejero académico hacer un pronóstico del rendimiento del estudiante basado en las materias que este desea registrarse, ni en características sociodemográficas del mismo.

Este estudio pretende desarrollar un modelo que pueda en un futuro cercano ser implementado en un sistema informático, para que el estudiante del primer

semestre de una carrera de ingeniería pueda ser mejor orientado por su consejero académico.

Este documento está conformado por cuatro capítulos que se detallan a continuación de manera breve:

En el primer capítulo, se presenta la introducción del estudio, la revisión de los antecedentes, la justificación del tema propuesto, el alcance y los objetivos.

En el segundo capítulo, se señala el marco teórico relacionado con este estudio, la importancia de las consejerías académicas, descripción de indicadores para medir el rendimiento académico y la revisión de estudios realizados por estos investigadores referente a este tema. Además, la metodología aplicada: la descripción las variables, las técnicas a utilizar y la descripción de un indicador que permite medir el grado de dificultad de las materias tomadas por el estudiante.

En el tercer capítulo, se realiza el análisis descriptivo de las variables que se pretenden investigar y su distribución al compararla con el número de materias aprobadas, además de la aplicación del Algoritmo de CHAID con la finalidad de definir perfiles de estudiantes que aprueban todas las materias y también mediante un modelo de regresión múltiple se pretende identificar qué factores influyen significativamente en el número de materias aprobadas.

Finalmente, en el último capítulo se presentan las conclusiones y recomendaciones de esta investigación.

# **CAPÍTULO 1**

## **INTRODUCCIÓN**

### **1.1. Antecedentes**

La Escuela Superior Politécnica del Litoral (ESPOL) fue creada como respuesta a la gran cantidad de demanda de educación científico-técnica en la Costa el 29 de octubre de 1958.

Desde sus inicios ha cumplido con su misión de formar profesionales de excelencia, fue la primera universidad pública de la región Costa calificada con categoría “A”, sus principales factores de éxito son la calidad de sus carreras y programas, su liderazgo nacional en ciencia y tecnología, su infraestructura física y tecnológica, el acceso a las TICs y la conectividad, entre otros.

Por esto, tuvo diferentes cambios en los procesos de reforma curricular para actualizar su quehacer académico a una enseñanza basada en competencias, y una educación centrada en el estudiante.

Debido a que los estudiantes están expuestos a constantes cambios nació la necesidad de brindarles una guía, lo que motivó la generación de consejerías académicas en 1991.

En el año 2008 se aprobó la primera versión del Instructivo General de Consejerías Académicas, donde se definió quienes están involucrados en el proceso y sus responsabilidades, además se establecieron indicadores de seguimiento.

En el año 2012, el Vicerrectorado Académico (VRA) emitió lineamientos para el uso del sistema de consejerías, donde definió condiciones que deberían cumplir los estudiantes que estaban obligados a asistir a sesiones de consejerías y el rol del consejero.

En el año 2014, el VRA agregó a las consejerías el rol de centralizador de información para los estudiantes y a partir de aquí se derivaba hacia otros servicios que brinda la institución.

La consejería académica en la ESPOL tiene una influencia positiva en el promedio general de calificaciones de los estudiantes que asisten a la misma, mientras que los que no asisten (52%) no presentan cambios en su rendimiento (Rodríguez, Bermúdez & Paredes, 2016).

## **1.2. Planteamiento del problema**

El proceso de consejerías en la ESPOL no cuenta con una aplicación informática que le permita hacer un pronóstico del rendimiento del estudiante basado en las materias que este desea registrarse, ni en características sociodemográficas del mismo.

Actualmente el consejero basa su guía y recomendaciones según el comportamiento que ha venido teniendo el estudiante en los periodos de estudios pasados, variables como el promedio de calificaciones, materias reprobadas o el nivel de avance de la carrera.

Sin embargo, en el caso de los estudiantes novatos no cuenta el consejero con información histórica sobre rendimiento académico para basar su criterio y guiar a este en la elección de sus materias.

El proceso de consejerías en la ESPOL para los estudiantes novatos se centra revisar simplemente las políticas, reglamentos, y prácticas habituales entre estudiantes, profesores y trabajadores.

### **1.3. Justificación**

Es importante analizar el rendimiento académico de los estudiantes de ingeniería, debido a que representa el mayor porcentaje de la población politécnica y además que debido a los cambios generados por el proceso de Reforma Curricular el primer nivel de las mallas es igual para todas estas carreras.

Marina Míguez (2008), en su tesis Doctoral aplicada a estudiantes de la Universidad de la República de Chile evidenció que en el primer año de estudios superiores es donde se muestran limitaciones en las estrategias de aprendizaje, carencia de hábitos de estudio, además de una elevada tendencia a la deserción. El estudiante que carece de buenos hábitos de estudio, tendrá como consecuencia un lento avance en las carreras, lo cual genera un bajo índice de egreso de las carreras.

En ESPOL, un estudiante con buen rendimiento académico tiene mayor probabilidad de ganar algún tipo de beca. Esta institución, en el periodo académico 2016 – 2017 entregó 1.778 becas a estudiantes que cumplen con los requisitos reglamentarios, el monto invertido en este rubro es de US\$ 1'488.748. De los cuales el 34,25% fueron becas por distinción académica o distinción de actividades científicas, tecnológicas, de innovación, culturales y artísticas. En relación con las ayudas económicas se beneficiaron a 1.703 estudiantes que realizan actividades de docencia, investigación y gestión, a quienes además se les otorgaron subsidios en los servicios de transporte y alimentación.

Además, se conoce que, la eficiencia terminal en la ESPOL en el año 2016 fue de 40,15%, lo que significó un incremento de 3,78 puntos porcentuales en relación con la tasa del 2015. La eficiencia terminal es la proporción de estudiantes que terminan sus estudios dentro del tiempo normativo o ideal establecido.

Los consejeros académicos podrían ofrecer a los estudiantes de los primeros niveles de ingeniería, a través de una aplicación informática, un pronóstico del número de materias aprobadas por el estudiante basado en sus características sociodemográficas, calificaciones del proceso de admisiones y las materias que pretende tomar, su grado de complejidad, etc.

## **1.4. Alcance**

El presente proyecto de titulación es parte de una investigación de los estudiantes de la ESPOL a las carreras de Ingeniería, que ingresaron a la Institución del término 2016- IIT al 2018 - IT.

Se estudian los factores socioeconómicos, calificaciones obtenidas en el proceso de admisiones y materias tomadas en el término académico de los estudiantes antes mencionados.

Se pretende que esta investigación pueda ser aplicada en el futuro en el proceso de Consejería Académica de la Institución con la finalidad pronosticar el rendimiento académico del estudiante.

Se propone el análisis del rendimiento académico mediante análisis descriptivo y la utilización de una de las técnicas del análisis multivariado, la regresión lineal múltiple y un análisis de segmentación jerárquica mediante CHAID.

## **1.5. Objetivos del Proyecto**

### **1.5.1. Objetivo General**

Diseñar un modelo estadístico predictivo que permita estimar el rendimiento académico de los estudiantes de primer nivel de las carreras de Ingeniería de la ESPOL.

### **1.5.2. Objetivos Específicos**

- Identificar qué factores socioeconómicos, calificaciones obtenidas en el proceso de admisiones y materias tomadas inciden en el rendimiento académico de los estudiantes.
- Analizar la relación entre el rendimiento académico y cada uno de los factores en una muestra aleatoria de la población estudiantil de primer nivel de las carreras de Ingeniería.
- Desarrollar un modelo que permita pronosticar el rendimiento académico y sirva de base para el proceso de consejerías a los estudiantes de primer nivel de las carreras de Ingeniería.

## **CAPÍTULO 2**

### **MARCO TEÓRICO Y METODOLÓGICO**

#### **2.1. Marco Teórico**

##### **2.1.1. Consejerías Académicas**

###### **2.1.1.1. ¿Qué son las consejerías académicas?**

La Consejería Académica es un recurso pedagógico, mecanismo y espacio de confianza y comunicación, donde se asigna un profesor a cada estudiante, el cual se desea que lo acompañe a lo largo de su carrera, y se preocupe por el cumplimiento de sus metas en su proceso de formación académica.

###### **2.1.1.2. Importancia de las consejerías académicas**

El proceso de consejería de los estudiantes en todos los niveles, no se trata solamente de que los jóvenes tengan la oportunidad de ingresar a los estudios universitarios, sino que además permanezcan en la institución a lo largo su carrera, y adquieran conocimientos sólidos y logren una vida profesional exitosa (Clinciu, 2013).

Para que un proceso efectivo de consejerías debería analizar diferentes factores que afectan el desarrollo del estudiante, los mismos que tienen diversas fuentes como: problemas financieros, fracasos académicos, problemas de salud, cambio en su estilo de vida, entre otras. (Elias, Ping & Abdullah, 2011)

### **2.1.1.3. Consejerías Académicas en ESPOL**

En la ESPOL existen cuatro períodos de consejerías establecidos durante un año académico, siendo uno antes del periodo de matriculación y uno después de la primera evaluación en cada término.

Todos los estudiantes en la ESPOL cuentan con un consejero académico, el mismo que es asignado desde el inicio de su carrera, y se procura que se mantenga el mismo durante los años de permanencia del estudiante en la institución.

Son obligatorias las consejerías para todos los estudiantes previo al periodo de matriculación del primer término del año académico. En los demás periodos, sólo es obligatoria para: los estudiantes novatos, con materias en segunda matrícula en el periodo y con promedio general de calificaciones menor a 7 sobre 10.

El consejero trata durante la sesión temas relacionados según el porcentaje de créditos aprobados de la malla de la carrera del estudiante. Para el caso de los estudiantes que recién ingresan a la institución se revisan los siguientes puntos:

- ✓ Entregar acta de grado o título refrendado en la STA.
- ✓ Entregar exámenes médicos en Bienestar Estudiantil
- ✓ Realizar valoración nutricional en el PROTAL.
- ✓ Completar la ficha INSOEC en el sistema académico
- ✓ Instrucciones en el caso que el estudiante decida cambiarse de carrera.
- ✓ Consultar si aprobó el requisito de Herramientas de Colaboración Digital.

Además, temas como: gratuidad, técnicas para mejorar los hábitos de estudios, administración del tiempo, o algún otro tema que el consejero considere importante tratar con el estudiante.

## **2.1.2. Rendimiento académico**

### **2.1.2.1. ¿Qué es el rendimiento académico?**

El rendimiento académico es un indicador de gran importancia en el sistema educativo ya que permite medir el nivel de aprendizaje alcanzado por el estudiante en clase a lo largo de un período establecido.

En el año 1997, Garnica expuso que una nota es sólo una parte de la definición del rendimiento académico. Hablar sobre el rendimiento académico es un tema complejo, debido a que hay autores que indican que el rendimiento estudiantil está basado en una nota. Lo definen como el nivel demostrado de conocimientos en un área o materia, evidenciado a través de indicadores cuantitativos, expresados mediante una calificación bajo el supuesto que es un "grupo social calificado" el que fija los rangos de aprobación, para las materias (Tonconi, 2010).

Otros autores defienden que hay varias variables inherentes en un proceso de aprendizaje, las mismas que deben ser identificadas para que este indicador pueda ser manejado y controlado para su mejora. (Mendoza, Cadavid & Herrera, 2013).

### **2.1.2.2. Métodos para medir el rendimiento académico**

La UNESCO (1997) señala que en diferentes países los resultados del rendimiento académico de los estudiantes son medidos mediante calificaciones, porcentajes de logros, etc. Dado que expresan los resultados de diferente manera, hay que interpretarlo de manera distinta, lo que lo vuelve no comparable entre países.

Es de gran importancia que las universidades tengan un indicador con el cual medir el rendimiento académico del estudiante, y de esta manera evaluarlo y buscar la forma de mejorarlo.

A continuación, se muestran algunos indicadores utilizados para medir el rendimiento académico:

✓ **Número de materias aprobadas**

Este indicador permite determinar el éxito o fracaso de un estudiante en las materias registradas. Si el estudiante aprueba la mayor parte de las materias registradas en un periodo determinado se consideraría como éxito. En el caso que apruebe menos de la mitad, indicaría que el estudiante tuvo un bajo rendimiento, y se podría considerar como fracaso.

Di Gresia & Porto (2005) señalan que número de materias aprobadas es el mejor indicador para medir el rendimiento académico, al compararlo con el promedio de notas presenta una mayor variabilidad.

El porcentaje de materias aprobadas (% AP) por el estudiante en un periodo académico determinado calcularía como:

$$\% AP = \frac{\text{Número de materias aprobadas}}{\text{Número de materias registradas}} \times 100$$

✓ **Créditos acumulados**

El indicador de créditos acumulados representa el avance que ha tenido en la carrera. Compara el número de créditos acumulados por el estudiante durante cierto tiempo de estudio y el número de créditos que debió acumular de acuerdo al plan de estudio de la carrera. (Rodríguez & Ruiz, 2011).

El porcentaje de créditos acumulados se calcularía igual que el porcentaje de materias aprobadas, pero con créditos aprobados y total de créditos registrados.

Otro indicador que utiliza el número de créditos acumulados es el Índice de Eficiencia del Rendimiento Académico, el cual sirve para medir el avance de los estudios con relación a la permanencia del estudiante. Para su cálculo se

necesita el total de créditos requeridos según la carrera que se encuentra el estudiante, y se expresa de la siguiente manera:

$$IR = \left[ \left( \frac{\text{Número de créditos aprobados al semestre } t}{\text{Número de créditos requeridos por la carrera}} \right) * \left( \frac{\text{Número de semestres de duración de la carrera}}{\text{Número de semestres registrado}} \right) \right] * 100$$

### ✓ Notas o calificaciones

Las notas o calificaciones dado que es una variable que se puede interpretar cuantitativa o cualitativa, es fácil de analizar, mientras mayor sea la nota indica que tuvo un excelente rendimiento el estudiante y viceversa. El problema está en que no existe un criterio estandarizado para todos los centros educativos, para todas las materias o para todos los profesores (Page, 1990).

Las notas o calificaciones están condicionadas no sólo por la calidad del estudiante sino también por el profesor en el momento de enseñar y de calificar. Son un indicador de la enseñanza, más no estrictamente de su calidad (Escudero, 1999).

Un indicador que utiliza las calificaciones es el promedio aritmético de las calificaciones finales de todas las materias registradas ( $n$ ).

Otro indicador utilizado es el promedio aritmético ponderado acumulado (PAPA), el cual considera las calificaciones finales de todas las materias registradas ( $Calificación_i$ ), teniendo en cuenta el número de créditos correspondientes a cada materia registrada ( $Creditos_i$ ), y se calcula de la siguiente manera:

$$PAPA = \frac{\sum_{i=1}^n Calificación_i * Creditos_i}{\sum_{i=1}^n Creditos_i}$$

En la ESPOL, el rendimiento académico de una materia para un estudiante está basado en un proceso de evaluación constante donde se realizan actividades

como deberes, lecciones, trabajos y exámenes durante el término académico. Al finalizar el término académico se calcula una calificación final de la misma, si esta calificación es mayor o igual a 6 sobre 10 entonces aprueba la materia el estudiante, caso contrario no.

Y en el caso que se quiera medir el rendimiento académico del estudiante en un término académico específico, la ESPOL utiliza el indicador promedio aritmético de las calificaciones finales de las materias aprobadas en dicho término, sin importar el número de créditos que tengas cada una.

### **2.1.3. Estado del Arte**

A continuación, se mencionan algunas investigaciones realizadas en diferentes países en las que se utilizaron datos similares y aplicaron técnicas multivariadas:

Ramírez & Peláez (2001) realizaron un análisis de componentes principales para medir el rendimiento académico de los estudiantes de la carrera estadística informática de la ESPOL, período 1995-2000, donde se encontró que los estudiantes nivel 100 – I tienen un alto nivel de mortalidad en las materias: Cálculo I, Estructuras algebraicas I y Fundamentos de Computación.

Carvajal, Mosquera & Artamonova (2009) realizaron un estudio a los estudiantes que entraron a primer semestre, especialmente a ingenierías y tecnologías de la Universidad Tecnológica de Pereira (UTP) donde aplicaron un modelo de regresión logística para determinar qué factores influyen en el rendimiento académico en la materia Matemáticas I, donde obtuvieron que las variables influyentes eran: calificación del examen del ICFES (Instituto Colombiano para el Fomento de la Educación Superior), el nivel de lectura y riesgo de uso de sustancias psicoactivas y nivel del pensamiento lógico abstracto. Se descartaron variables como: edad, género, estrato socio-económico y tipo de colegio, además factores de riesgo de salud física y mental, cobertura de salud, entre otros.

Ibarra & Michalus (2010) mediante el análisis del rendimiento académico en un modelo logit de los estudiantes de la Facultad de Ingeniería de la Universidad Nacional de Misiones en Argentina; de las cohortes 1999 a 2003 obtuvieron que las variables significativas del rendimiento académico son: el promedio de calificaciones de la secundaria, el tipo de Institución de la secundaria y el número de asignaturas aprobadas en el primer año de carrera, siendo este el más relevante, destacando la importancia de la primera etapa de la carrera en los posteriores resultados académicos del estudiante.

En un estudio realizado por Rodríguez & Ruiz (2011) a los estudiantes de la carrera de Química a una universidad de Uruguay, mediante un análisis de regresión múltiple encontraron que las variables que tienen efectos significativos en el promedio de calificaciones fueron el sexo, el colegio, si trabaja, calificaciones de lucimiento, capacidad percibida y rendimiento previo. Realizaron también otro modelo con las mismas variables explicativas donde se obtuvo que el porcentaje de avance en la carrera está afectado significativamente por si trabaja, calificaciones de lucimiento y rendimiento previo. Sin embargo, indican que el promedio de calificaciones sería un mejor indicador del rendimiento, dado que es más amplio que el avance en la carrera.

Cardona & Hernández (2013) utilizaron un modelo de regresión logística para predecir el IAR (Índice de Rendimiento Académico), donde obtuvieron factores con influencia a las calificaciones en el examen ICFES (Instituto Colombiano para el Fomento de la Educación Superior). Otro de los factores que aparece con una influencia positiva es “No provienen de la Capital”, sexo y edad.

McArdle, Paskus & Boker (2013), aplicaron modelos de regresión multinivel estándar con predicción no lineal aplicado a estudiantes de primer año de universidad para predecir sus calificaciones, utilizaron variables demográficas, del colegio, variables académicas de la secundaria, donde concluyeron que las variables más influyentes fueron las de la secundaria.

Meyer & Marx (2014), en su investigación encontró que la mayoría de los estudiantes desertan de las carreras de ingeniería debido a un déficit en una de

las cuatro áreas siguientes: académica, desencanto con la profesión, profesores de ingeniería, plan de estudios, y la falta de preparación en la secundaria.

Planck Barahona (2014) realizó un modelo de regresión múltiple para determinar los factores que influyen en el rendimiento académico de los estudiantes de primer año de las carreras de Trabajo Social, Ingeniería, Derecho y Humanidades de la Universidad de Atacama, Chile. Las variables predictoras estadísticamente significativas fueron: género, si trabaja, su conformidad con la carrera, y su calificación de matemática.

Estudios realizados en instituciones españolas señalan que generalmente el mayor porcentaje de deserción estudiantil se presenta en los estudiantes de primer año, dado que presentan grandes dificultades en su proceso de adaptación a la vida universitaria, y suele estar relacionado con la edad, exploración de la identidad, nuevas experiencias, limitaciones cognitivas, malos hábitos y disposición hacia el aprendizaje, entre otras (Martínez, Fernández, Couñago, Vacas, da Silva & González, 2014).

Tinisaray & Karina (2016) realizaron un análisis logístico bivalente del rendimiento académico de los estudiantes de una de las universidades ecuatorianas con más número de estudiantes a nivel de educación superior a distancia en Latinoamérica. Donde obtuvieron que ni el género ni la religión son estadísticamente significativos, mientras que la edad sí. Obtuvieron que los estudiantes jóvenes tienen menos ventaja de conseguir un rendimiento académico óptimo en relación a los estudiantes en edad adulta. Además, que los estudiantes que participan de manera mediana o activamente en actividades en línea como: chat, foro y video-colaboración tienen mayor ventaja de conseguir un rendimiento académico óptimo.

## 2.2. Marco Metodológico

### 2.2.1. Datos

Se utilizará la información de los estudiantes de primer nivel de las carreras de Ingeniería del término 2016- IIT al 2018 – IT de las bases de datos:

- ✓ Datos personales estudiante del Académico ESPOL
- ✓ Calificaciones de las materias del Proceso de Admisiones
- ✓ Materias registradas del Académico ESPOL.

### 2.2.2. Variables

En la *Tabla 1*, se describen las variables que se van a utilizar en esta investigación.

*Tabla 1. Descripción de las variables de análisis*

Grupo de Variable	Nombre de la Variable	Categorías
Características Sociodemográficas	Género del estudiante	0 Femenino 1 GLBTI 2 Masculino
	Edad del estudiante	-
	Estado Civil del estudiante	0 Casado 1 Soltero
	Autoidentificación étnica de estudiante	0 Blanco 1 Indígena 2 Mestizo 3 Montubio 4 Negro 5 Otro
	Estado Civil familiar	0 Casado/a 1 Divorciado/a 2 Soltero/a 3 Unión Libre 4 Viudo/a
	Grupo que reside	0 Amigos 1 Grupo Familiar 2 Parientes 3 Solo
	Parentesco con el jefe de hogar	0 Padre 1 Madre 2 Otro
	Nivel de Instrucción del jefe de hogar	0 Ninguno 1 Primaria 2 Secundaria 3 Tercer nivel 4 Postgrado
	Total de Ingresos del Hogar	-

	Estrato Socioeconómico	0 Bajo 1 Medio Bajo 2 Medio 3 Medio Alto 4 Alto
	Número de celulares en el hogar	0 Ninguno 1 Tiene 1 2 Tiene 2 3 Tiene 3 4 Tiene 4 o más
	Número de televisores en el hogar	0 Ninguno 1 Tiene 1 2 Tiene 2 3 Tiene 3 o más
	Número de vehículos en el hogar	0 Ninguno 1 Tiene 1 2 Tiene 2 3 Tiene 3 o más
	Tipo de vivienda	0 Cuarto 1 Departamento 2 Vivienda
	Forma de pago de la vivienda	0 Alquilado 1 Cedido 2 Hipotecado 3 Propio
	Tiene celular	0 No 1 Si
	Tiene teléfono convencional	0 No 1 Si
	Tiene servicio de Internet	0 No 1 Si
	Tienen servicio de tv. Pagada	0 No 1 Si
	Tiene Pc escritorio	0 No 1 Si
	Tiene laptop	0 No 1 Si
	Colegio ubicado en la Ciudad de Guayaquil	0 No 1 Si
	Tipo de colegio	0 Fiscal 1 Fisco misional 2 Particular
	Tiene familiares afiliados al IESS	0 No 1 Si
	Tiene familiares afiliados a un seguro privado	0 No 1 Si
Proceso de Admisiones	Física en el proceso de Admisiones	-
	Matemáticas en el proceso de Admisiones	-
	Química en el proceso de Admisiones	-
Materias registradas	Aprueba todas las materias	0 No 1 Si
	Número de materias aprobadas	-
	Número de materias tomadas	-
	Tasa de materias aprobadas	-
	Número de materias STEM	-

Número de materias Sociales	-
Índice del grado de dificultad de las materias STEM	-
Índice del grado de dificultad de las materias Sociales	-

*Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I*  
*Elaborado por: J. Barzola*

### 2.2.3. Regresión lineal múltiple

La regresión lineal múltiple es utilizada para pronosticar/predecir el valor de una variable dependiente, en este caso el rendimiento académico, en función de posibles variables de explicación o factores que influyan significativamente en la misma (Zeltermán, 2015).

Se denomina como variable de respuesta a  $y$ , la misma que será explicada por  $p$  variables  $x_1, x_2, \dots, x_p$ , en un conjunto de  $n$  observaciones. El modelo de regresión lineal múltiple está dado por:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i$$

Los  $\beta_0, \beta_1, \beta_2, \dots, \beta_p$  son los  $p + 1$  parámetros del modelo que miden la influencia que las variables explicativas tienen sobre la variable dependiente. Se entiende como  $\beta_0$  es la intersección o término "constante", los demás  $\beta_i$  son los parámetros respectivos a cada variable independiente.

El término  $\varepsilon_i$  estimaciones de la perturbación aleatoria o errores, que asume aquellos factores que no son controlables u observables, se asocian con el azar, y es confiere al modelo su carácter estocástico.

#### 2.1.1.1. Supuestos del modelo de regresión lineal

En un modelo de regresión lineal, se deben considerar los siguientes supuestos:

- La relación entre las variables debe ser lineal.
- Los errores deben ser independientes entre sí.
- Los errores tengan varianza constante. (Homocedasticidad)

- El valor esperado de los errores es igual a cero.
- El error total es la suma de todos los errores.

#### **2.2.4. Árboles de clasificación CHAID**

CHAID (Chi-Squared Automatic Interaction Detector) es un paradigma del aprendizaje automático que se enmarca dentro de los algoritmos de aprendizaje supervisado y que sirve para generar modelos de clasificación predictivos. Además, consiste en un algoritmo para la construcción de árboles de decisión mediante estadísticos de chi-cuadrado para identificar divisiones óptimas.

Esta técnica estadística es utilizada para analizar una gran cantidad de datos y crear perfiles de usuarios, ayuda a identificar patrones de conducta, comportamiento de los consumidores, reglas de decisión de incluir elementos a un determinado grupo. Es decir, sirve para segmentar a una población en subconjuntos homogéneos entre los elementos que los componen, ya que cumplen unas determinadas características, pero heterogéneos entre sí.

Está limitado por una variable dependiente, puede ser tipo nominal, ordinal, categórica, discreta o discontinua. No se hace suposiciones de normalidad.

Se sugiere trabajar con al menos 1000 casos y con un gran número de variables predictoras. Las variables predictoras con varias categorías tienen una probabilidad reducida de ser seleccionadas.

Ahorra de tiempo, evita que se analicen cientos de tabulaciones cruzadas e identifica de manera rápida y fácil las relaciones significativas entre las variables. Tiene más ventajas que las técnicas tradicionales de Análisis Multivariado como: regresión múltiple, regresión log-Lineal, análisis discriminante, y análisis de cluster.

##### **2.2.4.1. Procedimiento**

CHAID divide a la población en dos o más grupos distintos dependiendo del número de categorías del mejor predictor de una variable dependiente seleccionada. Para esto se utiliza el valor  $p$  de la distribución Chi-Cuadrado. Se forman todos los pares posibles de categorías. Para cada posible par se calcula el correspondiente cruce con la variable dependiente.

El par con el valor más bajo Chi-Cuadrado, y siempre que no sea significativo, formará una nueva categoría de dos valores fusionados. Si se ha fusionado un determinado par de categorías, se procede a realizar nuevas fusiones de los valores del pronosticador, pero esta vez con una categoría menos, pues dos de las antiguas han sido reducidas a una sola.

Luego divide cada uno de estos en grupos más pequeños basados en las demás variables predictoras. No se continuará dividiendo a la población, en el caso que el valor  $p$  ajustado de la variable independiente que mejor explique a la variable dependiente no es menor que dicho criterio. Este proceso de división continúa y termina hasta que no se encuentren más predictores estadísticamente significativos o con un tamaño mínimo del grupo predefinido.

CHAID despliega los subgrupos finales en un diagrama de árbol fácil de entender. En cada rama aparece un nodo indicando el número de elementos del segmento, la leyenda de la variable explicativa y algún estadístico o test de significación.

## CAPÍTULO 3

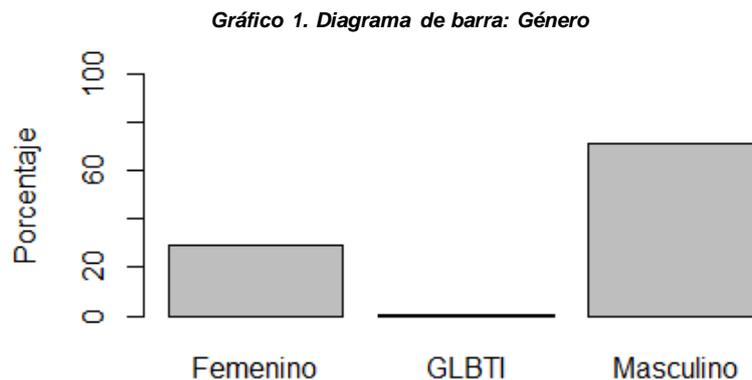
### RESULTADOS

#### 3.1. Análisis descriptivo

##### 3.1.1. Características Sociodemográficas

###### ➤ Género

En el *Gráfico 1*, se puede observar que aproximadamente el 71% de los estudiantes son hombres, alrededor del 0% son declarados GLBTI y el porcentaje restante son mujeres.

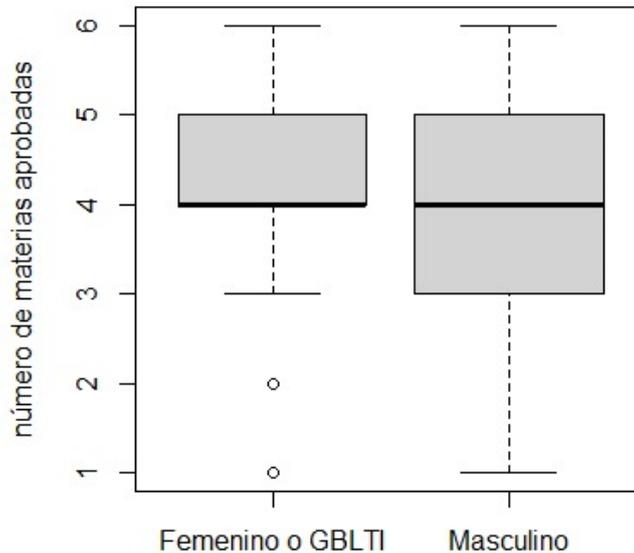


*Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola*

En el *Gráfico 2*, se puede observar que se agruparon en una misma categoría Femenino y GLBTI. En ambas categorías la mediana es igual a 4 materias aprobadas. Generalmente los estudiantes de género femenino o GLBTI, aprueban entre 3 y 6 materias. El 50% de ellos aprueban entre 4 a 5 materias, hay valores atípicos de estudiantes que aprueban una o dos materias. Mientras

que generalmente los estudiantes de género masculino, aprueban entre 1 y 6 materias. El 50% de ellos aprueban entre 3 a 5 materias.

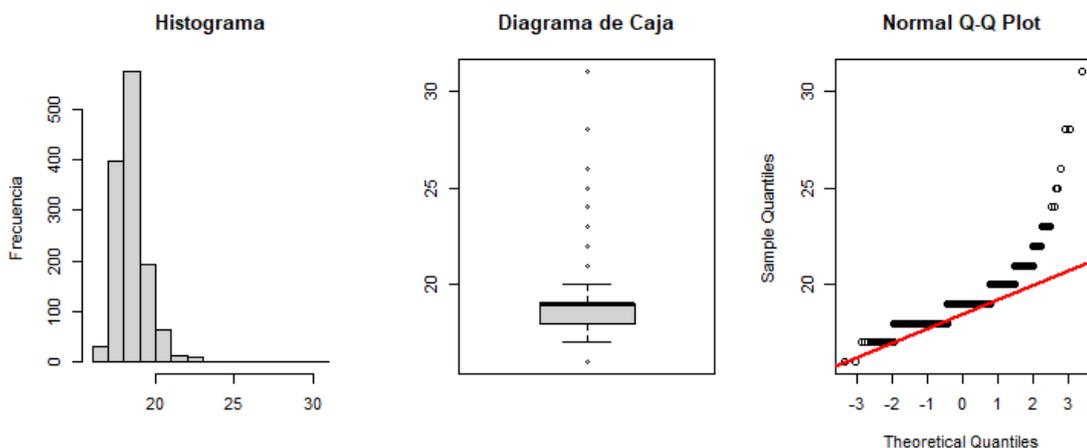
**Gráfico 2. Boxplot: Materias aprobadas vs. Género**



**Fuente:** Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
**Elaborado por:** J. Barzola

➤ **Edad**

**Gráfico 3. Histograma, Boxplot, Normal Q-Q Plot: Edad**



**Fuente:** Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
**Elaborado por:** J. Barzola

En el *Gráfico 3*, se puede observar que la edad no sigue una distribución normal, está sesgada hacia la derecha. Alrededor del 80% tienen menos de 20 años. El 50% de las observaciones tiene edad entre 18 y 19 años. En la *Tabla 4*, se puede

observar que en promedio los estudiantes tienen 19 años. Tiene una desviación estándar de 1.16. La edad mínima registrada fue 16 años y la máxima 31 años.

**Tabla 2. Estadística descriptiva: Edad**

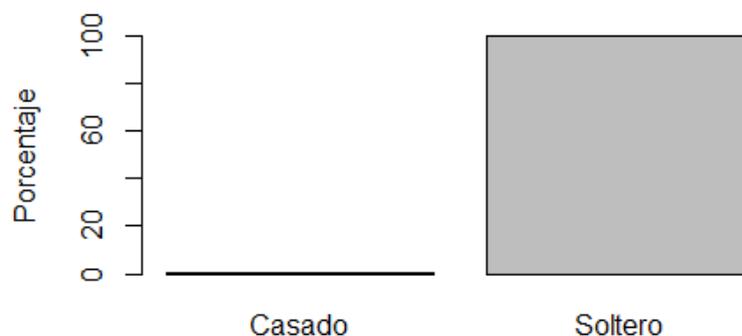
Statistics	edad	
Mean	19,00	
Std. Error of Mean	0,03	
Median	19,00	
Mode	19,00	
Std. Deviation	1,16	
Variance	1,34	
Skewness	2,62	
Kurtosis	16,99	
Minimum	16,00	
Maximum	31,00	
Percentiles	25	18,00
	50	19,00
	75	19,00

*Fuente:* Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
*Elaborado por:* J. Barzola

### ➤ Estado Civil

En el *Gráfico 4*, se puede observar que aproximadamente el 100% de los estudiantes son solteros.

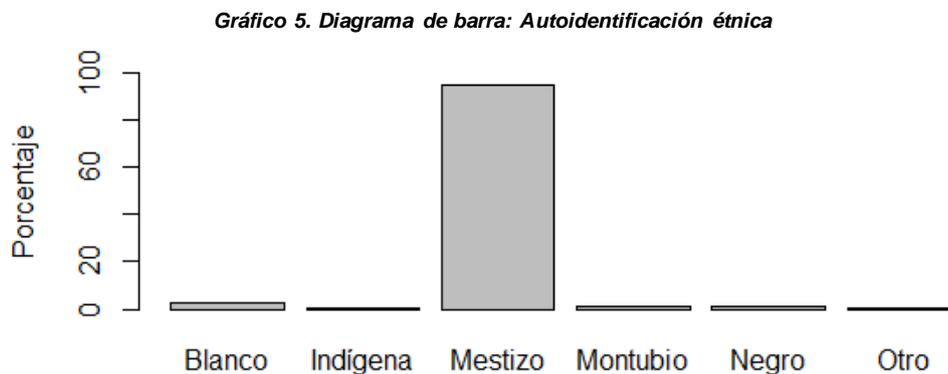
**Gráfico 4. Diagrama de barra: Estado Civil**



*Fuente:* Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
*Elaborado por:* J. Barzola

### ➤ Autoidentificación étnica

En el *Gráfico 5*, se puede observar que aproximadamente el 95% de los estudiantes se autoidentifican como mestizos, el 5% restante se autoidentifican como blancos, indígenas, montubios, negros u otros.

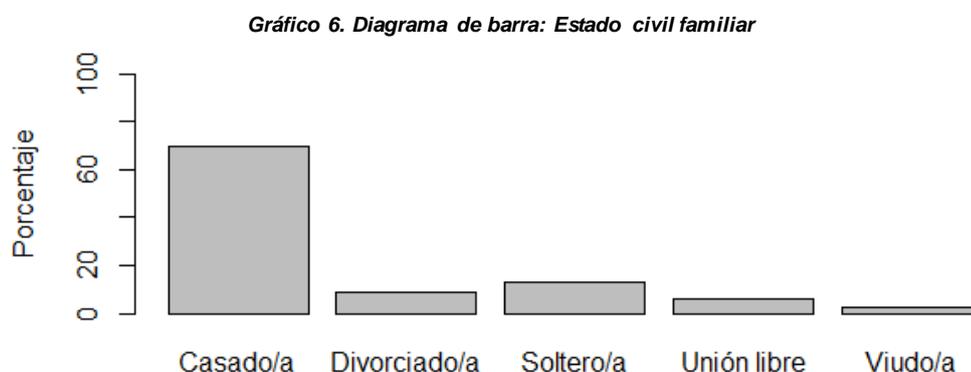


Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

### 3.1.2. Grupo familiar

#### ➤ Estado civil familiar

En el *Gráfico 6*, se puede observar que aproximadamente el 76% de los padres de los estudiantes tienen estado civil casados o unión libre. El 24% restante son divorciados, viudos o solteros.

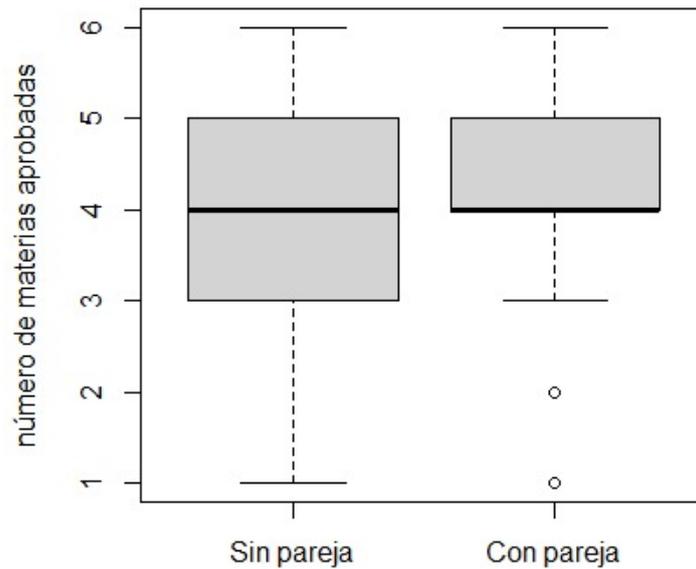


Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

En el *Gráfico 7*, se puede observar que se agruparon en la categoría Sin Pareja las opciones Soltero, Divorciado o Viudo. También se agruparon en la categoría

Con pareja las opciones Casado o Unión Libre. En ambas categorías la mediana es igual a 4 materias aprobadas. Generalmente los estudiantes que el estado civil de sus padres es Con pareja, aprueban entre 3 y 6 materias. El 50% de ellos aprueban entre 4 a 5 materias, hay valores atípicos de estudiantes que aprueban una o dos materias. Mientras que los que tienen estado civil Sin pareja, aprueban entre 1 y 6 materias. El 50% de ellos aprueban entre 3 a 5 materias.

Gráfico 7. Boxplot: Materias aprobadas vs. Estado Civil familiar

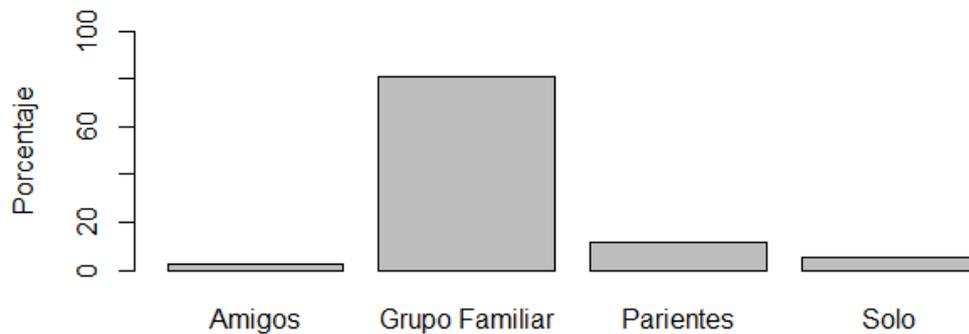


Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

### ➤ Grupo que reside

En el Gráfico 8, se puede observar que aproximadamente el 80% de los estudiantes en su lugar de residencia viven con sus familias, el 11% con parientes o familiares y el restante con sus amigos o solos.

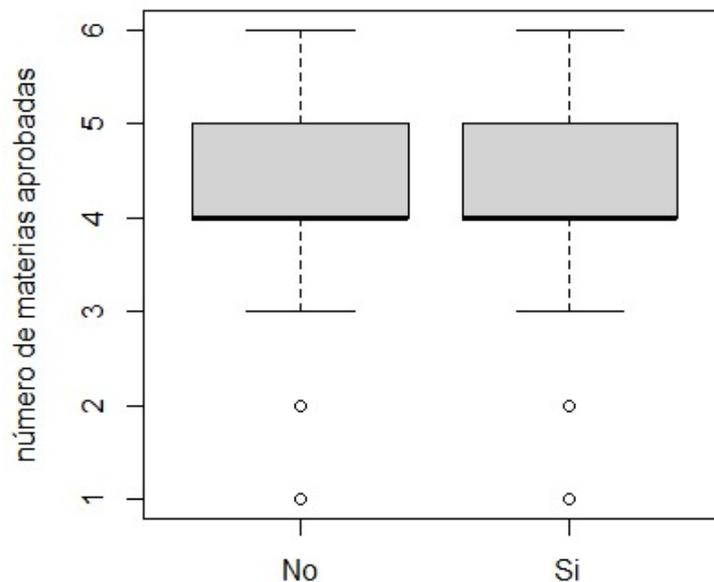
Gráfico 8. Diagrama de barra: Grupo que reside



Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

En el Gráfico 9, se puede observar que se recodificó la variable, a una nueva denominada Reside con la familia con sólo dos categorías Si o No. Presentan una distribución similar ambas categorías.

Gráfico 9. Boxplot: Materias aprobadas vs. Reside con la familia



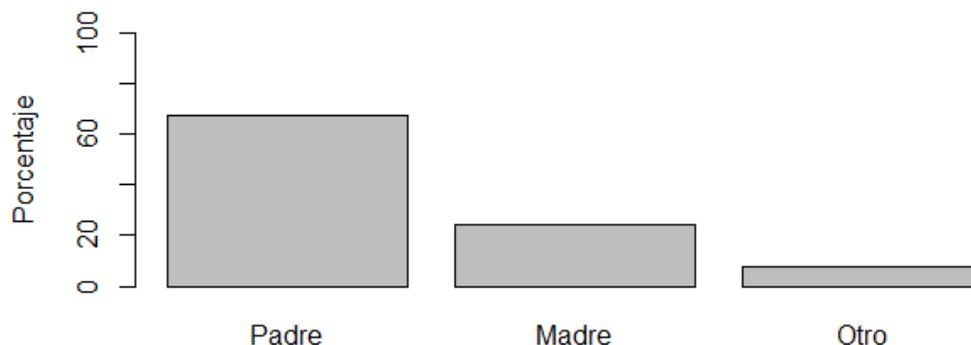
Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

### 3.1.3. Jefe del Hogar

#### ➤ Parentesco con el jefe del hogar

En el Gráfico 10, se puede observar que el mayor porcentaje corresponde al padre como jefe de hogar con el 67%. Mientras que en el 24% es la madre.

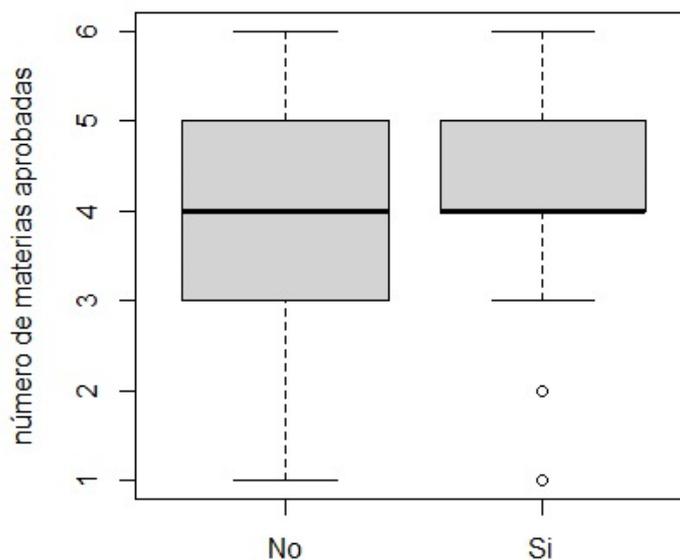
**Gráfico 10. Diagrama de barra: Parentesco con el jefe del hogar**



Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

En el Gráfico 11, se puede observar que se recodificó la variable, a una nueva denominada “El padre es el jefe de hogar” con sólo dos categorías Si o No. Los estudiantes que tienen como jefe del hogar al padre aprueban entre 3 y 6 materias, el 50% de las observaciones se concentran entre 4 y 5. Mientras que los que no tienen como jefe del hogar al padre aprueban entre 1 y 6 materias, el 50% de las observaciones se concentran entre 3 y 5.

**Gráfico 11. Boxplot: Materias aprobadas vs. El padre es el jefe del hogar**

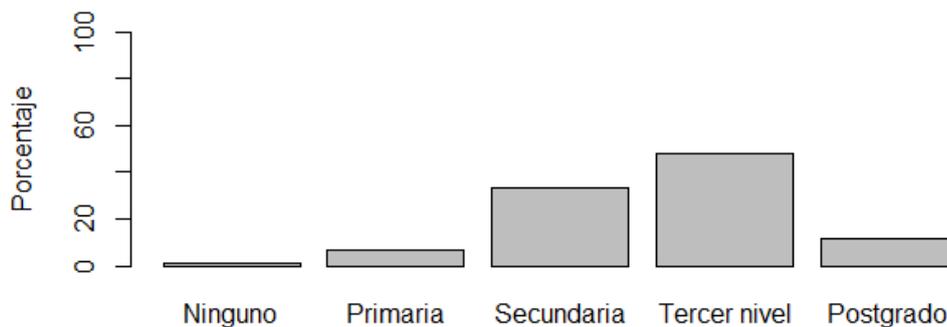


Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

➤ **Nivel de instrucción del jefe de hogar**

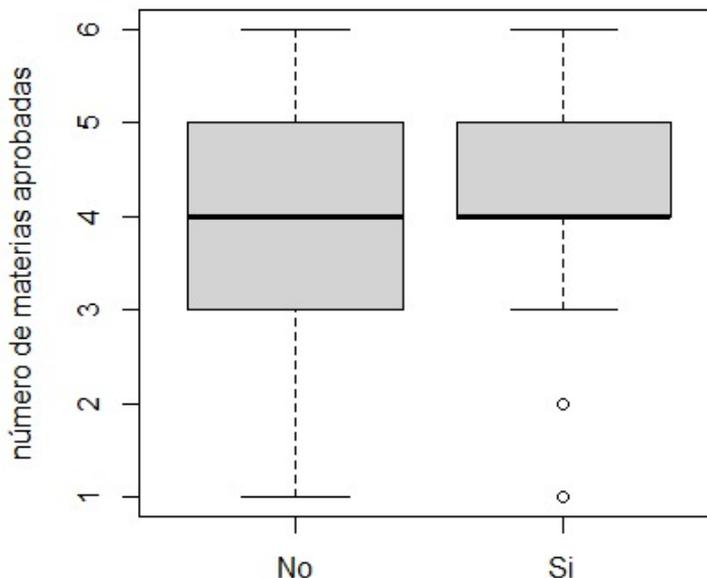
En el *Gráfico 12*, se puede observar que el nivel de instrucción del jefe del hogar en su mayoría corresponde al tercer nivel con un 48% de la muestra estudiada, seguido por el nivel secundaria con el 33%. Cabe recalcar que el nivel de postgrados en los jefes de hogar es del 11%.

*Gráfico 12. Diagrama de barra: Nivel de instrucción del jefe del hogar*



Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

*Gráfico 13. Boxplot: Mat. aprobadas vs. Jefe del hogar tiene tercer nivel de instrucción*



Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

En el *Gráfico 13*, se puede observar que se recodificó la variable, a una nueva denominada “Jefe del hogar tiene tercer nivel de instrucción” con sólo dos categorías Si o No. Se agruparon las categorías Tercer Nivel o Postgrado en Si, y las demás categorías en No. Los estudiantes que tienen como al jefe del con

nivel de instrucción tercer nivel aprueban entre 3 y 6 materias, el 50% de las observaciones se concentran entre 4 y 5. Mientras que los no aprueban entre 1 y 6 materias, el 50% de las observaciones se concentran entre 3 y 5.

### 3.1.4. Nivel Socioeconómico

#### ➤ Ingresos en el hogar

En la *Tabla 3*, se puede observar que el ingreso promedio en los hogares de los estudiantes fue de 1387.25 dólares. Tiene una desviación estándar de 1,176.81. El ingreso mínimo registrado fue de 200.00 dólares y el máximo fue de 12,500.00 dólares.

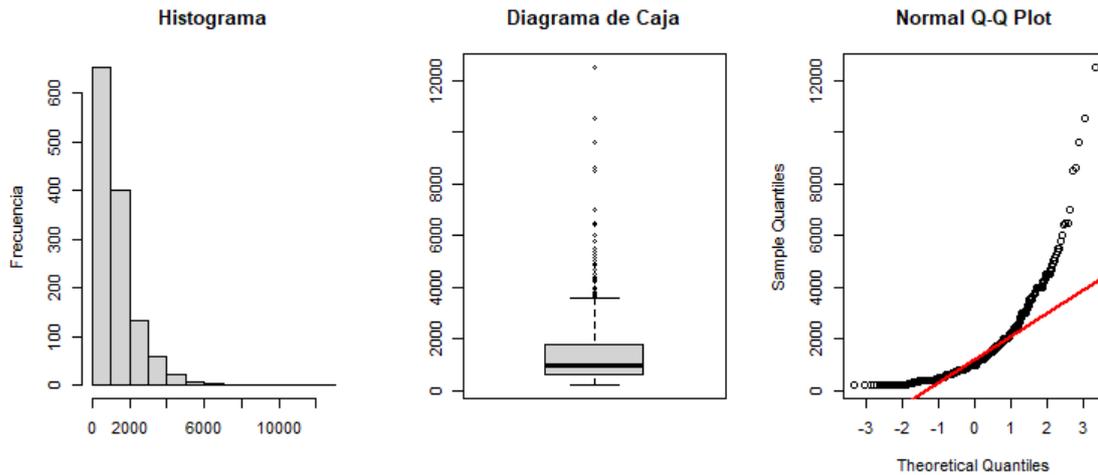
**Tabla 3. Estadística descriptiva: Ingresos en el hogar**

Statistics	Ingresos
Mean	1387,25
Std. Error of Mean	32,79
Median	1000,00
Mode	1000,00
Std. Deviation	1176,81
Variance	1384874,33
Skewness	2,80
Kurtosis	14,09
Minimum	200,00
Maximum	12500,00
	25
Percentiles	600,00
	50
	1000,00
	75
	1800,00

*Fuente:* Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
*Elaborado por:* J. Barzola

En el *Gráfico 14*, se puede observar que los ingresos no siguen una distribución normal, está sesgada hacia la derecha. El 50% de las observaciones tienen ingresos entre 600 y 1800 dólares.

**Gráfico 14. Histograma, Boxplot, Normal Q-Q Plot: Ingresos en el hogar**

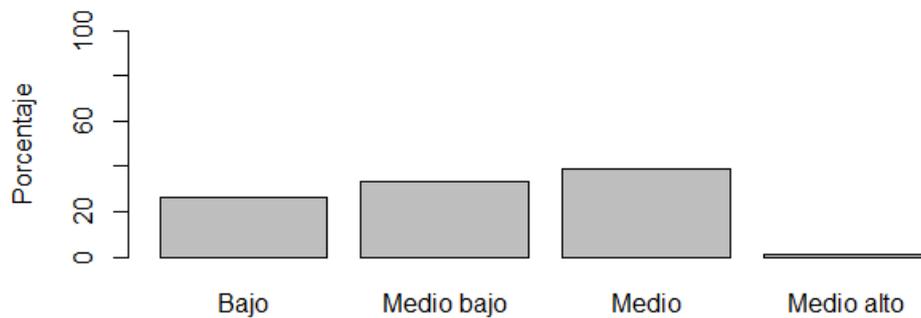


*Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola*

➤ **Estrato Socioeconómico**

En el *Gráfico 15*, se puede observar que el mayor porcentaje corresponde al estrato socioeconómico medio con un 38%. Mientras que el menor porcentaje pertenece a la clase media alta con un 1.24%.

**Gráfico 15. Diagrama de barra: Estrato Socioeconómico**

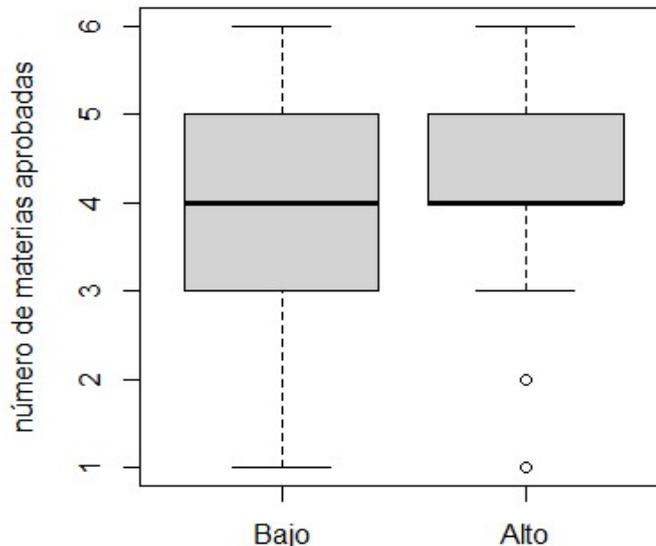


*Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola*

En el *Gráfico 16*, se puede observar que se recodificó las categorías de la variable Estrato Socioeconómico. Se agruparon las categorías Bajo o Medio Bajo como Bajo, y las demás categorías Medio o Medio Alto como Alto. Los estudiantes con estrato socioeconómico Alto aprueban entre 3 y 6 materias, el 50% de las observaciones se concentran entre 4 y 5. Mientras que los que tienen

estrato Bajo aprueban entre 1 y 6 materias, el 50% de las observaciones se concentran entre 3 y 5.

Gráfico 16. *Boxplot: Materias aprobadas vs. Estrato Socioeconómico*

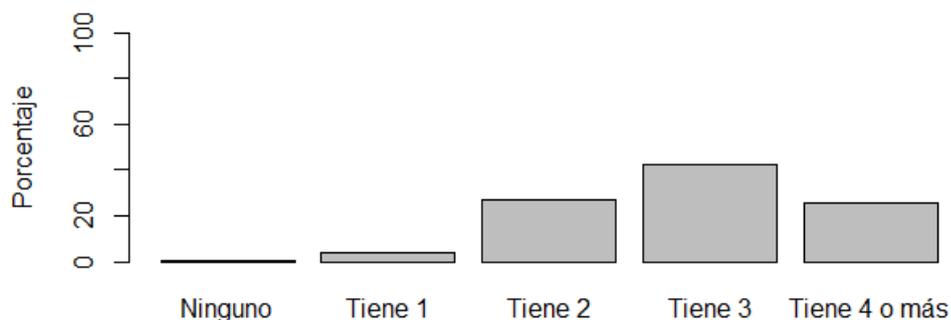


Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
 Elaborado por: J. Barzola

### ➤ Número de celulares en el hogar

En el Gráfico 17, se puede observar que la mayoría posee al menos 3 celulares en el hogar, corresponde aproximadamente al 68%.

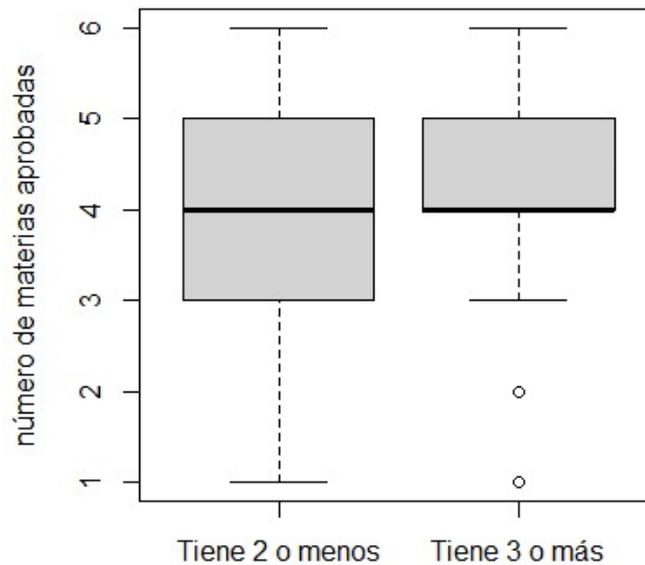
Gráfico 17. *Diagrama de barra: Número de Celulares en el Hogar*



Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
 Elaborado por: J. Barzola

En el Gráfico 18, se puede observar que aquellos que tienen 3 o más celulares Alto aprueban entre 3 y 6 materias, el 50% de las observaciones se concentran entre 4 y 5. Mientras que los que tienen 2 o menos celulares aprueban entre 1 y 6 materias, el 50% de las observaciones se concentran entre 3 y 5.

Gráfico 18. Boxplot: Materias aprobadas vs. Número de Celulares en el Hogar

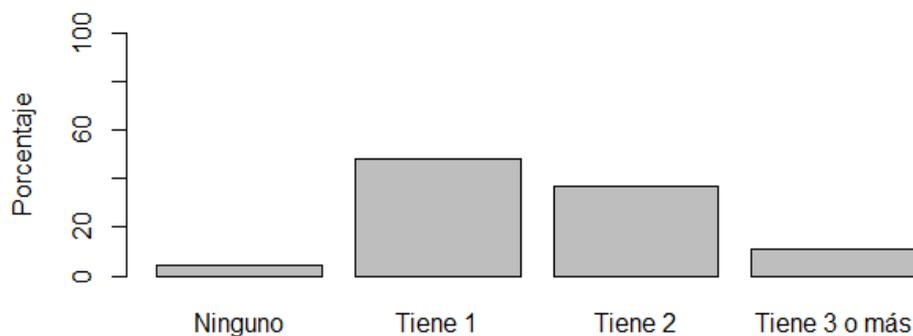


Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
 Elaborado por: J. Barzola

### ➤ Número de televisores en el hogar

En el Gráfico 19, se puede observar que la mayoría de los estudiantes poseen entre 1 o 2 televisores en el hogar, correspondiente al 85%. El 4% no tienen televisores en el hogar.

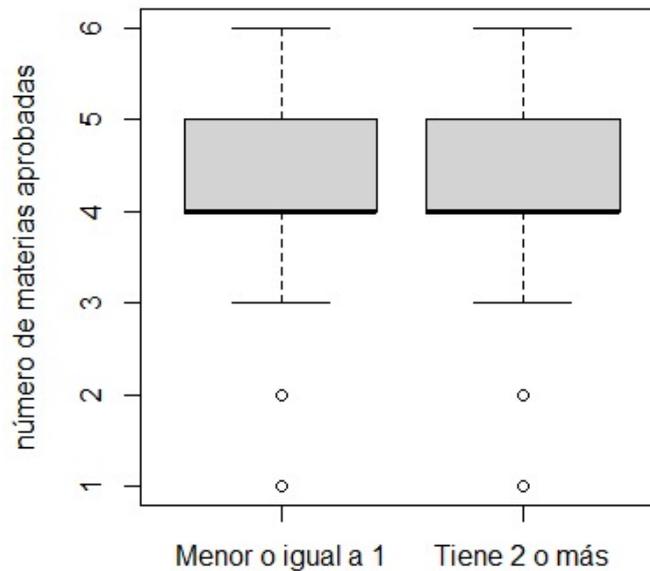
Gráfico 19. Diagrama de barra: Número de televisores en el hogar



Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
 Elaborado por: J. Barzola

En el Gráfico 20, se puede observar que aquellos que tienen 2 o más televisores tienen una distribución similar a los que tienen 1 o menos televisores con respecto al número de materias aprobadas.

Gráfico 20. Boxplot: Materias aprobadas vs. Número de televisores en el hogar

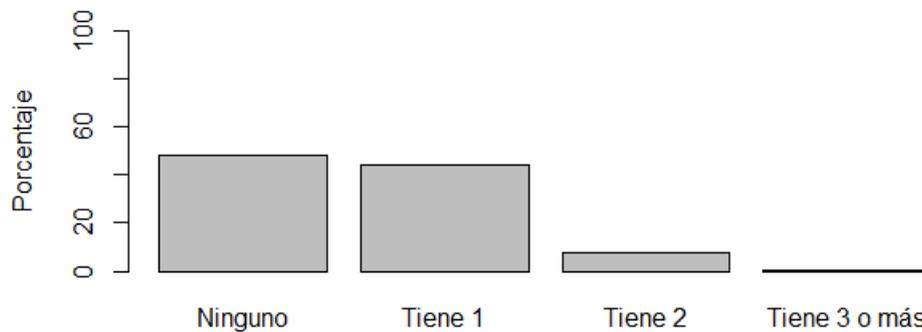


Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

### ➤ Número de vehículos en el Hogar

En el Gráfico 21, se puede observar que el 48% no tienen vehículos en sus hogares, aproximadamente el 44% tienen uno.

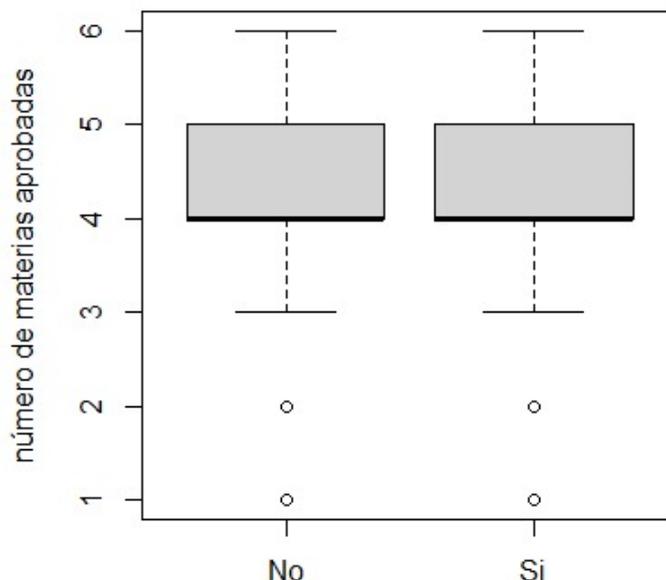
Gráfico 21. Diagrama de barra: Número de vehículos en el hogar



Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

En el Gráfico 22, se puede observar que aquellos que tienen vehículo tienen una distribución similar a los que no con respecto al número de materias aprobadas.

Gráfico 22. Boxplot: Materias aprobadas vs. Tiene vehículo en el hogar



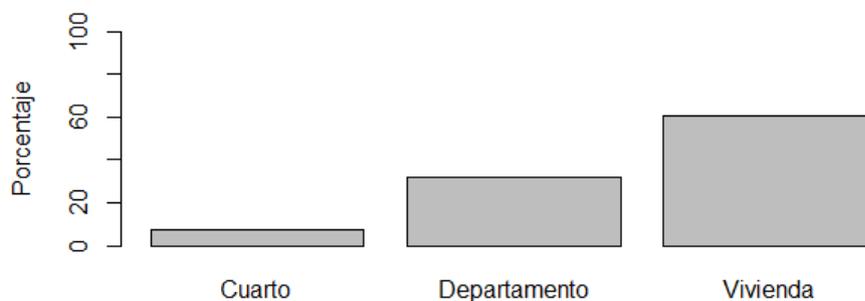
Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

### 3.1.5. Vivienda

#### ➤ Tipo de vivienda

En el Gráfico 23, se puede observar que la mayor parte de los estudiantes viven en un vivienda o casa, aproximadamente el 61%, seguido por el 32% que viven en departamentos y el restante viven en cuartos.

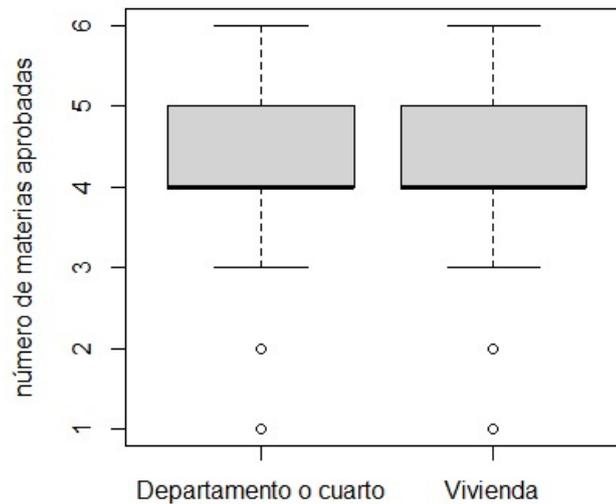
Gráfico 23. Diagrama de barra: Tipo de vivienda



Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

En el Gráfico 24, se puede observar que aquellos que habitan en una vivienda tienen una distribución similar a los que habitan en un departamento o un cuarto con respecto al número de materias aprobadas.

Gráfico 24. Boxplot: Materias aprobadas vs. Tipo de vivienda

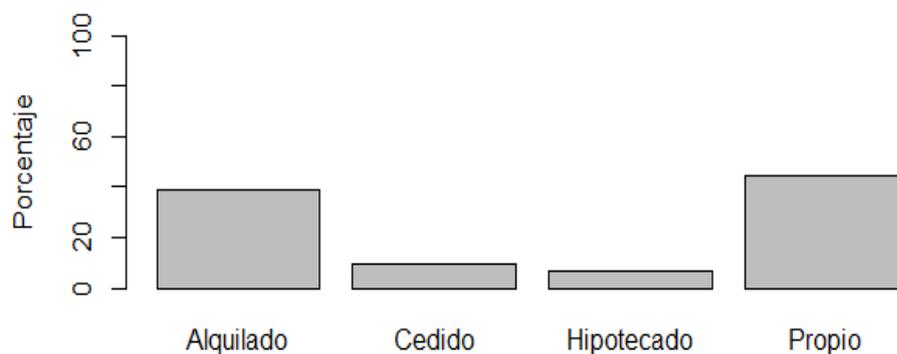


Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
 Elaborado por: J. Barzola

### ➤ Forma de pago de la vivienda

En el Gráfico 25, se puede observar que la mayoría de personas; el 44% poseen vivienda propia, seguida por el 39% perteneciente a personas que alquilan vivienda.

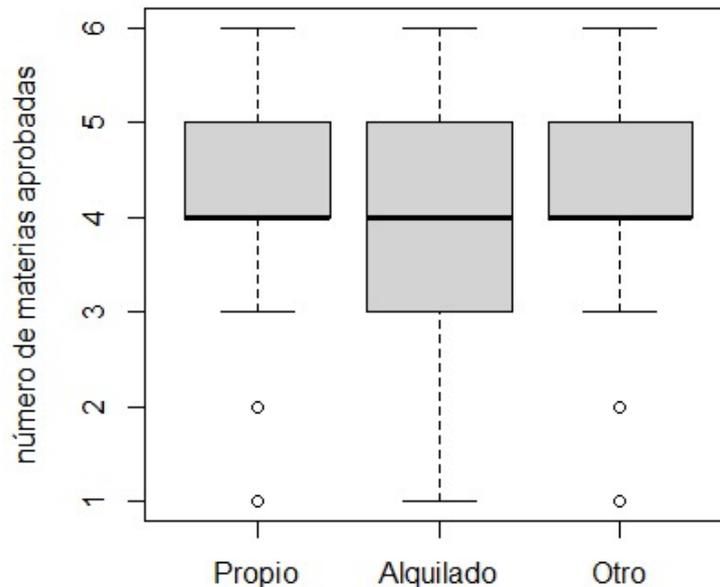
Gráfico 25. Diagrama de barra: Forma de pago de la vivienda



Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
 Elaborado por: J. Barzola

En el Gráfico 26, se puede observar que los estudiantes que viven en casa propia u otro, aprueban entre 3 y 6, el 50% de las observaciones aprueban entre 4 y 5 materias. Mientras que los estudiantes que alquilan su rango de materias aprobadas va desde 1 a 6 materias, el 50% de las observaciones se concentra entre 3 y 5 materias aprobadas.

Gráfico 26. Boxplot: Materias aprobadas vs. Forma de pago de la vivienda



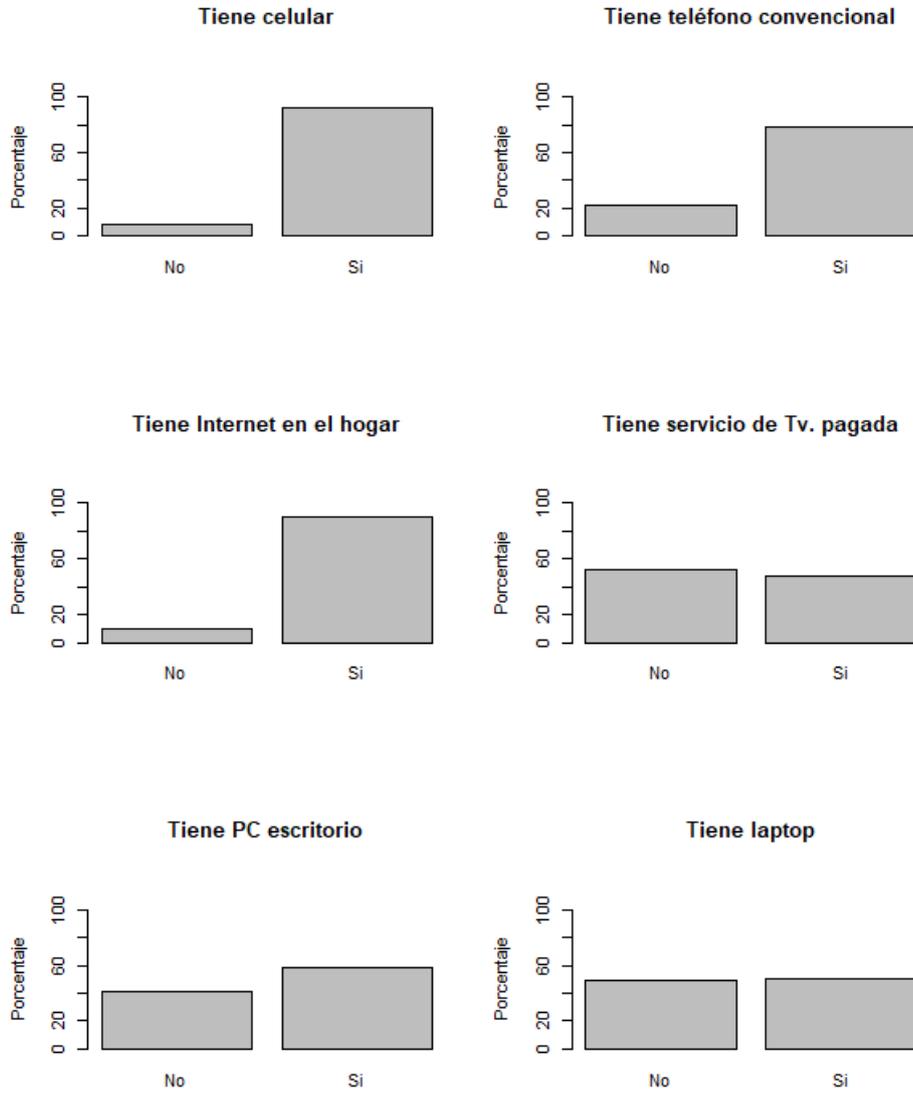
Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

### 3.1.5.1. Tecnologías de Información y Comunicación (TICs)

En el Gráfico 27, se puede observar que el 92% tiene celular, el 78% manifestó tener teléfono convencional. En cuanto al internet, el 89% posee este servicio en su hogar. A demás el 58% posee pc de escritorio, el 50% tiene laptop.

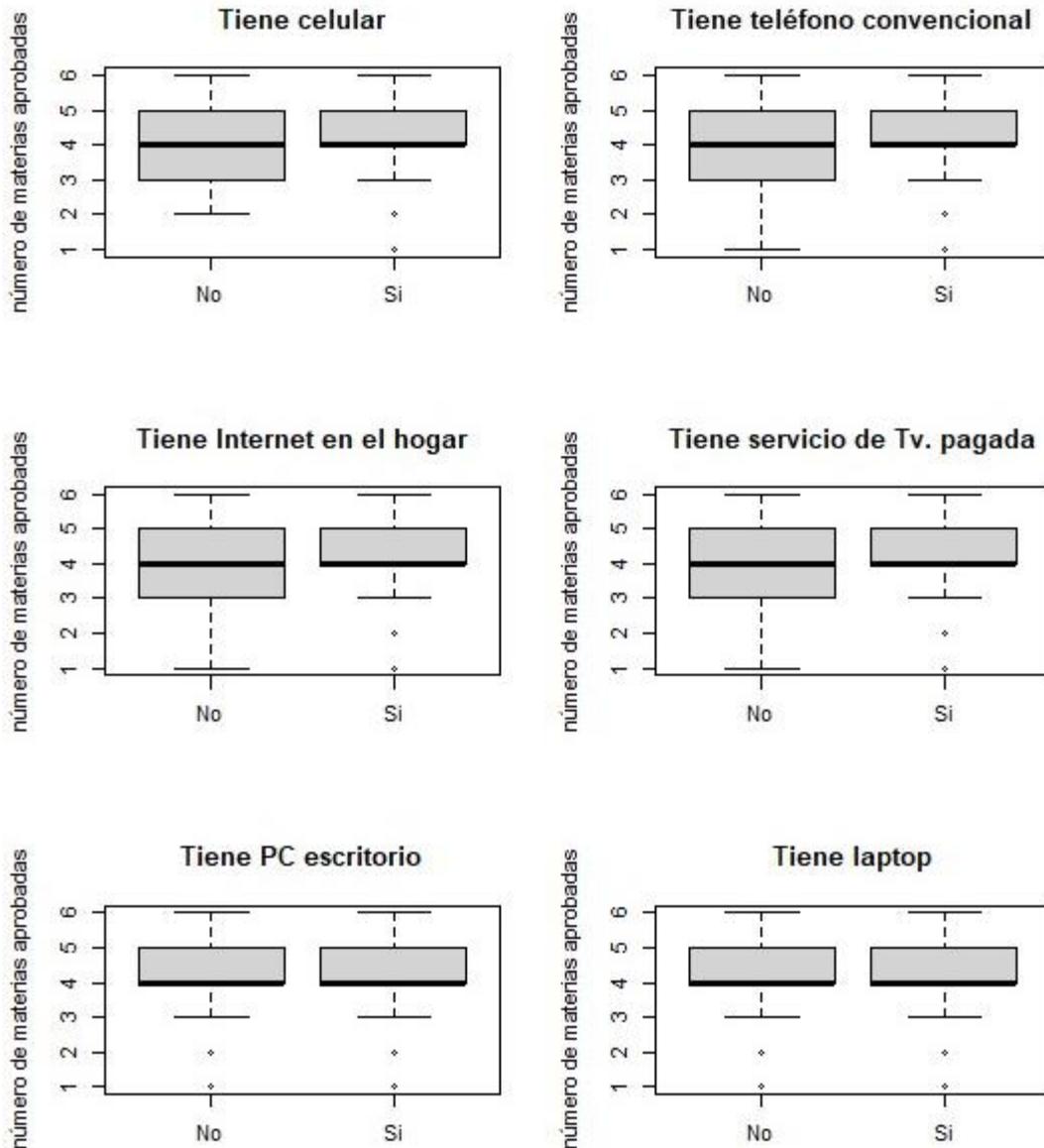
En el Gráfico 28, se puede observar que el número de materias aprobadas se distribuyen de manera similar las variables Tiene celular, tiene teléfono convencional, tiene internet en el hogar y tiene servicio de tv. Pagada. Mientras que las variables tienen pc escritorio y tiene laptop tienen una distribución similar.

Gráfico 27. Diagrama de barra: Tecnologías de Información y Comunicación



Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

Gráfico 28. Boxplot: Materias aprobadas vs. Tecnologías de Información y Comunicación



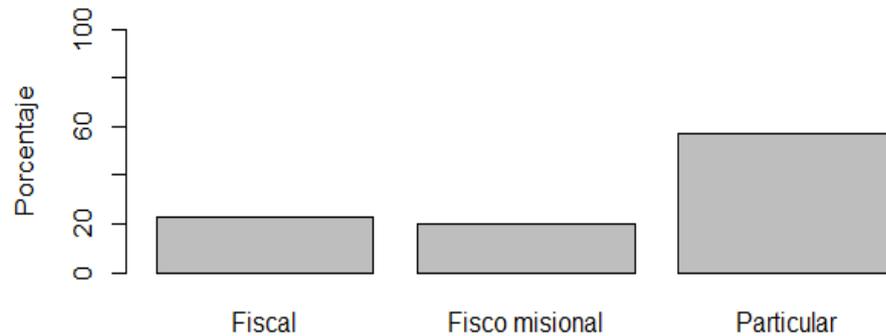
Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

### 3.1.5.2. Educación secundaria

#### ➤ Tipo de Colegio

En el Gráfico 29, se puede observar que la mayoría estudió en un colegio particular, aproximadamente el 56%.

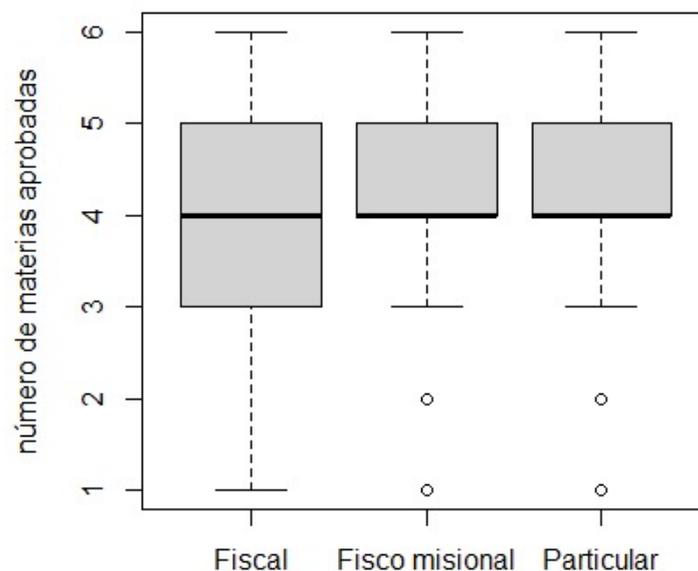
Gráfico 29. Diagrama de barra: Tipo de Colegio



Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

En el Gráfico 30, se puede observar que aquellos que estudiaron en colegios Fisco misionales o Particulares aprueban entre 3 y 6 materias, el 50% de las observaciones se concentran entre 4 y 5. Mientras que los que estudiaron en colegios fiscales aprueban entre 1 y 6 materias, el 50% de las observaciones se concentran entre 3 y 5.

Gráfico 30. Boxplot: Materias aprobadas vs. Tipo de Colegio

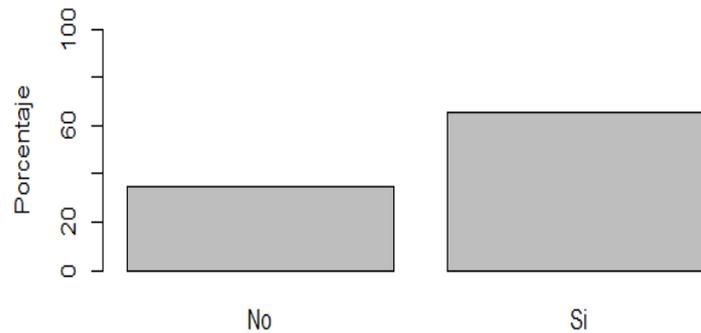


Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

### ➤ Colegio ubicado en Guayaquil

En el Gráfico 31, se puede observar que alrededor del 65% estudió en un colegio ubicado en la ciudad de Guayaquil.

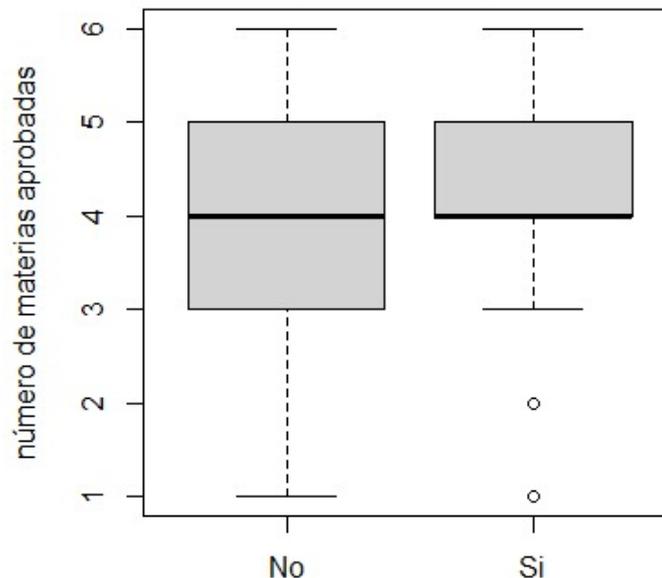
Gráfico 31. Diagrama de barra: Colegio ubicado en Guayaquil



Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

En el Gráfico 32, se puede observar que aquellos que estudiaron en colegios ubicados en Guayaquil aprueban entre 3 y 6 materias, el 50% de las observaciones se concentran entre 4 y 5. Mientras que los que no aprueban entre 1 y 6 materias, el 50% de las observaciones se concentran entre 3 y 5.

Gráfico 32. Boxplot: Materias aprobadas vs. Colegio ubicado en Guayaquil



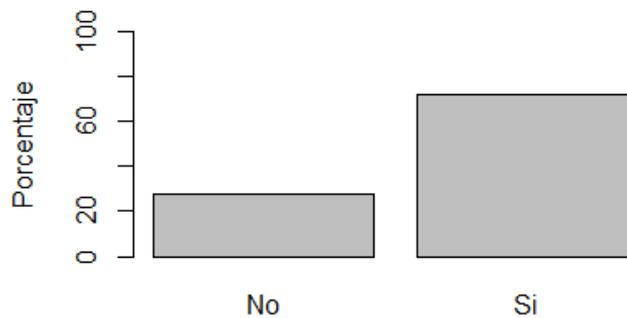
Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

### 3.1.5.3. Salud

#### ➤ Familiares afiliados al IESS

En el Gráfico 33, se observa que el 72% de la muestra posee familiares afiliados al IESS.

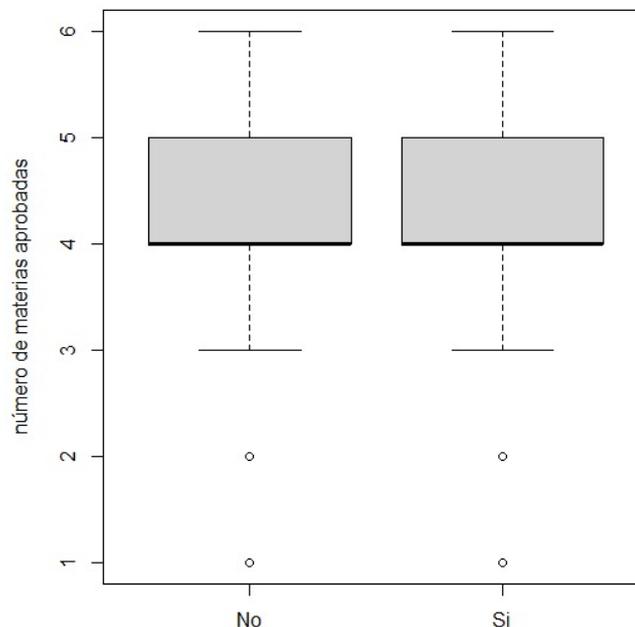
Gráfico 33. Diagrama de barra: Familiares afiliados al IESS



Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

En el Gráfico 34, se puede observar que aquellos que tienen familiares afiliados al IESS tienen una distribución similar a los que no con respecto al número de materias aprobadas.

Gráfico 34. Boxplot: Materias aprobadas vs. Familiares afiliados al IESS

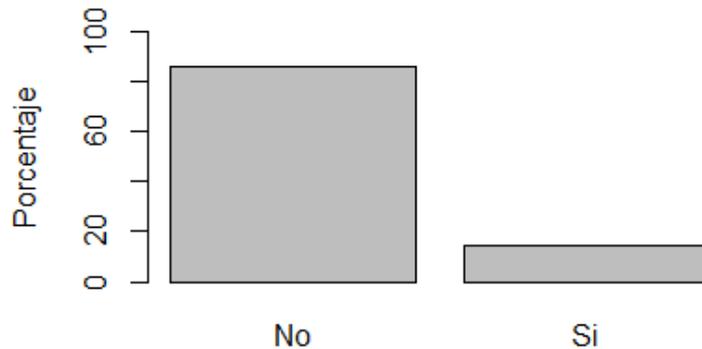


Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

### ➤ Familiares afiliados a un seguro privado

En el Gráfico 35, se observa que el 85% de la muestra no posee familiares afiliados a un seguro privado.

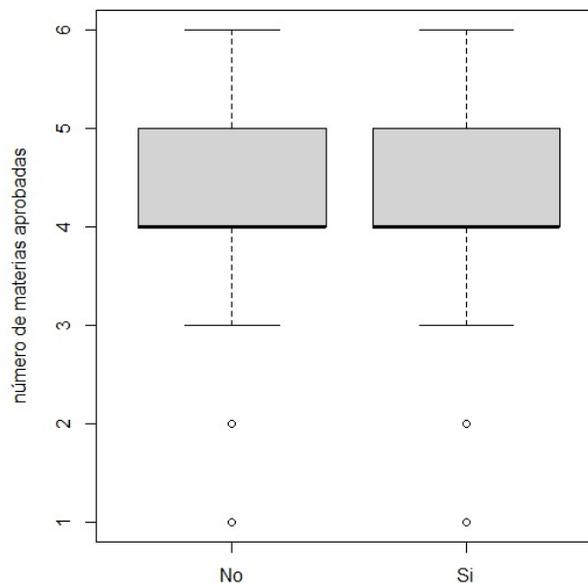
Gráfico 35. Familiares afiliados a un seguro privado



Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

En el Gráfico 36, se puede observar que aquellos que tienen familiares afiliados a un seguro privado tienen una distribución similar a los que no con respecto al número de materias aprobadas.

Gráfico 36. Boxplot: Materias aprobadas vs. Familiares afiliados a un seguro privado



Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

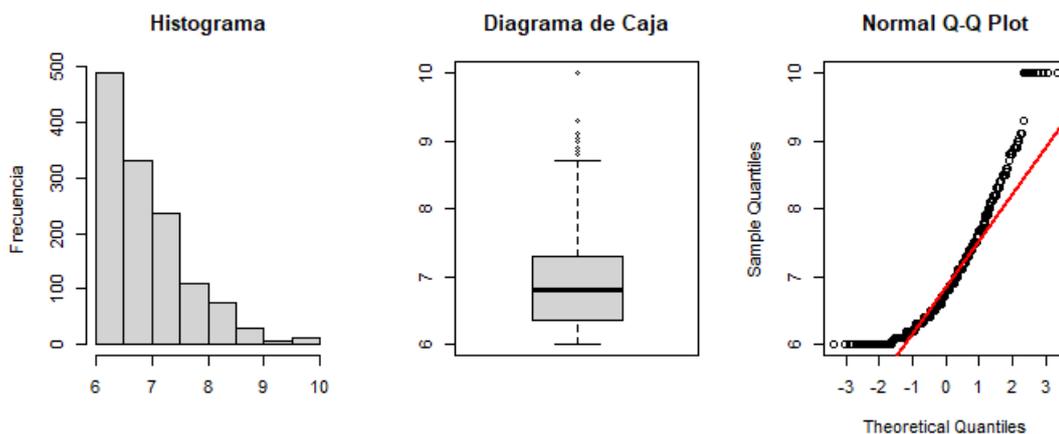
### 3.1.6. Materias del proceso de Admisiones

#### ➤ Calificación de física

En el Gráfico 37, se puede observar que las calificaciones de la materia de Física de los estudiantes de las carreras de ingeniería de la ESPOL en el primer nivel

están sesgadas hacia la izquierda, aproximadamente el 65% de las observaciones tienen calificación entre 6 y 7 como se muestra en el histograma. Las observaciones no siguen una distribución normal, los valores extremos se alejan más de la recta como se observa en el Normal Q-Q Plot. Las calificaciones 10/10 son considerados valores atípicos como se muestra en el diagrama de Caja.

**Gráfico 37. Física en el Proceso de Admisiones**

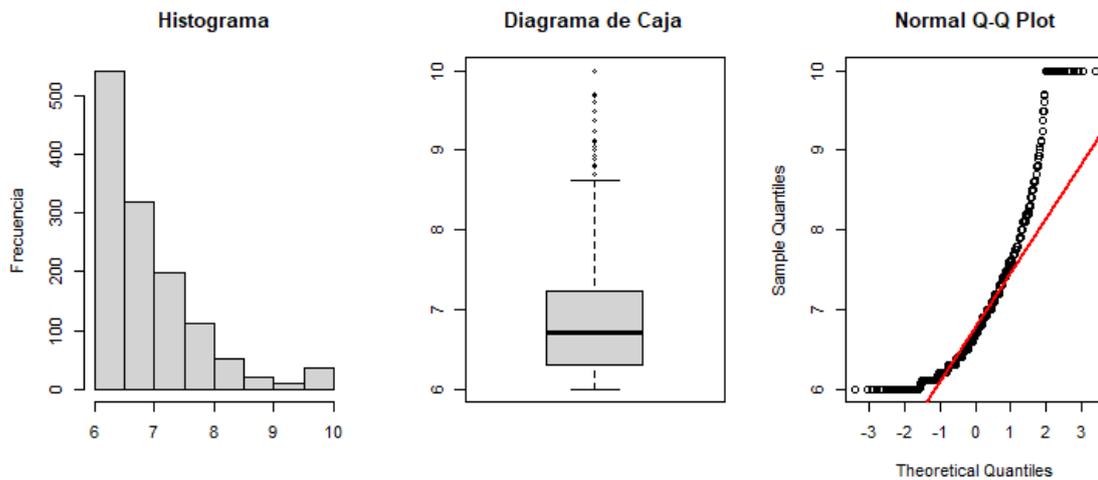


**Fuente:** Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
**Elaborado por:** J. Barzola

### ➤ **Calificación de Matemáticas**

En el *Gráfico 38*, se puede observar que las calificaciones de la materia de Matemáticas de los estudiantes de las carreras de ingeniería de la ESPOL en el primer nivel están sesgadas hacia la izquierda, aproximadamente el 65% de las observaciones tienen calificación entre 6 y 7 como se muestra en el histograma. Las observaciones no siguen una distribución normal, sobre todo en las calificaciones mayores a 8 como se observa en el Normal Q-Q Plot.

Gráfico 38. Matemáticas en el Proceso de Admisiones

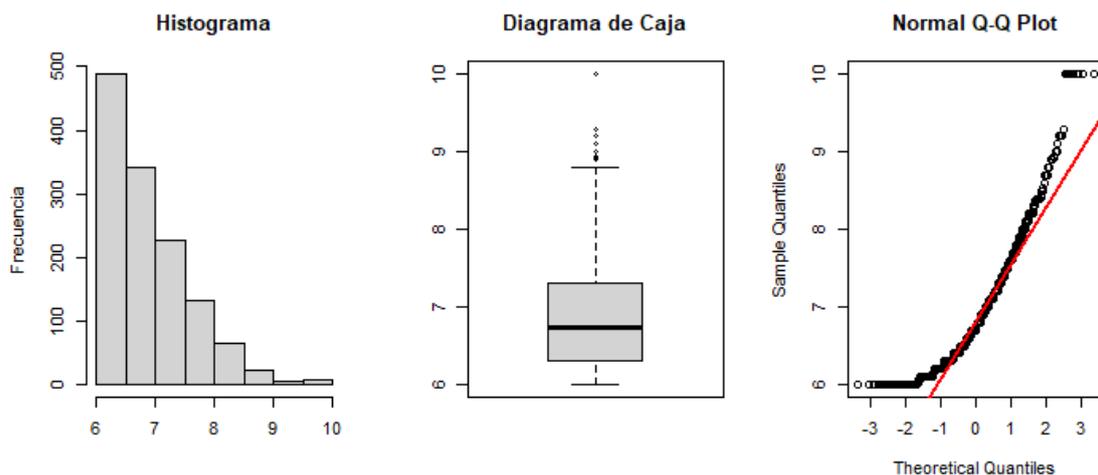


Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

### ➤ Calificación de Química

En el Gráfico 39, se puede observar que las calificaciones de la materia de Química de los estudiantes de las carreras de ingeniería de la ESPOL en el primer nivel están sesgadas hacia la izquierda, aproximadamente el 65% de las observaciones tienen calificación entre 6 y 7 como se muestra en el histograma. Tiene una distribución similar a las calificaciones de Física. Las observaciones no siguen una distribución normal, sobre todo en las calificaciones mayores a 8 como se observa en el Normal Q-Q Plot.

Gráfico 39. Calificación de la materia de Química en el Proceso de Admisiones



Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

### 3.1.7. Materias registradas en el término académico

#### ➤ Número de materias registradas por estudiante

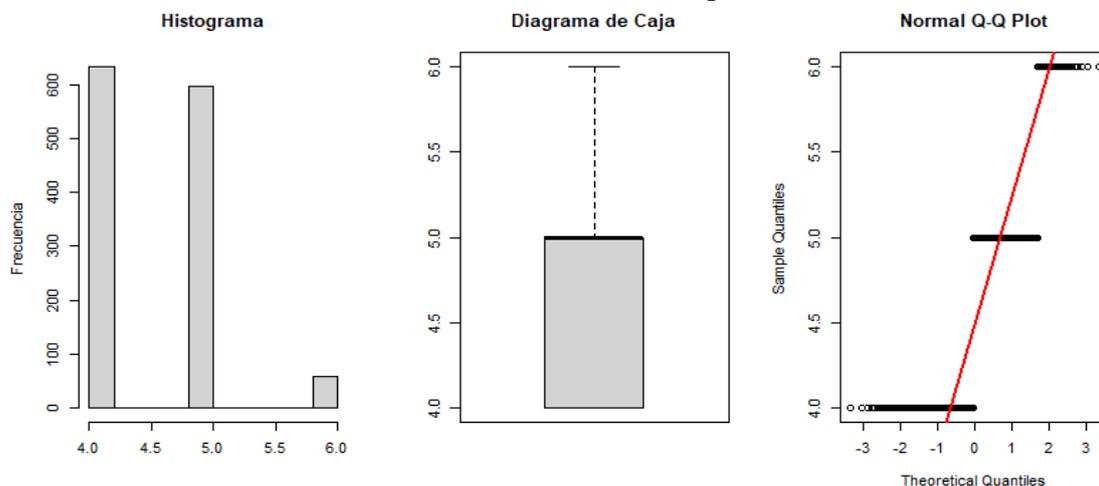
En la *Tabla 4*, se puede observar que en promedio los estudiantes toman 4.55 materias en el término académico, 2.42 materias STEM y 2.13 materias Sociales. Tiene una desviación estándar de 0.58. El máximo número de materias que se registran es 6. Aproximadamente el 75% de los estudiantes se registra en 2 materias Sociales.

**Tabla 4. Estadística descriptiva: Número de materias registradas**

Statistics	n_tom	n_stem	n_social
Mean	4,55	2,42	2,13
Std. Error of Mean	0,02	0,01	0,01
Median	5,00	2,00	2,00
Mode	4,00	2,00	2,00
Std. Deviation	0,58	0,52	0,38
Variance	0,34	0,27	0,14
Skewness	0,49	0,51	2,04
Kurtosis	-0,69	-1,14	4,66
Minimum	4,00	1,00	1,00
Maximum	6,00	4,00	4,00
	25	4,00	2,00
Percentiles	50	5,00	2,00
	75	5,00	3,00

*Fuente:* Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
*Elaborado por:* J. Barzola

**Gráfico 40. Número de materias registradas**



*Fuente:* Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
*Elaborado por:* J. Barzola

En el *Gráfico 40*, se puede observar que el número de materias registradas no sigue una distribución normal. Aproximadamente el 60% de los estudiantes se registran en 4 materias y alrededor de un 4% en 6 materias.

➤ **Número de materias aprobadas por estudiante**

En la *Tabla 5*, se puede observar que en promedio los estudiantes aprueban 4.05 materias en el término académico. Tiene una desviación estándar de 0.98. El mínimo de materias aprobadas es 1 materia y el máximo es 6.

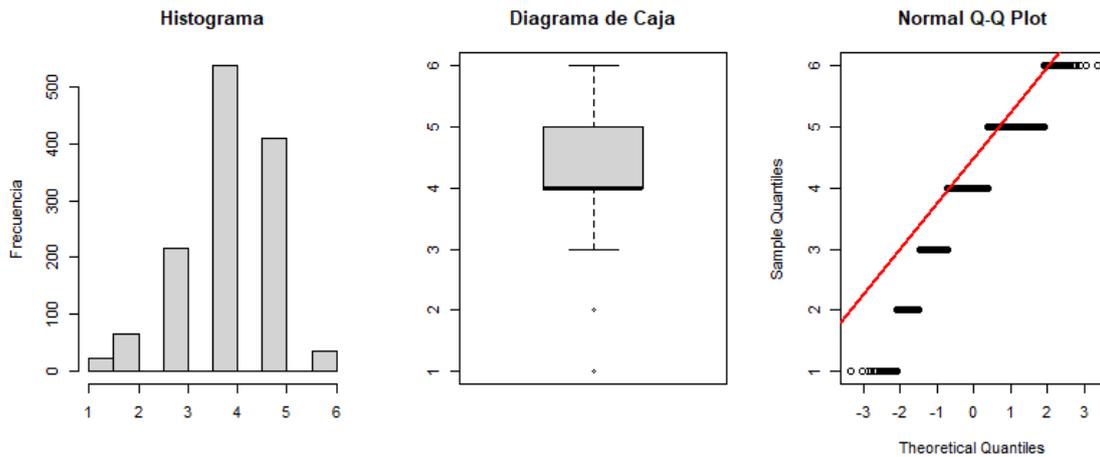
*Tabla 5. Estadística descriptiva: Número de materias aprobadas*

Statistics		n_ap
Mean		4,05
Std. Error of Mean		0,03
Median		4,00
Mode		4,00
Std. Deviation		0,98
Variance		0,97
Skewness		-0,71
Kurtosis		0,63
Minimum		1,00
Maximum		6,00
	25	4,00
Percentiles	50	4,00
	75	5,00

*Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola*

En el *Gráfico 41*, se puede observar que el número de materias registradas no sigue una distribución normal. Aproximadamente el 40% de los estudiantes aprueban 4 materias y alrededor de un 30% aprueban 5 materias.

Gráfico 41. Número de materias aprobadas



Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

➤ **Porcentaje de materias aprobadas por estudiante**

En la *Tabla 6*, se puede observar que en promedio los estudiantes aprueban el 89% de materias que toman en el término académico. Tiene una desviación estándar de 18%. El porcentaje mínimo de materias aprobadas es de 20% y el máximo es 100%.

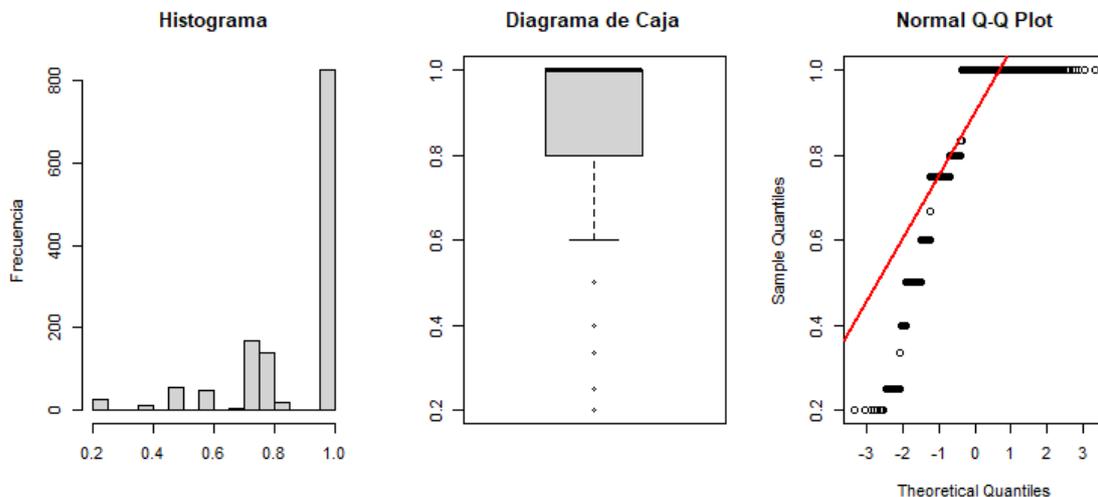
Tabla 6. Estadística descriptiva: Porcentaje de materias aprobadas

Statistics	Porcentaje_ap
Mean	0,89
Std. Error of Mean	0,00
Median	1,00
Mode	1,00
Std. Deviation	0,18
Variance	0,03
Skewness	-1,68
Kurtosis	2,55
Minimum	0,20
Maximum	1,00
Percentiles	
25	0,80
50	1,00
75	1,00

Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

En el *Gráfico 42*, se puede observar que el porcentaje de materias aprobadas no sigue una distribución normal. Aproximadamente el 65% de los estudiantes aprueban todas las materias registradas.

*Gráfico 42. Porcentaje de materias aprobadas*



*Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola*

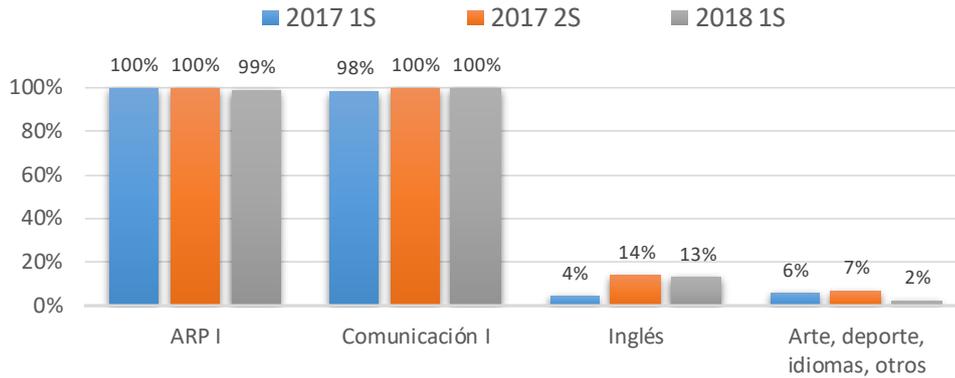
### 3.2. Índice del grado de dificultad del término académico.

Los estudiantes generalmente se registran en todas las materias disponibles o que ellos consideran fáciles, sin contar con un indicador que les permita medir la dificultad de su elección de materias.

Según la Unidad Académica que dicta la materia en la ESPOL se la asignó a un grupo por su similitud. En el grupo STEM (Science, Technology, Engineering y Mathematics) se incluyen: FCNM, FCV, FICT y FIEC. Y en el grupo Sociales se incluyen: EDCOM, ESPOL y FCSH.

En el *Gráfico 43*, se puede observar como los estudiantes se registran en materias sociales: aproximadamente el 100% de ellos en ARP I y Comunicación I. Mientras que aproximadamente el 13% busca registrarse en inglés. Alrededor del 5% de los estudiantes busca registrarse en las materias complementarias como: arte, deporte, idiomas u otros, en el último término académico se registró el 2% de los estudiantes.

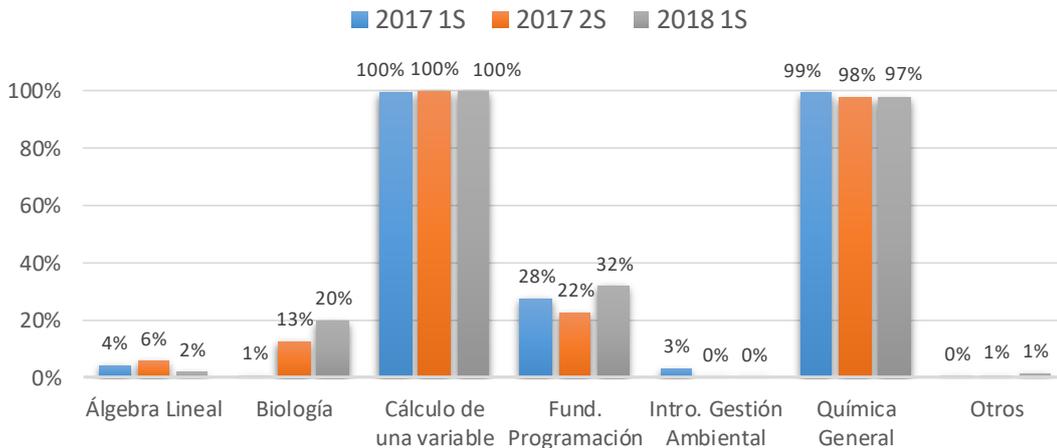
**Gráfico 43. Porcentaje de registrados en materias sociales por término académico**



Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

En el Gráfico 44, se puede observar que los estudiantes de las carreras de ingeniería de la ESPOL en el primer nivel dentro de las materias STEM se registra aproximadamente el 100% de ellos en Cálculo de una variable y Química General. Aproximadamente el 30% de los estudiantes busca registrarse en Fundamentos de Programación y alrededor del 15% se registra en Biología. En Álgebra Lineal se registra menos del 5% de los estudiantes. Y el registro en Introducción a la Gestión Ambiental es casi nulo.

**Gráfico 44. Porcentaje de registrados en materias STEM por término académico**



Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

Los estudiantes de la ESPOL no cuentan con un indicador que les permita medir el grado de dificultad de las materias que pretenden tomar en el término académico de manera global.

*Tabla 7. Tasa de reprobación de las materias por término académico*

Tipo de materia	Materia	Término Académico		
		2016 2S	2017 1S	2017 2S
<b>Sociales</b>	ARP I	3%	4%	4%
	Comunicación I	7%	13%	11%
	Inglés	10%	12%	9%
	Arte, deporte, idiomas, otros	9%	8%	8%
<b>Stem</b>	Álgebra Lineal	46%	36%	52%
	Biología	6%	4%	8%
	Cálculo de una variable	44%	30%	63%
	Intro. Gestión Ambiental	4%	0%	0%
	Química General	26%	22%	24%
	Fund. Programación	34%	30%	34%
	Otros	30%	8%	32%

*Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola*

En la *Tabla 7*, se puede observar las tasas de reprobación de las materias que toman los estudiantes de primer nivel de las carreras de ingeniería de la ESPOL por término académico:

Se puede observar que las materias sociales que los estudiantes más reprueban son Comunicación I e inglés, alrededor del 10%. Aproximadamente el 8% reprueba en las materias complementarias. Alrededor del 4% reprueban en Análisis y Resolución de Problemas.

También se puede observar que la materia STEM que los estudiantes más reprueban es Cálculo de una variable, en el último término académico reprobaron aproximadamente el 60% de los registrados. Seguido de la materia Álgebra Lineal, donde aproximadamente reprueban el 50% de los estudiantes. Los Porcentajes de reprobación para las materias Química General y Fundamentos de programación son alrededor de 25% y 35% respectivamente. Las materias que tuvieron los Porcentajes de reprobación más bajas fueron Introducción a la Gestión Ambiental (0%) y Biología (8%).

Se propone establecer un indicador con el cual los estudiantes tengan claro la complejidad de las materias que pretenden tomar y les ayude analizar el esfuerzo

que conllevará si se registra en esa combinación de materias, y de esta manera poder reducir el número de materias reprobadas por el estudiante. Además, es un indicador comparable entre estudiantes.

Este indicador se denominará “Índice del grado de dificultad del término académico” y se representará como " $gd$ ", este indicador variará en cada término académico, dado que utilizará como insumo las tasas de reprobación de materias del término anterior. Además, variará para cada estudiante dependiendo de las materias que pretenda tomar.

Es la media aritmética de las tasas de reprobación que el estudiante pretende tomar del término académico anterior. A continuación, se muestra la expresión matemática:

$$gd_t = \frac{\sum_{i=1}^n Rp_{i_{t-1}}}{n}$$

Donde:

- $Rp_i$  : tasa de reprobación de la materia tomada  $i$
- $t$  : término académico
- $n$  : número de materias tomadas por el estudiante

Es decir, en el caso que el estudiante quiera registrarse en el término 2018 – 1S en las siguientes materias:

Tipo de materia	Materia	2017 2S
Sociales	ARP I	4%
	Comunicación I	11%
Stem	Álgebra Lineal	52%
	Biología	8%
	Cálculo de una variable	63%

Se tomará las tasas de reprobación de las materias del término 2017 – 2S. En este caso el  $gd_t$  para el estudiante será 0,276. Es decir que la tasa de reprobación promedio de las materias tomadas es de 27.6%.

También se puede calcular el  $gd_t$  por grupo de materias STEM o Sociales. Para el caso de las materias STEM, se tuvo un  $gd_t$  de 0,41. Es decir que la tasa de reprobación promedio de las materias STEM tomadas es de 41%. Para el caso de las materias Sociales, se tuvo un  $gd_t$  de 0,075. Es decir que la tasa de reprobación promedio de las materias Sociales tomadas es de 7,5%.

En la *Tabla 8*, se puede observar que en promedio el grado de dificultad de las materias STEM registradas por los estudiantes es de 0.34, con una desviación estándar de 0.07. El valor mínimo fue de 0.16 y el máximo 0.50. La mediana y la moda fue de 0.35. También se puede observar que en promedio el grado de dificultad de las materias Sociales registradas por los estudiantes fue de 0.07, con una desviación estándar de 0.01. El valor mínimo fue de 0.03 y el máximo 0.14. La mediana y la moda fue de 0.07.

En el *Gráfico 45*, se puede observar que el grado de dificultad de las materias STEM no sigue una distribución normal, sin embargo, los valores están cerca de la recta. Alrededor del 55% de los estudiantes toman materias con un grado de dificultad mayor a 0.35. El 50% de las observaciones tiene un grado de dificultad STEM entre 0.27 y 0.40.

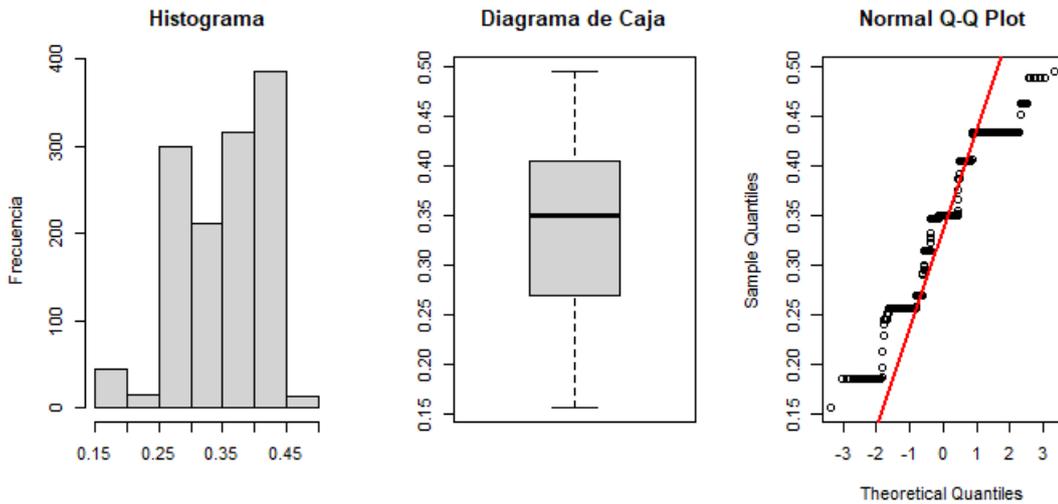
En el *Gráfico 46*, se puede observar que el grado de dificultad de las materias Sociales no sigue una distribución normal, sin embargo, los valores están cerca de la recta. Alrededor del 65% de los estudiantes toman materias con un grado de dificultad mayor a 0.07. El 50% de las observaciones tiene un grado de dificultad de las materias Sociales entre 0.06 y 0.08.

**Tabla 8. Estadística descriptiva: Porcentaje de materias aprobadas**

Statistics	gd_stem	gd_social
Mean	0,34	0,07
Std. Error of Mean	0,00	0,00
Median	0,35	0,07
Mode	0,35	0,07
Std. Deviation	0,07	0,01
Variance	0,00	0,00
Skewness	-0,27	0,22
Kurtosis	-0,74	0,44
Minimum	0,16	0,03
Maximum	0,50	0,14
	25	0,27
Percentiles	50	0,35
	75	0,40

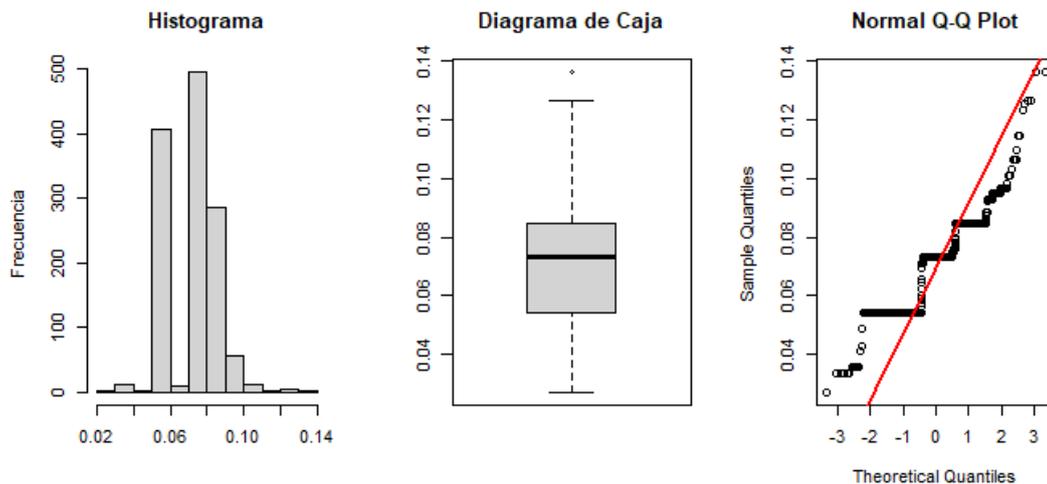
*Fuente:* Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
*Elaborado por:* J. Barzola

**Gráfico 45. Histograma, Diagrama de Caja y QQ-Normal: Grado de dificultad M. STEM**



*Fuente:* Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
*Elaborado por:* J. Barzola

Gráfico 46. Histograma, Diagrama de Caja y QQ-Normal: Grado de dificultad M. Sociales



Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

### 3.3. Aplicación de Algoritmo CHAID.

El algoritmo de CHAID se utilizó para identificar que variables influyen significativamente para que un estudiante de ingeniería apruebe todas las materias registradas en su primer semestre en la ESPOL o no. Es decir, analizar qué características tienen los estudiantes con más probabilidad de aprobar todas las materias registradas. La variable de interés para este modelo es “Aprobó todas las materias registradas”, teniendo como categorías “Si” o “No”. Se definió el nivel de significación para la división de nodos y la fusión de categorías de 0.05.

En el Gráfico 47, se puede observar en el nodo 0, que de manera general el 64% del estudiante de ingeniería que están cursando su primer semestre en la ESPOL aprueban todas las materias que toman. También se puede observar que la variable cuantitativa “Calificación de Física en el Proceso de Admisiones” es la variable que principalmente determina la aprobación de las materias de un estudiante. De esta variable se dividen 3 nuevos nodos, los cuales han sido agrupados por rangos de calificación. Entre los perfiles de los estudiantes que

aprueban todas las materias con una probabilidad aprobación mayor al 70%, los cuales se muestran a continuación:

- ✓ Con una probabilidad del 100%, son estudiantes que obtuvieron una calificación en Física mayor a 8 puntos y una calificación de Matemáticas mayor a 7,4 puntos.
- ✓ Con una probabilidad del 82,2%, son estudiantes que obtuvieron una calificación en Física mayor a 8 puntos y una calificación de Matemáticas menor o igual a 7,4 puntos.
- ✓ Con una probabilidad del 87,3%, son estudiantes que obtuvieron una calificación en Física mayor a 6,78 y menor o igual a 8 puntos, y una calificación de Matemáticas mayor a 7,91 puntos.
- ✓ Con una probabilidad del 75,2%, son estudiantes que obtuvieron una calificación en Física mayor a 6,78 y menor o igual a 8 puntos, una calificación de Matemáticas mayor a 6,37 y menor o igual a 7,91 puntos y asistieron a un colegio ubicado en la ciudad de Guayaquil.
- ✓ Con una probabilidad del 74,5%, son estudiantes que obtuvieron una calificación en Física menor o igual a 6,78, el grado de dificultad de las materias STEM mayor o igual a 0,258 y la edad es menor o igual a 18 años.

**Tabla 9. CHAID: Tabla de Clasificación**

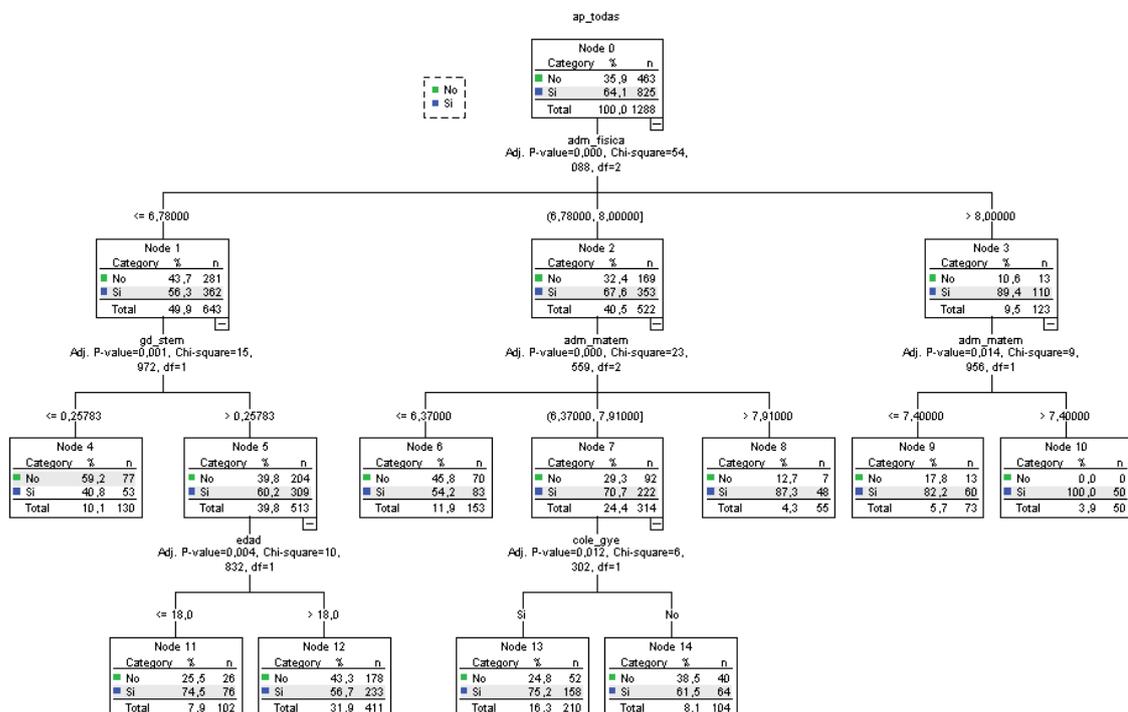
Observed	Predicted		Percent Correct
	No	Si	
No	77	386	16,6%
Si	53	772	93,6%
Overall Percentage	10,1%	89,9%	65,9%

*Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola*

La estimación de riesgo es de 0,341. Es decir, el “riesgo” de clasificar de manera errónea a un estudiante es de aproximadamente el 34,1%. En la *Tabla 9*, se puede observar que el modelo clasifica de forma correcta, aproximadamente, al 66% de los estudiantes. Además, revela un potencial problema con el modelo, dado que para los estudiantes que no aprobaron todas las materias, este modelo

pronostica correctamente sólo el 16,6% de ellos. El 83,4% de los estudiantes que no aprobaron todas las aparecen inapropiadamente clasificados como estudiantes que aprobaron todas las materias.

**Gráfico 47. Diagrama de árbol: CHAID para modelar el éxito académico en el primer nivel de las carreras de ingeniería de la ESPOL.**



Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

### 3.4. Matriz de correlación

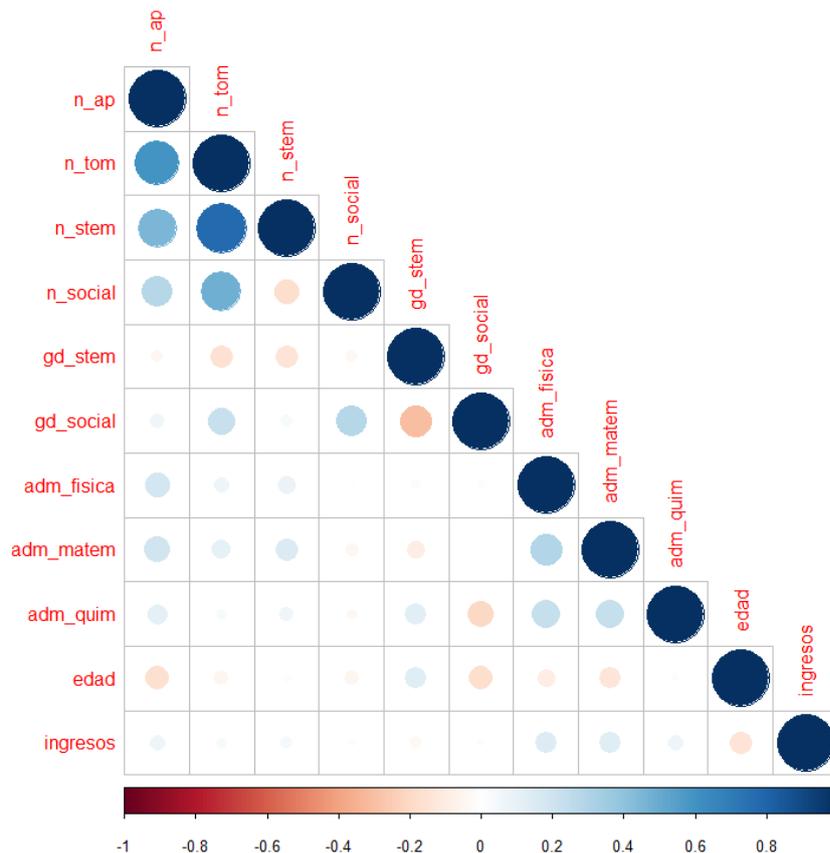
En el Gráfico 48, se muestra un Corplot con la correlación entre las variables cuantitativas que se pretenden incluir en el modelo. A medida que el valor tiende al valor -1 la esfera se muestra de color más rojiza, mientras que si tiende al valor 1 se muestra de color más azul.

La diagonal principal se muestra de color azul intenso dando que la correlación con la misma variable es 1. Según la escala de color la correlación entre el número de materias aprobadas y el número de materias registradas es

aproximadamente 0.6. El grado de dificultad de las materias stem y las sociales tienen una correlación negativa, a medida que aumenta una disminuye la otra.

La variable edad muestra correlaciones menores entre -0.2 y 0 con las variables número de materias aprobadas, calificaciones de física y química de admisiones, el grado de dificultad de las materias sociales e ingresos. Es decir, a medida que aumenta la edad, estas variables disminuyen.

Gráfico 48. Corplot de las variables del modelo



Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

En el Gráfico 49, se muestra debajo de la diagonal principal los gráficos de dispersión entre los pares de las variables. A simple vista pareciera que sólo el número de materias aprobadas está relacionado con el número de materias registradas, dado que no se muestra una tendencia en los gráficos de las demás variables.

Sin embargo, al analizar la *Tabla 10* se observó que sólo la correlación entre el número de materias aprobadas y el grado de dificultad de las materias STEM, tienen un valor de -0,04 y un valor p igual a 0,12, es decir la correlación no es estadísticamente significativa y no se puede concluir que es diferente de 0. Por lo que excluiríamos esta variable del modelo, dado que no existe relación lineal entre dichas variables.

**Tabla 10. Matriz de Correlación y valor p**

Correlación	n_ap	n_tom	n_stem	n_social	gd_stem	gd_social	adm_fisica	adm_matem	adm_quim	edad	ingresos
n_ap	1,00	0,59	0,46	0,29	-0,04	0,06	0,18	0,20	0,12	-0,17	0,07
n_tom	0,59	1,00	0,77	0,49	-0,15	0,22	0,07	0,11	0,03	-0,06	0,03
n_stem	0,46	0,77	1,00	-0,18	-0,14	0,04	0,09	0,15	0,06	-0,02	0,05
n_social	0,29	0,49	-0,18	1,00	-0,04	0,29	-0,01	-0,05	-0,03	-0,06	-0,01
gd_stem	-0,04	-0,15	-0,14	-0,04	1,00	-0,31	-0,02	-0,09	0,13	0,13	-0,04
gd_social	0,06	0,22	0,04	0,29	-0,31	1,00	0,02	-0,01	-0,21	-0,17	0,02
adm_fisica	0,18	0,07	0,09	-0,01	-0,02	0,02	1,00	0,29	0,24	-0,10	0,14
adm_matem	0,20	0,11	0,15	-0,05	-0,09	-0,01	0,29	1,00	0,23	-0,14	0,13
adm_quim	0,12	0,03	0,06	-0,03	0,13	-0,21	0,24	0,23	1,00	0,01	0,07
edad	-0,17	-0,06	-0,02	-0,06	0,13	-0,17	-0,10	-0,14	0,01	1,00	-0,14
ingresos	0,07	0,03	0,05	-0,01	-0,04	0,02	0,14	0,13	0,07	-0,14	1,00

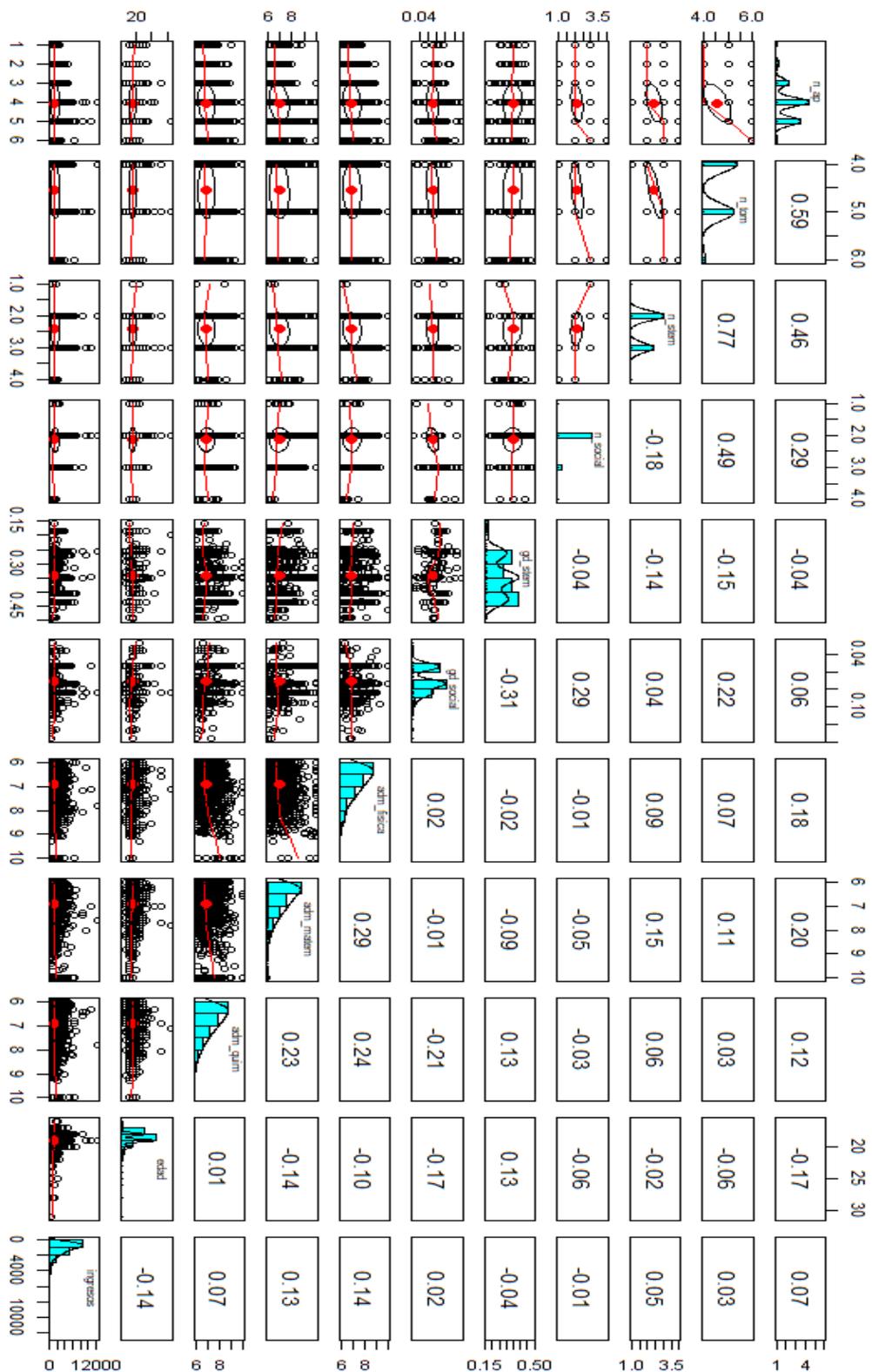
  

Valor p	n_ap	n_tom	n_stem	n_social	gd_stem	gd_social	adm_fisica	adm_matem	adm_quim	edad	ingresos
n_ap		0,00	0,00	0,00	0,11	0,03	0,00	0,00	0,00	0,00	0,01
n_tom	0,00		0,00	0,00	0,00	0,00	0,01	0,00	0,22	0,05	0,24
n_stem	0,00	0,00		0,00	0,00	0,17	0,00	0,00	0,03	0,47	0,10
n_social	0,00	0,00	0,00		0,15	0,00	0,72	0,10	0,27	0,04	0,66
gd_stem	0,11	0,00	0,00	0,15		0,00	0,37	0,00	0,00	0,00	0,15
gd_social	0,03	0,00	0,17	0,00	0,00		0,50	0,71	0,00	0,00	0,55
adm_fisica	0,00	0,01	0,00	0,72	0,37	0,50		0,00	0,00	0,00	0,00
adm_matem	0,00	0,00	0,00	0,10	0,00	0,71	0,00		0,00	0,00	0,00
adm_quim	0,00	0,22	0,03	0,27	0,00	0,00	0,00	0,00		0,63	0,01
edad	0,00	0,05	0,47	0,04	0,00	0,00	0,00	0,00	0,63		0,00
ingresos	0,01	0,24	0,10	0,66	0,15	0,55	0,00	0,00	0,01	0,00	

*Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I*

*Elaborado por: J. Barzola*

Gráfico 49. Matriz de Dispersión, Histograma y Correlación



Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

### 3.5. Aplicación de Modelos de Regresión Multivariante.

En esta sección se analizarán las variables que se pretenden incluir en el modelo de regresión. Se tiene como variable objetivo el número de materias que aprueban los estudiantes de primer nivel de las carreras de ingeniería de la ESPOL.

En la *Tabla 11*, se enlista las variables que se incluirán en el modelo de regresión multivariante luego de haber sido analizadas de manera univariante y bivariante.

**Tabla 11. Variables a Incluir en el modelo de regresión multivariada**

Nombre de la Variable	Categorías	Categorías recodificadas	Nueva Variable	Código Nueva Variable
Género del estudiante	0 Femenino 1 GLBTI 2 Masculino	0 Femenino o GLBTI 1 Masculino	-	género
Edad del estudiante	-	-	-	edad
Estado Civil familiar	0 Casado/a 1 Divorciado/a 2 Soltero/a 3 Unión Libre 4 Viudo/a	0 Sin pareja: soltero, divorciado o viudo 1 Con pareja: casado o unión libre	Estado Civil familiar	est_civil_fa
Parentesco con el jefe de hogar	0 Padre 1 Madre 2 Otro	0 No: madre u otro 1 Si: padre	El padre es el jefe del hogar	padre_jefeh
Nivel de Instrucción del jefe de hogar	0 Ninguno 1 Primaria 2 Secundaria 3 Tercer nivel 4 Postgrado	0 No: Ninguno, primaria, secundaria. 1 Si: tercer nivel, postgrado.	Jefe del hogar tiene tercer nivel de instrucción	jefeh_tercer_niv
Total de Ingresos del Hogar	-	-	-	ingresos
Estrato Socioeconómico	0 Bajo 1 Medio Bajo 2 Medio 3 Medio Alto 4 Alto	0 Bajo: bajo, medio bajo 1 Alto: medio, medio alto, alto	-	estrat_socioec
Número de celulares en el hogar	0 Ninguno 1 Tiene 1 2 Tiene 2 3 Tiene 3 4 Tiene 4 o más	0 Tiene 2 o menos 1 Tiene 3 o más	-	n_cell
Tipo de vivienda	0 Cuarto 1 Departamento 2 Vivienda	0 Cuarto o departamento 1 Vivienda	-	tipo_viv_resid

Tiene celular	0 No 1 Si	-	-	tcell
Tiene teléfono convencional	0 No 1 Si	-	-	ttelf_conv
Tiene servicio de Internet	0 No 1 Si	-	-	tinternet
Tienen servicio de tv. Pagada	0 No 1 Si	-	-	ttv_pagad
Colegio ubicado en la Ciudad de Guayaquil	0 No 1 Si	-	-	cole_gye
Tipo de colegio	0 Fiscal 1 Fisco misional 2 Particular	-	-	tipo_cole
Calificación de Física en el proceso de Admisiones	-	-	-	adm_fisica
Calificación de Matemáticas en el proceso de Admisiones	-	-	-	adm_matem
Calificación de Química en el proceso de Admisiones	-	-	-	adm_quim
Número de materias tomadas	-	-	-	n_tom
Índice del grado de dificultad de las materias Sociales	-	-	-	gd_social

*Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola*

En primero se separó aleatoriamente las observaciones en dos grupos, el 70% de ellas fueron usadas para modelar y el 30% restante para probar el mismo.

### 3.5.1. Modelo original

A continuación, se muestra el modelo original correspondiente al número de materias aprobadas, al cual se realizó un análisis de regresión lineal múltiple con las siguientes variables:

$$\begin{aligned}
 n_{ap} = & n_{tom} + gd_{social} + adm_{fisica} + adm_{matem} + adm_{quim} \\
 & + género + edad + est_{civil\_fam} + padre_{jefeh} \\
 & + jefeh_{tercer\_niv} + ingresos + estrat_{socioec} + n_{cell} \\
 & + tipo_{viv\_resid} + tcell + ttelf_{conv} + tinternet \\
 & + ttv_{pagad} + cole_{gye} + tipo_{cole}
 \end{aligned}$$

En la *Tabla 12*, se puede observar que el modelo de regresión lineal obtenido no tiene un buen ajuste, dado que el coeficiente de determinación es  $R^2 = 0,3903$ . Es decir, el 39 % de la variabilidad de los datos queda explicado, por otra parte, el coeficiente de determinación ajustado alcanzó el 37,57 %.

Además, dado que el valor  $p$  es aproximadamente 0, existe evidencia estadística para rechazar la hipótesis nula que todos los coeficientes son igual a 0. También se muestran los coeficientes del modelo de regresión obtenidos con el método de mínimos cuadrados ordinarios, con ayuda del estadístico *t de Student* con sus correspondientes valores  $p$ .

En la *Tabla 13*, se pueden observar las variables que tienen influencia significativa en la variable número de materias aprobadas, entre estas variables se obtuvieron: número de materias registradas, índice del grado de dificultad de las materias sociales, calificaciones de física y matemáticas en el proceso de admisiones, edad y el jefe del hogar es el padre.

*Tabla 12. Método de mínimos cuadrados: Modelo Original*

Coefficients:	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.100867490	0.624328665	0.162	0.871688
n_tom	0.940817574	0.046350789	20.298	< 2e-16 ***
gd_social	-5.408231939	1.914493342	-2.825	0.004837 **
adm_fisica	0.133913280	0.035901872	3.730	0.000204 ***
adm_matem	0.071466299	0.032479717	2.200	0.028042 *
adm_quim	0.062311301	0.038672396	1.611	0.107482
género[T.Masculino]	-0.086727203	0.056847702	-1.526	0.127467
edad	-0.104736491	0.022399929	-4.676	0.00000339 ***
est_civil_fam [T.Con_parej]	-0.052299735	0.051993612	-1.006	0.314746
padre_jefeh[T.Si]	0.088669717	0.066110779	1.341	0.180192
jefeh_tercer_niv[T.Si]	-0.006736014	0.056403917	-0.119	0.904966
ingresos	-0.000007682	0.000024061	-0.319	0.749600
estrat_socioec[T.Alto]	0.067422666	0.042096676	1.602	0.109600
n_cell[T.Tiene 3 o más]	-0.001428661	0.061487789	-0.023	0.981468
tipo_viv_resid[T.vivienda]	-0.005225989	0.055648139	-0.094	0.925201
tcell[T.Si]	0.153359066	0.093157762	1.646	0.100074
ttelf_conv[T.Si]	0.052014968	0.064731453	0.804	0.421874
tinternet[T.Si]	0.074094662	0.087826133	0.844	0.399093
ttv_pagad[T.Si]	-0.049602837	0.059035287	-0.840	0.401011
cole_gye[T.Si]	0.055988443	0.057622851	0.972	0.331499
tipo_cole[T.Fiscomisional]	-0.067524269	0.048230360	-1.400	0.161855
tipo_cole[T.Particular]	-0.070545453	0.053760898	-1.312	0.189792

Residual standard error: 0.7592 on 880 degrees of freedom  
Multiple R-squared: 0.3903, Adjusted R-squared: 0.3757  
F-statistic: 26.83 on 21 and 880 DF, p-value: < 2.2e-16

Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

Tabla 13. Análisis de varianza del Modelo Original

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
n_tom	1	270.92	270.925	470.0291	< 2.2e-16	***
gd_social	1	5.00	5.002	8.6787	0.003304	**
adm_fisica	1	16.14	16.143	28.0065	0.000001527	***
adm_matem	1	5.88	5.881	10.2037	0.001451	**
adm_quim	1	1.29	1.291	2.2406	0.134785	
género	1	2.26	2.259	3.9190	0.048055	*
edad	1	15.00	14.997	26.0179	0.0000004144	***
est_civil_fam	1	0.02	0.018	0.0319	0.858233	
padre_jefeh	1	0.99	0.990	1.7176	0.190338	
jefeh_tercer_niv	1	0.00	0.003	0.0059	0.938546	
ingresos	1	0.00	0.002	0.0027	0.958700	
estrat_socioec	1	1.44	1.444	2.5056	0.113801	
n_cell	1	0.08	0.083	0.1447	0.703744	
tipo_viv_resid	1	0.06	0.055	0.0955	0.757339	
tcell	1	1.49	1.486	2.5781	0.108709	
ttelf_conv	1	0.35	0.351	0.6089	0.435416	
tinternet	1	0.23	0.229	0.3981	0.528237	
ttv_pagad	1	0.60	0.601	1.0419	0.307651	
cole_gye	1	0.21	0.206	0.3566	0.550578	
tipo_cole	2	2.74	1.368	2.3726	0.093835	.
Residuals	880	507.23	0.576			

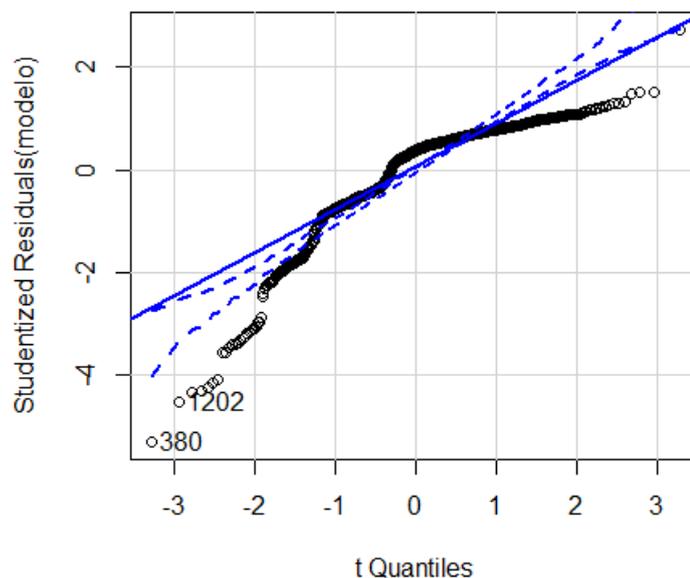
Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

### 3.5.1.1. Supuestos del modelo

Se procedió a la validación de los supuestos inherentes al modelo de regresión estimado, los resultados fueron los siguientes:

#### a. Normalidad:

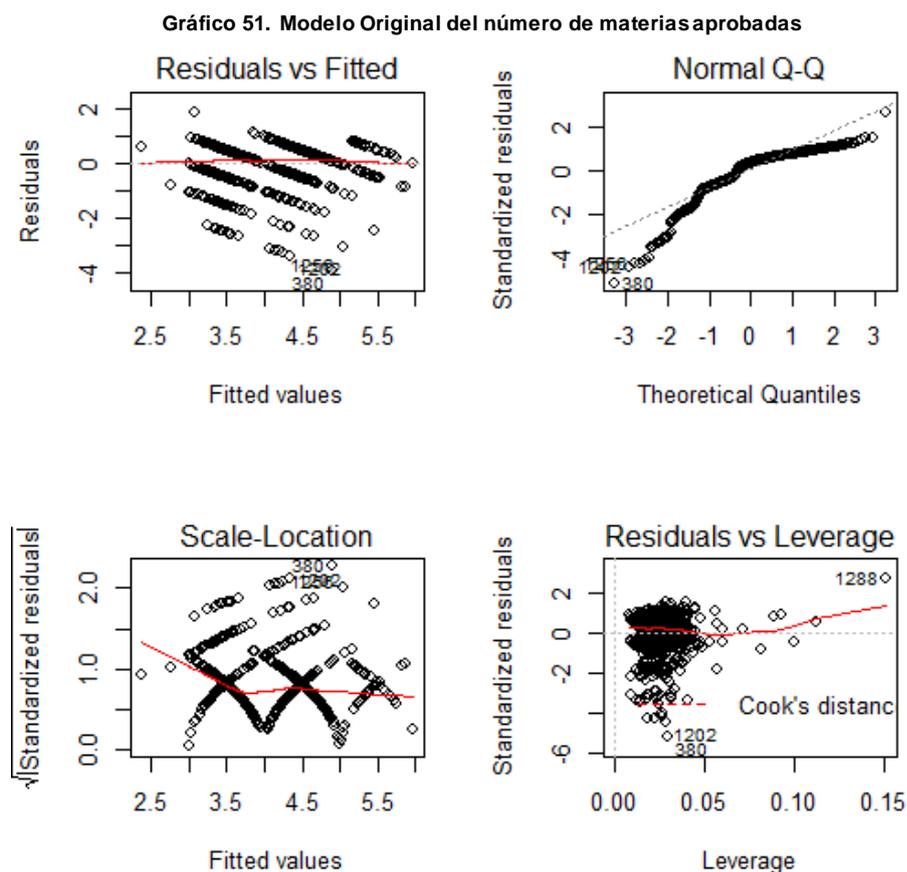
Gráfico 50. Normal Q-Q de los residuos del Modelo Original



Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

En el *Gráfico 50*, los puntos no se ubican a lo largo de la línea, es decir, para este caso no muestran una distribución normal. También se utilizó el test de Shapiro Wilks, donde su hipótesis nula sostiene que los errores del modelo siguen una distribución normal, se tuvo un estimado de 0.849, y dado que su valor  $p$  es aproximadamente 0, se corrobora que no se cumple con el supuesto.

### b. Homogeneidad de los residuos



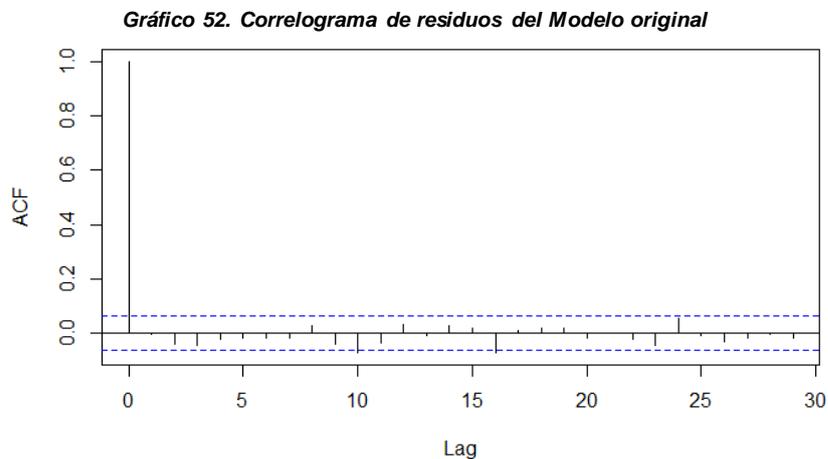
Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

En el *Gráfico 51*, “Residuals vs. Fitted” debería aparecer un patrón de dispersión aleatoria de puntos. Sin embargo, se observó una curva que decrece los residuos para los valores predichos más grandes. La varianza de los residuos debe ser similar a lo largo de las predicciones del modelo. Se utilizó el test studentized Breusch-Pagan para probar este supuesto, se obtuvo un estimado de 47,71 , y

un valor p aprox. 0. Por lo tanto, existió evidencia estadística para rechazar dicha hipótesis, existe heterocedasticidad de los residuos.

### c. Independencia de los residuos:

El test de Durbin-Watson, se utilizó para se utilizó para detectar la autocorrelación de los residuos, tiene la hipótesis nula de que la autocorrelación de los residuos es 0, se tuvo un estimado de 2.01, y dado que su valor p es aproximadamente 0.93, si se cumple dicha condición. Además, se analizó el *Gráfico 52*, donde no se mostró picos más largos de residuos por encima, ni por debajo de la línea de correlación, es decir que no existe autocorrelación significativa en los residuales.

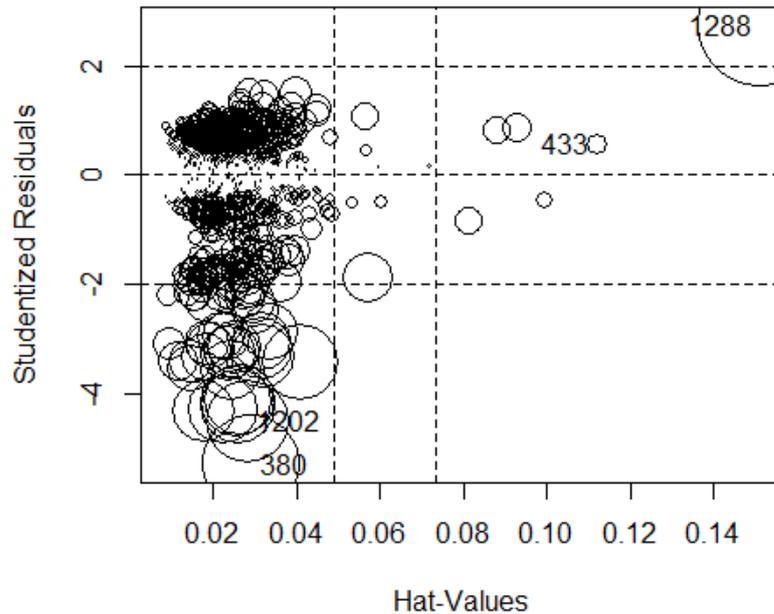


*Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola*

### d. Puntos influyentes y puntos perdidos.

En la exploración de las observaciones para detectar puntos influyentes y perdidos se utilizó un Influenceplot como se muestra en el *Gráfico 53*. Se encontró que las observaciones que tuvieron mayor influencia fueron: 380, 433, 1202, 1288.

Gráfico 53. Influenceplot del Modelo original



Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

### e. Multicolinealidad

En la *Tabla 14*, se puede observar los valores de *VIF* para cada variable. No se encontró multicolinealidad entre las variables.

Tabla 14. Modelo Original: Factor de Inflación de Varianza (VIF)

n_tom	1.096630	gd_social	1.174175
adm_fisica	1.174822	adm_matem	1.183536
adm_quim	1.177665	género[T.Masculino]	1.049254
edad	1.117584	est_civil_fam.L	1.501143
padre_jefeh[T.Si]	1.478237	jefeh_tercer_niv[T.Si]	1.197246
ingresos	1.434096	estrat_socioec.L	1.351251
n_cell[T.Tiene 3 o más]	1.258946	tipo_viv_resid[T.Vivienda]	1.149533
tcell[T.Si]	1.072826	ttelf_conv[T.Si]	1.094468
tinternet[T.Si]	1.137402	ttv_pagad[T.Si]	1.360515
cole_gye[T.Si]	1.182616	tipo_cole.L	1.247411
tipo_cole.Q	1.083686		

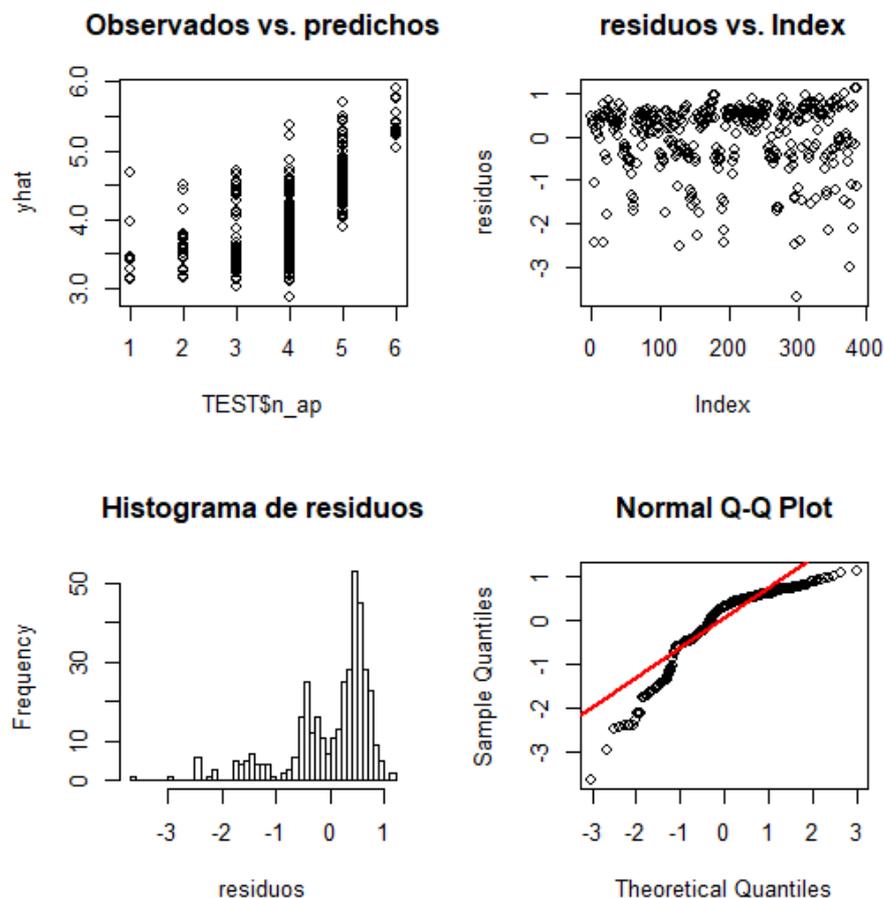
Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

### 3.5.1.2. Validación del modelo

Una vez generado el modelo se estima el número de materias aprobadas con el 30% de las observaciones restantes no empleadas. La suma cuadrática de los residuos de la validación del modelo es 235,70.

En el *Gráfico 54*, se muestra cómo se distribuyen los residuos del modelo validado. Donde se observa que no hay un buen ajuste.

*Gráfico 54. Validación del Modelo original*



*Fuente:* Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
*Elaborado por:* J. Barzola

### 3.5.2. Modelo Final

Luego de haber probado varios modelos, se muestra a continuación el modelo Final correspondiente al número de materias aprobadas:

$$n_{ap}^3 \sim \beta_0 n_{tom} + \beta_1 gd_{social} + \beta_2 adm_{fisica} + \beta_3 edad$$

En la *Tabla 15*, se puede observar que el modelo de regresión lineal obtenido tiene un buen ajuste, dado que el coeficiente de determinación es  $R^2 = 0.5133$ , es decir, el 51% de la variabilidad de los datos queda explicado; por otra parte, el coeficiente de determinación ajustado también alcanzó el 51%, aproximadamente 15% más que el modelo original. Además, dado que el valor  $p$  es aproximadamente 0, existe evidencia estadística para rechazar la hipótesis nula que todos los coeficientes son igual a 0. También se muestra los coeficientes del modelo de regresión obtenidos con el método de mínimos cuadrados ordinarios, con ayuda del estadístico  $t$  de *Student* con sus correspondientes valores  $p$ .

En la *Tabla 16*, se puede observar que todas las variables tienen influencia significativa en la variable número de materias aprobadas estas variables son: número de materias registradas, calificaciones de matemáticas y física de admisiones, edad y grado de dificultad de las materias sociales.

*Tabla 15. Método de mínimos cuadrados: Modelo Final*

Coefficients:	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-100.3334	22.3425	-4.491	0.0000080246 ***
n_tom	54.2302	1.8755	28.915	< 2e-16 ***
gd_social	-315.2579	75.4943	-4.176	0.0000325751 ***
adm_fisica	6.9345	1.3717	5.055	0.0000005211 ***
edad	-5.0113	0.8859	-5.657	0.0000000208 ***
---				
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Residual standard error: 31.23 on 897 degrees of freedom				
Multiple R-squared: 0.5133, Adjusted R-squared: 0.5111				
F-statistic: 236.5 on 4 and 897 DF, p-value: < 2.2e-16				

Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

A continuación, se interpreta el valor de los betas:

- ✓ El coeficiente  $\beta_1$  indica que, por cada materia tomada adicional, el número de materias aprobadas aumenta en  $\sqrt[3]{54,23} = 3,79$ .

- ✓ El coeficiente  $\beta_2$  indica que por cada punto porcentual que aumenta el índice del grado de dificultad de las materias sociales, el número de materias aprobadas aumenta en  $\sqrt[3]{-315,26} / 100 = -0,0681$ .
- ✓ El coeficiente  $\beta_3$  indica que por cada punto adicional en la calificación de Física en el proceso de admisiones, el número de materias aprobadas aumenta en  $\sqrt[3]{6,93} = 1,91$ .
- ✓ El coeficiente  $\beta_4$  indica que la edad por cada año adicional, el número de materias aprobadas aumenta en  $\sqrt[3]{-5,01} = -1,71$ .

Tabla 16. Análisis de varianza del Modelo Final

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
n_tom	1	850600	850600	872.187	< 2.2e-16	***
gd_social	1	10805	10805	11.079	0.0009087	***
adm_fisica	1	29823	29823	30.580	0.00000004203	***
edad	1	31206	31206	31.998	0.00000002075	***
Residuals	897	874799	975			
---						
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						

Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

### 3.5.2.1. Supuestos del modelo

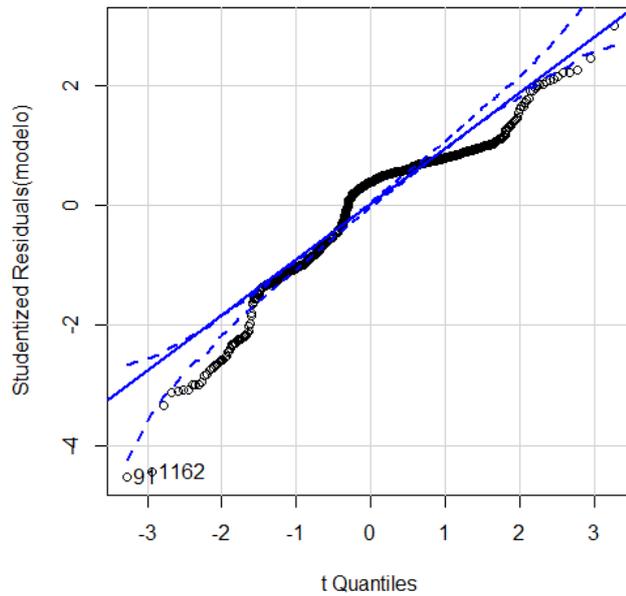
Se procedió a la validación de los supuestos inherentes al modelo de regresión estimado, los resultados fueron los siguientes:

#### a. Normalidad:

En el *Gráfico 55*, se puede observar que los puntos no se ubican a lo largo de la línea, es decir, para este caso no muestran una distribución normal. Sin embargo, se ajusta mejor que el modelo original, los puntos están más cercanos a la línea.

También se utilizó el test de Shapiro Wilks, donde su hipótesis nula sostiene que los errores del modelo siguen una distribución normal, se tuvo un estimado de 0.92, y dado que su valor  $p$  es aproximadamente 0, se corrobora que no se cumple con el supuesto.

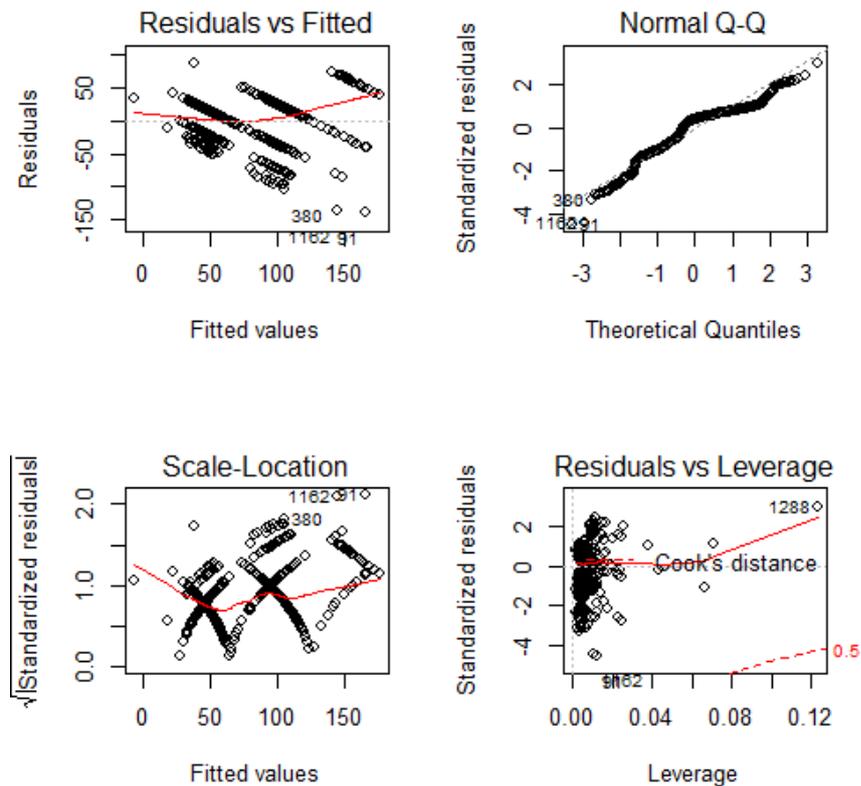
Gráfico 55. Normal Q-Q de los residuos del Modelo Final



Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

**b. Homogeneidad de los residuos:**

Gráfico 56. Modelo Final del número de materias aprobadas



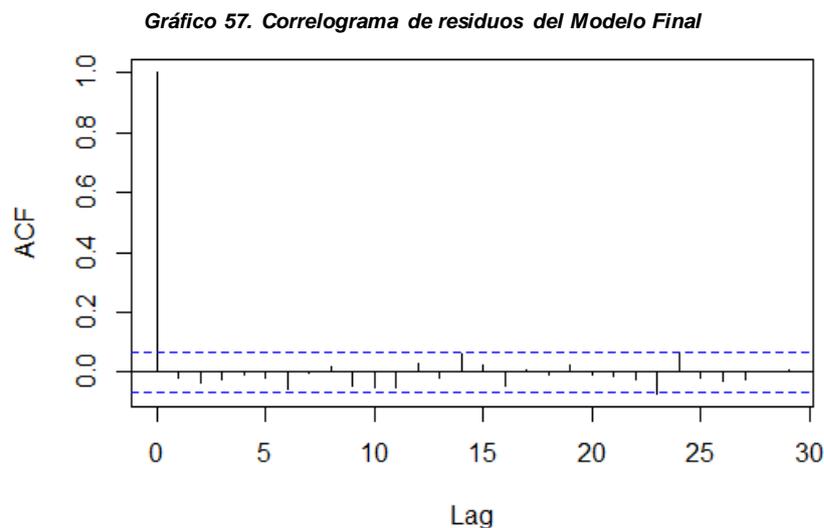
Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

En el *Gráfico 56* “Residuals vs. Fitted”, se puede observar que una curva donde los residuos crecen para los valores predichos más grandes. Se utilizó el test studentized Breusch-Pagan para probar el supuesto de homogeneidad, se obtuvo un estimado de 178.48, y un valor p aproximadamente 0. Por lo tanto, existió evidencia estadística para rechazar dicha hipótesis, existe heterocedasticidad de los residuos.

### c. Independencia de los residuos:

El test de Durbin-Watson, se utilizó para se utilizó para detectar la autocorrelación de los residuos, tiene la hipótesis nula de que la autocorrelación de los residuos es 0, se tuvo un estimado de 2.04, y dado que su valor p es aproximadamente 0.56, si se cumple dicha condición.

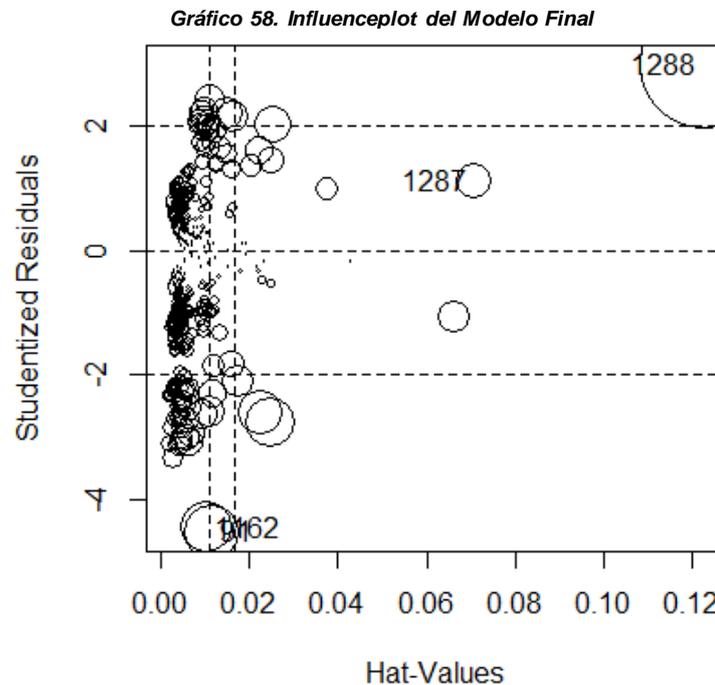
Además, en el *Gráfico 57* se puede observar en el Correlograma de los residuos que no se mostró picos más largos de residuos por encima, ni por debajo de la línea de correlación, es decir que no existe autocorrelación significativa en los residuales.



Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

**d. Puntos influyentes y puntos perdidos.**

En la exploración de las observaciones para detectar puntos influyentes y perdidos se utilizó un Influenceplot como se muestra en el *Gráfico 58*. Se encontró que las observaciones que tuvieron mayor influencia fueron: 91, 1162, 1287, 1288.



Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

**e. Multicolinealidad**

En la *Tabla 18*, se puede observar los valores de *VIF* para cada variable. No se encontró multicolinealidad entre las variables.

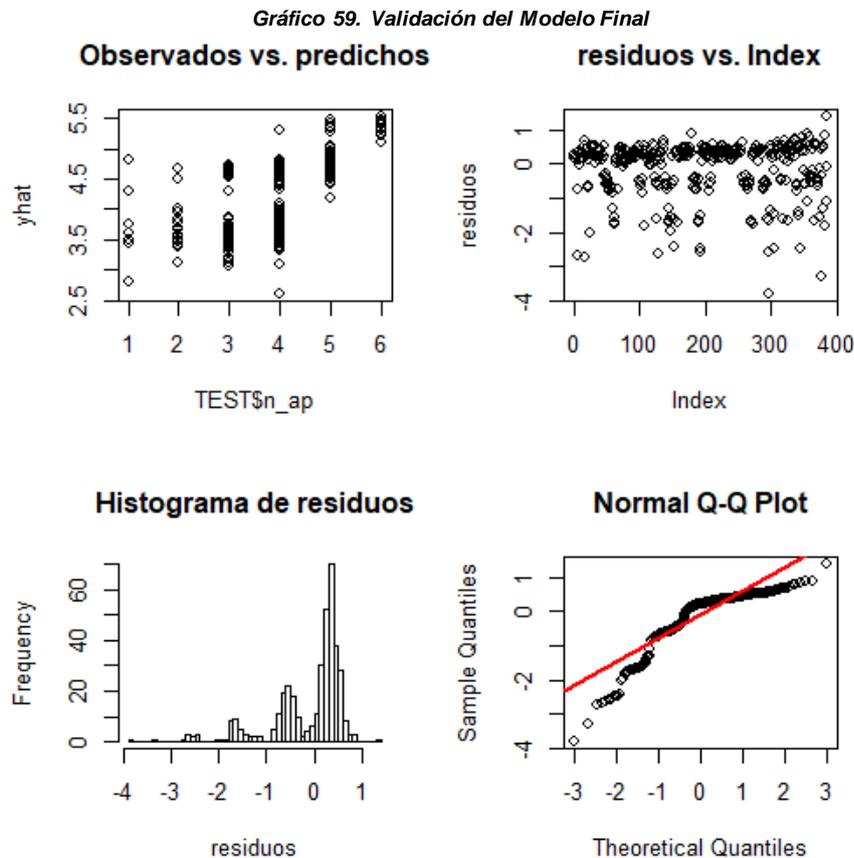
Tabla 17. Factor de Inflación de Varianza (VIF)

n_tom	gd_social	adm_fisica	edad
1.061154	1.079101	1.013664	1.033191

Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola

### 3.5.2.2. Validación del modelo

Se calculó la raíz cúbica de los valores estimados, con la finalidad de poder compararlos con el modelo original. La suma cuadrática de los residuos de la validación del modelo es 241. En el *Gráfico 59*, se muestra cómo se distribuyen los residuos del modelo validado. Donde se observa que no hay un buen ajuste.



*Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola*

### 3.5.3. Resumen de los modelos

En la *Tabla 19*, que se compara el modelo Original versus al modelo Final. Se puede concluir que el modelo Final es el mejor, dado que con cuatro variables tiene un porcentaje de variabilidad del 51% de los datos, mientras que el modelo Original con 20 variables representa el 39% de la variabilidad de los datos.

*Tabla 18. Comparación de los modelos*

	<b>Modelo Original</b>	<b>Modelo Final</b>
Variab <sup>l</sup> es	20	4
R <sup>2</sup>	0.39	0.513
Adjusted R <sup>2</sup>	0.376	0.511
Residual Std. Error	0.759	31.2
F Statistic	26.8	236

*Fuente: Base de datos STA – ESPOL desde el término 2016-I al 2018-I  
Elaborado por: J. Barzola*

A pesar de que el modelo de regresión representa el 51% de la variabilidad total de la variable número de materias aprobadas y no se cumplen todos los supuestos, sirve para tener una referencia de las variables que influyen significativamente en ella.

Algunos de los factores que influyen significativamente en modelo son las calificaciones obtenida de Física y Matemáticas en el proceso de admisiones. Así mismo en el algoritmo de CHAID que se aplicó se encontró que estas variables son las que generalmente determinan si el estudiante aprueba o no todas las materias registradas. A pesar que los estudios realizados por Carvajal, Mosquera & Artamonova (2009), Rodríguez & Ruiz (2011) y Cardona & Hernández (2013) no incluye en su modelo la variable calificación de física, se observa que es significativamente influyente los conocimientos adquiridos previamente antes de ingresar a una universidad.

La variable edad es uno de los factores que influye significativamente en modelo que predice el número de materias aprobadas por el estudiante. Así mismo en el algoritmo de CHAID que se aplicó se encontró que esta variable influye significativamente para que el estudiante apruebe todas sus materias cuando ha obtenido una baja calificación de Física en el proceso de Admisiones. Se obtuvo de manera similar este resultado en el estudio realizado por Cardona & Hernández (2013), Martínez, Fernández, Couñago, Vacas, da Silva & González (2014), Tinsaray & Karina (2016). Lo que contradice al estudio realizado por Carvajal, Mosquera & Artamonova (2009).

Similar al estudio realizado por Carvajal, Mosquera & Artamonova (2009) se descartaron variables como: estrato socio-económico, tipo de colegio, cobertura de salud.

La variable género influye significativamente en el número de materias aprobadas por el estudiante. Se obtuvo de manera similar este resultado en el estudio realizado por Rodríguez & Ruiz (2011), Cardona & Hernández (2013), Planck Barahona (2014), Martínez, Fernández, Couñago, Vacas, da Silva & González (2014). Lo que contradice al estudio realizado por Tinsaray & Karina (2016).

También el índice del grado de dificultad de las materias Sociales influye significativamente de manera negativa en el número de materias aprobadas. Es decir, mientras mayor sea el promedio de las tasas de reprobación de las materias Sociales que se registra el estudiante, aumente la probabilidad que disminuya el número de materias aprobadas.

Las variables “tiene laptop” o “pc escritorio” no influyen en el número de materias aprobadas. Lo que contradice indirectamente al estudio realizado por Tinsaray & Karina (2016) el cuál encontró como factor estadísticamente influyente a la participación en actividades en línea como: chat, foro y video-colaboración.

En el algoritmo de CHAID, otra de las variables significativamente influyente fue la calificación de matemáticas en el proceso de Admisiones. Se obtuvo de manera similar este resultado en el estudio realizado por Planck Barahona (2014).

Las variables con respecto a los estudios secundarios no resultaron significativamente influyentes en el modelo. Sin embargo, dentro de uno de los perfiles de los estudiantes que aprueban todas las materias que se formaron mediante el algoritmo de CHAID, una variable significativa resultó los que estudiaron en un colegio ubicado en la Ciudad de Guayaquil. Estudios que obtuvieron que las variables con respecto a los estudios secundarios fueron

significativas fueron Ibarra & Michalus (2010), Rodríguez & Ruiz (2011), McArdle, Paskus & Boker (2013) y Meyer & Marx (2014).

Y otra variable que resultó ser influyentemente significativa dentro del modelo de regresión fue el índice del grado de dificultad de las materias sociales, a medida que este índice aumente, el número de materias aprobadas disminuye. En cambio, el índice del grado de dificultad de las materias STEM resultó significativo en el algoritmo de CHAID para los estudiantes que aprueban todas sus materias, pero obtuvieron una baja calificación de Física en el proceso de Admisiones. En el estudio realizado por Meyer & Marx (2014) encontró que el plan de estudio del estudiante influye en el rendimiento del estudiante. Es decir, que tiene importancia las materias que toma el estudiante durante el término académico.

## **CAPÍTULO 4**

### **CONCLUSIONES Y RECOMENDACIONES**

#### **4.1. Conclusiones**

Este estudio permitió identificar cuáles son los factores más afectan al rendimiento académico del estudiante son el número de materias tomadas, el índice de dificultad de las materias sociales, la calificación obtenida de Física en el proceso de Admisiones y la edad.

Se pudo evidenciar que el factor que principalmente influye de manera favorable para que el estudiante tenga un término académico exitoso es la calificación obtenida de Física en el proceso de Admisiones. Es decir, mientras mayor sea la calificación entonces mayor es su probabilidad de aprobar todas las materias tomadas.

Se pudo evidenciar que a pesar de que el estudiante así haya obtenido una baja calificación de Física en el proceso de admisiones, aún tiene una alta probabilidad de aprobar (74.5%) todas las materias si el índice de dificultad de las materias STEM tomadas es alto ( $\geq 0.258$ ) y tienen edad menor o igual a 18 años.

Se obtuvieron dos modelos que permiten pronosticar el rendimiento académico: uno que indica las características o factores que elevan la probabilidad de que el estudiante apruebe todas las materias que tomó a través del logaritmo de CHAID; y otro que ayuda a pronosticar el número de materias que aprobaría el estudiante.

## **4.2. Recomendaciones**

Revisar de manera general el formulario actualmente utilizado en la encuesta socioeconómica aplicado a los estudiantes, con la finalidad de incluir variables aún no consideradas y obtener información acerca de factores que podrían favorecer o no en el rendimiento académico de los mismo, de esta manera también se contribuiría a la creación de nuevas estrategias institucionales para mejorar este indicador.

Revisar la sección de estudios secundarios de la encuesta socioeconómica, se podría incluir variables como el promedio de graduación y en caso de ser posible la calificación obtenida en el examen Ser bachiller, el tipo de bachillerato, modalidad de estudio, entre otras.

Agregar más variables que indaguen sobre su desempeño con la tecnología, las redes sociales, etc.

Implementar una sección en la encuesta socioeconómica sobre los hábitos de estudio del estudiante; se podría incluir variables como: lugar de estudio en la casa, horario, número de personas con quienes estudia, nivel de concentración, otras actividades académicas o no adicionales a sus estudios, etc.

Analizar la posibilidad de levantar un estudio para levantar información de los estudiantes de ESPOL sobre el nivel de lectura, riesgo de uso de sustancias psicoactivas, nivel del pensamiento lógico abstracto, factores de riesgo de salud física y mental, entre otros como en el estudio realizado por Carvajal, Mosquera & Artamonova (2009).

Capacitar a los profesores para que puedan detectar a estudiantes con problemas académicos, psicológico o de salud que afecten al rendimiento del término académico.

Realizar un estudio más detallado sobre los valores atípicos presentados en el estudio e identificar factores que favorecen o desfavorecen al rendimiento de estos estudiantes.

## **REFERENCIAS BIBLIOGRÁFICAS**

- [1] Aroca, M. D. L. Á. R., Romero, J. E. B., & Paredes, C. A. (2016). Consejerías Académicas en la ESPOL: evolución e impacto en la comunidad politécnica.
- [2] Paredes, C., Zurita, G. and Bermúdez, J. (2014). CONSEJERÍAS ACADÉMICAS: Instructivo para el Consejero o Consejera. Guayaquil, Ecuador: Vicerrectorado Académico - ESPOL.
- [3] Clinciu, A. I. (2013). Adaptation and stress for the first year university students. *Procedia-Social and Behavioral Sciences*, 78, 718-722.
- [4] Narro Robles, J., & Arredondo Galván, M. (2013). La tutoría: un proceso fundamental en la formación de los estudiantes universitarios. *Perfiles educativos*, 35(141), 132-151.
- [5] Martínez-López, Z., Fernández, M. F. P., Couñago, M. A. G., Vacas, C. T., da Silva Almeida, L., & González, M. S. R. (2014). Apoyo social en universitarios españoles de primer año: propiedades psicométricas del Social Support Questionnaire-Short Form y el Social Provisions Scale. *Revista Latinoamericana de Psicología*, 46(2), 102-110.
- [6] Elias, H., Ping, W. S., & Abdullah, M. C. (2011). Stress and academic achievement among undergraduate students in Universiti Putra Malaysia. *Procedia-Social and Behavioral Sciences*, 29, 646-655.
- [7] Rodríguez Ayán, M. N., & Ruiz Díaz, M. Á. (2011). Indicadores de rendimiento de estudiantes universitarios: calificaciones versus créditos acumulados. *Revista de educación*.
- [8] Ramirez, J., & Pelaez, J. C. (2001). Análisis multivariante para medir rendimiento académico de los estudiantes de una carrera universitaria: caso ingeniería en estadística informática, período 1995-2000 (Bachelor's thesis).

- [9] Mendoza, A. A. M., Cadavid, D. A. V., & Herrera, T. J. F. (2013, August). PROPUESTA PARA LA MEDICIÓN DEL RENDIMIENTO ACADÉMICO DE LOS ESTUDIANTES DE LAS UNIVERSIDADES UTILIZANDO ANALISIS ENVOLVENTE DE DATOS (DEA). In WEEF 2013 Cartagena.
- [10] Arancibia, V. (1997). Los sistemas de medición y evaluación de la calidad de la educación. OREALC-UNESCO. Santiago de Chile.
- [11] Tinisaray, G., & Karina, D. (2016). Construcción de un modelo para determinar el rendimiento académico de los estudiantes basado en learning analytics (análisis del aprendizaje), mediante el uso de técnicas multivariantes.
- [12] UNIVERSIDAD NACIONAL DE COLOMBIA (2013). INDICADORES DE RENDIMIENTO ACADÉMICO SEMESTRE 2013-1. Medellín, Colombia.
- [13] McArdle, J. J., Paskus, T. S., & Boker, S. M. (2013). A multilevel multivariate analysis of academic performances in college based on NCAA student-athletes. *Multivariate behavioral research*, 48(1), 57-95.
- [14] Campbell, T. A., & Campbell, D. E. (1997) Faculty/student mentor program: Effects on academic performance and retention. *Research in higher education*, 38(6), 727-742.
- [15] Bermúdez, J. M. V. (2017). Estudio multivariante del síndrome del desgaste profesional y sus implicaciones en el desempeño de docentes en una universidad del Ecuador.
- [16] Carvajal Olaya, P., Mosquera, J. C., & Artamonova, I. (2009). Modelos de predicción del rendimiento académico en Matemáticas I en la Universidad Tecnológica de Pereira. *Scientia et technica*, 15(43).
- [17] Mercado, R. M. E. (1998). Las aplicaciones del análisis de segmentación: El procedimiento Chaid. *Empiria. Revista de metodología de ciencias sociales*, (1), 13-49.
- [18] Barbera, C. G., Niebla, J. C., López, K. D., & Ortega, M. L. (2012). Rendimiento académico y factores asociados. Aportaciones de algunas evaluaciones a gran escala. *Bordón*, 64(2), 51-68.
- [19] Casaravilla, A., DEL CAMPO, J. M., García, A., & TORRALBA, R. M. (2016). Análisis del abandono en estudios de ingeniería y arquitectura en la Universidad Politécnica de Madrid. In Congresos CLABES.

- [20] Castillo, A., & Paz, C. (2016). Superación académica en primer año de Ingeniería y Ciencias: mecanismos de permanencia y mejoramiento académico.
- [21] Cornejo, C. O., & San Martín, N. I. (2013). Estilos de aprendizaje y rendimiento académico en estudiantes de pedagogía de educación general básica (primaria) de una Universidad pública de Chile. *Journal of Learning Styles*, 6(11).
- [22] dDe Fanelli, A. M. G. (2014). Rendimiento académico y abandono universitario: Modelos, resultados y alcances de la producción académica en la Argentina. *Revista Argentina de Educación Superior*,(8), 9-38.
- [23] Edel Navarro, R. (2003). El rendimiento académico: concepto, investigación y desarrollo. *REICE: Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*.
- [24] Erazo-Santander, O. (2012). El rendimiento académico, un fenómeno de múltiples relaciones y complejidades. *Revista Vanguardia Psicológica Clínica Teórica y Práctica*, 2(2), 144-173.
- [25] Garbanzo Vargas, G. M. (2013). Factores asociados al rendimiento académico en estudiantes universitarios desde el nivel socioeconómico: Un estudio en la Universidad de Costa Rica. *Revista Electrónica Educare*, 17(3).
- [26] Garrido Macías, M., Jiménez Luque, N., Landa Sánchez, A., Páez Espinar, E., & Ruiz Barranco, M. (2013). Factores que influyen en el rendimiento académico: la motivación como papel mediador en las estrategias de aprendizaje y clima escolar.
- [27] González Cardona, J. C., & Hernández Cáceres, J. (2013). Sistema de apoyo para la acreditación de la calidad de programas académicos de la Universidad de Caldas, aplicando técnicas en minería de datos.
- [28] Iglesias, L., & Calmet, V. V. (2015). Factores psicológicos, sociales y demográficos asociados al rendimiento académico en estudiantes universitarios. *Revista de Psicología*, 12(1), 216-236.
- [29] Maris Vázquez, S., Noriega Biggio, M., & Maris García, S. (2013). Relaciones entre rendimiento académico, competencia espacial, estilos de aprendizaje y deserción. *Revista electrónica de investigación educativa*, 15(1), 29-44.

- [30] Meyer, M., & Marx, S. (2014). Engineering dropouts: A qualitative examination of why undergraduates leave engineering. *Journal of Engineering Education*, 103(4), 525-548.
- [31] Míguez, M. (2008). Análisis de las relaciones entre proceso motivacional, estrategias de aprendizaje y rendimiento académico en estudiantes del Área Científico-Tecnológica de la Universidad de la República (Doctoral dissertation, Tesis doctoral inédita).
- [32] Montes, I., & Lerner, J. (2011). Rendimiento académico de los estudiantes de pregrado de la Universidad EAFIT. Colombia: Universidad de EAFIT.
- [33] Moreno, B., Ramos, C., & García, A. S. (2017). Efectes de les estratègies d'ensenyament i aprenentatge universitaris en el rendiment acadèmic. *Revista d'Innovació Docent Universitària*, (9), 39-53.
- [34] Parra, C. M., Mejía, L. F., Valencia, A., Restrepo, G., Usuga, O., & Mendoza, R. (2016, November). Rendimiento académico de los estudiantes de primer semestre de pregrado de la Facultad de Ingeniería de la Universidad de Antioquia: cohorte 2012-2. In *Congresos CLABES*.
- [35] Tomás-Miquel, J. V., Expósito-Langa, M., & Sempere-Castelló, S. (2014). Determinantes del rendimiento académico en los estudiantes de grado. Un estudio en administración y dirección de empresas. *Revista de Investigación educativa*, 32(2), 379-392.

## ANEXOS

### i. Código SPSS: Algoritmo de CHAID.

\* Decision Tree.

```
TREE ap_todas [n] BY n_stem [s] n_social [s] gd_stem [s] gd_social [s] adm_fisica [s]  
adm_matem [s] adm_quim [s] edad [s] ingresos [s] cole_gye [n] tipo_cole [n] est_civil_fam [n]  
est_socioec [n] genero [n] grupo_habita [n] ninst_jefeh [n] forma_pago_resid [n] tip_viv_resid [n]  
n_celular [n] n_tvh [n] n_vehich [n] parentesco_jefh [n] tcell [n] tinternet [n] tpc_escr [n] tlaptop [n]  
fam_IESS [n] fam_seg_priv [n] ttv_pagad [n] ttelf_conv [n] etnia [n]
```

```
/TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES  
NODEDEFS=YES SCALE=AUTO  
/DEPCATEGORIES USEVALUES=[VALID]  
/PRINT MODELSUMMARY CLASSIFICATION RISK  
/METHOD TYPE=CHAID  
/GROWTHLIMIT MAXDEPTH=AUTO MINPARENTSIZE=100 MINCHILDSIZE=50  
/VALIDATION TYPE=NONE OUTPUT=BOTHSAMPLES  
/CHAID ALPHASPLIT=0.05 ALPHAMERGE=0.05 SPLITMERGED=NO  
CHISQUARE=PEARSON CONVERGE=0.001 MAXITERATIONS=100 ADJUST=BONFERRONI  
INTERVALS=10  
/COSTS EQUAL  
/MISSING NOMINALMISSING=MISSING.
```

## ii. Código R: Estadística Univariada

### # Cargar librerías

#=====

```
library(Rcmdr)
library(stats)
library(lmtest)
library(tseries)
```

### # Leer datos

#=====

```
datos <- readXL("C:/Users/jodebarz/base_datos.xlsx", rownames=FALSE,
header=TRUE, na="", sheet="datos", stringsAsFactors=TRUE)
```

### # Ordenar Categorías de las variables

#=====

```
datos$estrat_socioec <- with(datos, factor(estrat_socioec, levels=c('Bajo',
'Alto'), ordered=TRUE))
datos$est_socioec <- with(datos, factor(est_socioec, levels=c('Bajo','Medio
bajo','Medio','Medio alto'), ordered=TRUE))
datos$parentesco_jefh <- with(datos, factor(parentesco_jefh,
levels=c('Padre','Madre','Otro'), ordered=TRUE))
datos$pago_resid <- with(datos, factor(pago_resid, levels=c('Propio',
'Alquilado','Otro'), ordered=TRUE))
datos$est_civil_fam <- with(datos, factor(est_civil_fam, levels=c('Sin
pareja','Con pareja'), ordered=TRUE))
datos$tipo_cole <- with(datos, factor(tipo_cole, levels=c('Fiscal', 'Fisco
misional','Particular'), ordered=TRUE))
datos$ninst_jefeh2 <- with(datos, factor(ninst_jefeh2, levels=c('Ninguno',
'Primaria','Secundaria','Tercer nivel','Postgrado'), ordered=TRUE))
datos$ninst_jefeh <- with(datos, factor(ninst_jefeh, levels=c('Ninguno',
'Primaria Incompleta','Primaria
Completa','Secundaria Incompleta',
'Secundaria completa','Tercer nivel
(hasta 3 años)',
'Tercer nivel (4 o más
años)','Postgrado'), ordered=TRUE))
```

### # Variables cualitativas

#=====

# Características generales

```
#-----  
par(mfrow=c(1,1))  
with(datos, Barplot( genero , xlab=NULL, main= "Género",ylab="Porcentaje",  
ylim=c(0,100), scale="percent"))  
summary(datos$genero)  
  
with(datos, Barplot( est_civil , xlab=NULL, main= "Estado Civil",  
ylab="Porcentaje", ylim=c(0,100), scale="percent"))  
summary(datos$est_civil)  
  
with(datos, Barplot( etnia , xlab=NULL, main= "Autoidentificación  
étnica",ylab="Porcentaje", ylim=c(0,100),scale="percent"))  
summary(datos$etnia)  
  
#Nivel Socioeconómico  
#-----  
par(mfrow=c(1,1))  
with(datos, Barplot( est_socioec , xlab=NULL, main= "Estrato  
Socioeconómico",ylab="Porcentaje", ylim=c(0,100), scale="percent"))  
summary(datos$est_socioec)  
  
with(datos, Barplot( n_celular , xlab=NULL, main= "Número de celulares en el  
hogar",ylab="Porcentaje", ylim=c(0,100), scale="percent"))  
summary(datos$n_celular)  
  
with(datos, Barplot( n_tvh , xlab=NULL, main= "Número de televisores en el  
hogar",ylab="Porcentaje", ylim=c(0,100), scale="percent"))  
summary(datos$n_tvh)  
  
with(datos, Barplot( n_vehich , xlab=NULL, main= "Número de vehículos en el  
hogar",ylab="Porcentaje", ylim=c(0,100), scale="percent"))  
summary(datos$n_vehich)  
#Colegio  
#-----  
par(mfrow=c(1,2))  
with(datos, Barplot( tipo_cole , xlab=NULL, main= "Tipo de  
Colegio",ylab="Porcentaje", ylim=c(0,100), scale="percent"))  
summary(datos$tipo_cole)  
  
with(datos, Barplot( cole_gye , xlab=NULL, main= "Colegio ubicado en  
Guayaquil",ylab="Porcentaje", ylim=c(0,100), scale="percent"))  
summary(datos$cole_gye)
```

#### #Vivienda

#-----

```
par(mfrow=c(1,1))
with(datos, Barplot( tipo_pago_resid , xlab=NULL, main= "Forma de pago de la
vivienda",ylab="Porcentaje", ylim=c(0,100), scale="percent"))
summary(datos$tipo_pago_resid)
```

```
with(datos, Barplot( tip_viv_resid , xlab=NULL, main= "Tipo de
vivienda",ylab="Porcentaje", ylim=c(0,100), scale="percent"))
summary(datos$tip_viv_resid)
```

```
with(datos, Barplot( grupo_habita , xlab=NULL, main= "Grupo que
reside",ylab="Porcentaje", ylim=c(0,100), scale="percent"))
summary(datos$grupo_habita)
```

#### #Características del jefe del hogar

#-----

```
par(mfrow=c(1,1))
with(datos, Barplot( parentesco_jefh , xlab=NULL, main= "Parentesco con el
Jefe de hogar",ylab="Porcentaje", ylim=c(0,100), scale="percent"))
summary(datos$parentesco_jefh)
```

```
with(datos, Barplot( est_civil_fa , xlab=NULL, main= "Estado Civil
Familiar",ylab="Porcentaje", ylim=c(0,100), scale="percent"))
summary(datos$est_civil_fa)
```

```
with(datos, Barplot( ninst_jefeh2 , xlab=NULL, main= "Nivel de Instrucción del
Jefe de Hogar",ylab="Porcentaje", ylim=c(0,100), scale="percent"))
summary(datos$ninst_jefeh2)
```

#### #Tecnologías de Información y Comunicación (TICs)

#-----

```
par(mfrow=c(3,2))
with(datos, Barplot( tcell , xlab=NULL, main= "Tiene celular",ylab="Porcentaje",
ylim=c(0,100), scale="percent"))
summary(datos$tcell)
```

```
with(datos, Barplot( ttelf_conv , xlab=NULL, main= "Tiene teléfono
convencional",ylab="Porcentaje", ylim=c(0,100), scale="percent"))
summary(datos$ttelf_conv)
```

```
with(datos, Barplot( tinternet , xlab=NULL, main= "Tiene Internet en el  
hogar",ylab="Porcentaje", ylim=c(0,100), scale="percent"))  
summary(datos$tinternet)
```

```
with(datos, Barplot( ttv_pagad , xlab=NULL, main= "Tiene servicio de Tv.  
pagada",ylab="Porcentaje", ylim=c(0,100), scale="percent"))  
summary(datos$n_vehich)
```

```
with(datos, Barplot( tpc_escr , xlab=NULL, main= "Tiene PC  
escritorio",ylab="Porcentaje", ylim=c(0,100), scale="percent"))  
summary(datos$tpc_escr)
```

```
with(datos, Barplot( tlaptop , xlab=NULL, main= "Tiene  
laptop",ylab="Porcentaje", ylim=c(0,100), scale="percent"))  
summary(datos$tlaptop)
```

#### #Salud

```
#-----
```

```
par(mfrow=c(1,1))  
with(datos, Barplot( fam_IESS , xlab=NULL, main= "Familiares afiliados al  
IESS",ylab="Porcentaje", ylim=c(0,100), scale="percent"))  
summary(datos$fam_IESS)
```

```
with(datos, Barplot( fam_seg_priv , xlab=NULL, main= "Familiares afiliados a un  
Seguro Privado",ylab="Porcentaje", ylim=c(0,100), scale="percent"))  
summary(datos$fam_seg_priv)
```

#### # Variables cuantitativas: Histograma, Boxplot, QQ-norm

```
#=====
```

#### #Número de materias tomadas

```
#-----
```

```
par(mfrow=c(1,1))  
par(mfrow=c(1,3))  
hist(datos$n_tom , col = "lightgrey", main = "Histograma",  
xlab=NULL,ylab="Frecuencia")  
boxplot(datos$n_tom , col = "lightgrey", main = "Diagrama de Caja")  
qqnorm(datos$n_tom)  
qqline(datos$n_tom, col="red",lwd=2)
```

#### #Número de materias aprobadas

```
#-----
```

```
par(mfrow=c(1,1))
par(mfrow=c(1,3))
hist(datos$n_ap , col = "lightgrey", main = "Histograma",
xlab=NULL,ylab="Frecuencia")
boxplot(datos$n_ap , col = "lightgrey", main = "Diagrama de Caja")
qqnorm(datos$n_ap)
qqline(datos$n_ap, col="red",lwd=2)
```

*#Tasa de materias aprobadas*

*#-----*

```
par(mfrow=c(1,1))
par(mfrow=c(1,3))
hist(datos$tasa_ap , col = "lightgrey", main = "Histograma",
xlab=NULL,ylab="Frecuencia")
boxplot(datos$tasa_ap , col = "lightgrey", main = "Diagrama de Caja")
qqnorm(datos$tasa_ap)
qqline(datos$tasa_ap, col="red",lwd=2)
```

*#Grado de dificultad stem*

*#-----*

```
par(mfrow=c(1,1))
par(mfrow=c(1,3))
hist(datos$gd_stem , col = "lightgrey", main = "Histograma",
xlab=NULL,ylab="Frecuencia")
boxplot(datos$gd_stem , col = "lightgrey", main = "Diagrama de Caja")
qqnorm(datos$gd_stem)
qqline(datos$gd_stem, col="red",lwd=2)
```

*#Grado de dificultad social*

*#-----*

```
par(mfrow=c(1,1))
par(mfrow=c(1,3))
hist(datos$gd_social , col = "lightgrey", main = "Histograma",
xlab=NULL,ylab="Frecuencia")
boxplot(datos$gd_social , col = "lightgrey", main = "Diagrama de Caja")
qqnorm(datos$gd_social)
qqline(datos$gd_social, col="red",lwd=2)
```

*#Admisiones: Física*

*#-----*

```
par(mfrow=c(1,1))
par(mfrow=c(1,3))
```

```
hist(datos$adm_fisica , col = "lightgrey", main ="Histograma",  
xlab=NULL,ylab="Frecuencia")  
boxplot(datos$adm_fisica , col = "lightgrey", main = "Diagrama de Caja")  
qqnorm(datos$adm_fisica)  
qqline(datos$adm_fisica, col="red",lwd=2)
```

#### #Admisiones: Matemáticas

```
#-----  
par(mfrow=c(1,1))  
par(mfrow=c(1,3))  
hist(datos$adm_matem , col = "lightgrey", main ="Histograma",  
xlab=NULL,ylab="Frecuencia")  
boxplot(datos$adm_matem , col = "lightgrey", main = "Diagrama de Caja")  
qqnorm(datos$adm_matem)  
qqline(datos$adm_matem, col="red",lwd=2)
```

#### #Admisiones: Química

```
#-----  
par(mfrow=c(1,1))  
par(mfrow=c(1,3))  
hist(datos$adm_quim , col = "lightgrey", main ="Histograma",  
xlab=NULL,ylab="Frecuencia")  
boxplot(datos$adm_quim , col = "lightgrey", main = "Diagrama de Caja")  
qqnorm(datos$adm_quim)  
qqline(datos$adm_quim, col="red",lwd=2)
```

#### #Edad

```
#-----  
par(mfrow=c(1,1))  
par(mfrow=c(1,3))  
hist(datos$edad , col = "lightgrey", main ="Histograma",  
xlab=NULL,ylab="Frecuencia")  
boxplot(datos$edad , col = "lightgrey", main = "Diagrama de Caja")  
qqnorm(datos$edad)  
qqline(datos$edad, col="red",lwd=2)
```

#### #Ingresos

```
#-----  
par(mfrow=c(1,1))  
par(mfrow=c(1,3))  
hist(datos$ingresos , col = "lightgrey", main ="Histograma",  
xlab=NULL,ylab="Frecuencia")
```

```
boxplot(datos$ingresos , col = "lightgrey", main = "Diagrama de Caja")  
qqnorm(datos$ingresos)  
qqline(datos$ingresos, col="red",lwd=2)
```

### iii. Código R: Estadística Bivariada

```
# Características generales
```

```
#-----
```

```
Boxplot(n_ap ~ género ,data=datos, id=list(method="none"), xlab=NULL,  
ylab="número de materias aprobadas" ,main=" género ",col = "lightgray")
```

```
#Nivel Socioeconómico
```

```
#-----
```

```
par(mfrow=c(1,1))
```

```
Boxplot(n_ap ~ estrat_socioec ,data=datos, id=list(method="none"),  
xlab=NULL, ylab="número de materias aprobadas" ,main=" est_socioec ",col =  
"lightgray")
```

```
Boxplot(n_ap ~ n_cell ,data=datos, id=list(method="none"), xlab=NULL,  
ylab="número de materias aprobadas" ,main=" n_celular ",col = "lightgray")
```

```
Boxplot(n_ap ~ n_tv ,data=datos, id=list(method="none"), xlab=NULL,  
ylab="número de materias aprobadas" ,main=" n_tvh ",col = "lightgray")
```

```
Boxplot(n_ap ~ n_vehic ,data=datos, id=list(method="none"), xlab=NULL,  
ylab="número de materias aprobadas" ,main=" n_vehic ",col = "lightgray")
```

```
#Colegio
```

```
#-----
```

```
par(mfrow=c(1,1))
```

```
Boxplot(n_ap ~ cole_gye ,data=datos, id=list(method="none"), xlab=NULL,  
ylab="número de materias aprobadas" ,main=" cole_gye ",col = "lightgray")
```

```
Boxplot(n_ap ~ tipo_cole ,data=datos, id=list(method="none"), xlab=NULL,  
ylab="número de materias aprobadas" ,main=" tipo_cole ",col = "lightgray")
```

```
#Vivienda
```

```
#-----
```

```
par(mfrow=c(1,1))
```

```
Boxplot(n_ap ~ pago_resid ,data=datos, id=list(method="none"), xlab=NULL,  
ylab="número de materias aprobadas" ,main=" pago_resid ",col = "lightgray")
```

```
Boxplot(n_ap ~ tipo_viv_resid ,data=datos, id=list(method="none"), xlab=NULL,  
ylab="número de materias aprobadas" ,main=" tipo_viv_resid ",col = "lightgray")
```

```
Boxplot(n_ap ~ reside_familia ,data=datos, id=list(method="none"), xlab=NULL,  
ylab="número de materias aprobadas" ,main=" reside_familia ",col = "lightgray")
```

#### #Características del jefe del hogar

```
#-----
```

```
par(mfrow=c(1,1))
```

```
Boxplot(n_ap ~ padre_jefeh ,data=datos, id=list(method="none"), xlab=NULL,  
ylab="número de materias aprobadas" ,main=" padre_jefeh ",col = "lightgray")
```

```
Boxplot(n_ap ~ est_civil_fam ,data=datos, id=list(method="none"), xlab=NULL,  
ylab="número de materias aprobadas" ,main=" est_civil_fam ",col = "lightgray")
```

```
Boxplot(n_ap ~ jefeh_tercer_niv ,data=datos, id=list(method="none"),  
xlab=NULL, ylab="número de materias aprobadas" ,main=" jefeh_tercer_niv  
",col = "lightgray")
```

#### #Tecnologías de Información y Comunicación (TICs)

```
#-----
```

```
par(mfrow=c(3,2))
```

```
Boxplot(n_ap ~ tcell ,data=datos, id=list(method="none"), xlab=NULL,  
ylab="número de materias aprobadas" ,main=" Tiene celular",col = "lightgray")
```

```
Boxplot(n_ap ~ ttelf_conv ,data=datos, id=list(method="none"), xlab=NULL,  
ylab="número de materias aprobadas" ,main=" Tiene teléfono convencional",col  
= "lightgray")
```

```
Boxplot(n_ap ~ tinternet ,data=datos, id=list(method="none"), xlab=NULL,  
ylab="número de materias aprobadas" ,main=" Tiene Internet en el hogar", col  
= "lightgray")
```

```
Boxplot(n_ap ~ ttv_pagad ,data=datos, id=list(method="none"), xlab=NULL,  
ylab="número de materias aprobadas" ,main=" Tiene servicio de Tv.  
pagada",col = "lightgray")
```

```
Boxplot(n_ap ~ tpc_escr ,data=datos, id=list(method="none"), xlab=NULL,  
ylab="número de materias aprobadas" ,main=" Tiene PC escritorio",col =  
"lightgray")
```

```
Boxplot(n_ap ~ tlaptop ,data=datos, id=list(method="none"), xlab=NULL,  
ylab="número de materias aprobadas" ,main=" Tiene laptop", col = "lightgray")
```

#### #Salud

```
#-----
```

```
par(mfrow=c(1,1))
```

```
Boxplot(n_ap ~ fam_IESS ,data=datos, id=list(method="none"), xlab=NULL,  
ylab="número de materias aprobadas" ,main=" fam_IESS ",col = "lightgray")
```

```
Boxplot(n_ap ~ fam_seg_priv ,data=datos, id=list(method="none"), xlab=NULL,  
ylab="número de materias aprobadas" ,main=" fam_seg_priv ",col = "lightgray")
```

#### iv. Código R: Matriz de Correlación

```
# Leer datos
#=====
cuantitativas=datos[,8:16]

# Histograma múltiple
#-----
library(psych)
multi.hist(x = cuantitativas, dcol = c("blue", "red"), dlty = c("dotted", "solid"))

# Gráfica: Matriz de Dispersión, Histograma y Correlación
#-----
pairs.panels(cuantitativas, pch=21,main=NULL)

# Matriz de Correlación y valor p
#-----
library(Hmisc)
rcorr(as.matrix(cuantitativas))

# Corrplot
#-----
library(corrplot)
corrplot::corrplot.mixed(cor(cuantitativas), lower="ellipse", upper="number")
corrplot(round(cor(cuantitativas), digits = 3), type = "lower")
```

#### v. Código R: Modelo Regresión Lineal Multivariada

```
# Cálculo de la muestra
#=====
ix<-sample(x = nrow(datos),size=round(0.7*nrow(datos)),replace=FALSE)
TRAIN<-datos[ix,]
TEST<-datos[-ix,]

# Modelo de Regresión Lineal Multivariada
#=====
modelo <- lm(n_ap ~ n_tom + gd_social + adm_fisica + adm_matem +
adm_quim + género + edad + est_civil_fam + padre_jefeh + jefeh_tercer_niv +
```

```
ingresos + strat_socioec + n_cell + tipo_viv_resid + tcell + ttelf_conv +  
tinternet + ttv_pagad + cole_gye + tipo_cole,data=TRAIN)
```

```
summary(modelo)  
anova(modelo)
```

```
# STEP modelo
```

```
#=====
```

```
step(modelo)  
stepwise(modelo,direction = "backward","AIC")  
stepwise(modelo,direction = "backward","BIC")
```

```
# Supuestos de un modelo lineal:
```

```
#=====
```

```
# 1. Normalidad, los residuos provienen de una normal
```

```
residuo=residuals(modelo)  
shapiro.test(residuo)
```

```
library(leaps, pos=18)  
oldpar <- par(oma=c(0,0,3,0), mfrow=c(2,2))  
plot(modelo)  
par(oldpar)  
par(mfrow=c(1,1))  
qqPlot(modelo, simulate=TRUE, id=list(method="y", n=2))
```

```
# 2. Homocedasticidad
```

```
bptest(modelo)
```

```
# 3. Residuos no están correlacionados
```

```
dwtest(modelo,alternative="two.sided")
```

```
# Autocorrelación de los residuos y sus retardos
```

```
acf(residuo)  
durbinWatsonTest(modelo)
```

```
#4. Outliers en el modelo
```

```
outlierTest(modelo)  
influencePlot(modelo)  
avPlots(modelo, id=list(method="mahal", n=2))
```

```
#5. Prueba linealidad
```

```
crPlots(modelo, smooth=list(span=0.5))  
ceresPlots(modelo)
```

#### #6. Multicolinealidad

```
#INDICE DE INFLACION DE LA VARIANZA  
#Si alguno es >10, hay problema de multicolinealidad  
vif(modelo) # factores de inflación de la varianza
```

#### # Validación del modelo:

```
#=====
```

```
yhat <- predict(modelo, TEST)
```

```
par(mfrow=c(2,2))  
plot(TEST$n_ap,yhat,main="Observados vs. predichos")  
residuos<-TEST$n_ap-yhat
```

```
plot(residuos, main = "residuos vs. Index")  
hist(residuos,nclass=40,main="Histograma de residuos")
```

```
qqnorm(residuos)  
qqline(residuos, col="red", lwd=2)
```

```
RSS=sum(residuos**2)  
RSS
```

#### ## validación cruzada

```
library(DAAG)  
par(mfrow=c(1,1))  
cv.lm(datos, modelo, m = 5)
```