

**ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL**



**FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICAS**

**DEPARTAMENTO DE MATEMÁTICAS**

**PROYECTO DE TITULACIÓN**

PREVIO A LA OBTENCIÓN DEL TÍTULO DE:

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

**TEMA:**

**ANÁLISIS ESTADÍSTICO DE LAS CAUSAS QUE INFLUYEN EN EL  
RENDIMIENTO ACADÉMICO EN MATEMÁTICAS EN UN  
INSTITUTO SUPERIOR DE GUAYAQUIL**

**AUTOR:**

**DARWIN OMAR ESPÍN MONTESDEOCA**

Guayaquil - Ecuador

2019

## RESUMEN

Los estudiantes que ingresan a institutos tecnológicos por lo general llegan con un nivel bajo en el conocimiento de Matemáticas, lo cual tiene consecuencias negativas en el estudiantado, siendo esta asignatura fundamental para su desarrollo profesional, ya que permite complementar la teoría y la práctica. El objetivo de este estudio fue la construcción de un modelo utilizando la regresión lineal múltiple con factores cualitativos y cuantitativos. La variable dependiente fue la puntuación obtenida durante el examen final de un tecnológico y diecinueve variables explicativas. Para analizar las variables cualitativas se hizo uso de varias variables dummy. Para determinar el modelo predictor óptimo se utilizaron los criterios de significación backward, forward y el método stepwise mixto AIC. El modelo obtenido por el método stepwise mixto AIC dio un p-value de  $2.39 \times 10^{-7}$ ,  $R^2$  de 0.3996, SER de 2.309 y F-statics de 9.982, el análisis de residuo no muestra colinealidad ni heterocedasticidad, es decir que las variables predictoras se relacionan significativamente con la variable dependiente puntuación. Las variables predictoras fueron la edad, tiempo de dedicación al estudio, estado civil del estudiante, nivel de estudio del padre y situación laboral del padre. En base a estos resultados y al análisis de causa-efecto mediante espigas de pescado se elaboró un perfil del estudiante y una propuesta de mejora.

**Palabras claves:** stepwise, backward, forward, dummy, AIC

## ABSTRACT

Students who enter technological institutes usually arrive with a low level of knowledge of Mathematics, which has negative consequences in the student body, this being a fundamental subject for their professional development, since it allows to complement the theory and practice. The objective of this study was the construction of a predictive model using multiple linear regression with qualitative and quantitative factors. The dependent variable was the score obtained during the final examination of a technological and nineteen explanatory variables. To analyze the qualitative variables, several dummy variables were used. To determine the optimal predictor model, we used the criteria of significance backward, forward and mixed stepwise method AIC. The model obtained by the mixed stepwise method AIC gave a p-value of  $2.39 \times 10^{-7}$ ,  $R^2$  of 0.3996, SER of 2,309 and F-statics of 9,982, the residue analysis does not show collinearity or homoscedasticity, that is, the predictor variables they are significantly related to the dependent variable score. The predictor variables were age, time of study dedication, student's marital status, father's study level and father's work situation. Based on these results and the analysis of cause and effect using fish bones, a profile of the student and an improvement proposal were elaborated.

**Keywords:** stepwise, backward, forward, dummy, AIC

## **DEDICATORIA**

Este trabajo lo dedico principalmente a Dios todo poderoso, quien me ha provisto de fuerzas para obtener uno de los anhelos más deseados.

A mi familia, mi esposa e hijos quienes han sido el motor y fuente de inspiración que me han llevado a sacar adelante este proyecto y han estado conmigo en momentos difíciles, superando uno a uno los obstáculos presentados.

A mis padres, quienes han sabido guiarme y enseñarme el valor de las cosas, gracias a ustedes he logrado llegar hasta aquí y alcanzar una meta propuesta. Son los mejores padres.

A mi hermano por estar siempre presente acompañándome y por el apoyo moral que me ha brindado a lo largo de esta etapa de mi vida.

## **AGRADECIMIENTO**

Agradezco a Dios por su bendición y guiarme a lo largo de este tiempo de estudio y darme fuerzas en momentos de dificultad y debilidad.

Agradezco a mi esposa quien siempre estuvo a mi lado apoyándome, sacrificando tiempo y alentándome en momentos difíciles, gracias por eso.

Quiero agradecer también a los docentes de este Postgrado quienes aportaron sus conocimientos y vasta experiencia de forma muy profesional, cumpliendo las expectativas esperadas.

A todas las personas y amigos que me han apoyado y han hecho que el trabajo se realice con éxito, en especial aquellos que me abrieron las puertas y me dieron directrices compartiendo sus conocimientos.

## DECLARACIÓN EXPRESA

La responsabilidad por los hechos y doctrinas expuestas en este Proyecto de Titulación, me corresponde exclusivamente y ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría. El patrimonio intelectual del mismo, corresponde exclusivamente a la ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.



---

Nombre del Autor

# TRIBUNAL DE GRADUACIÓN



---

MARIA NELA PASTIUZACA FERNANDEZ

PRESIDENTE



---

GARCIA BUSTOS SANDRA LORENA

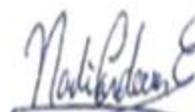
DIRECTOR



---

RUIZ BARZOLA OMAR HONORIO

VOCAL 1



---

CARDENAS ESCOBAR NADIA LORENA

VOCAL 2

## ABREVIATURAS O SIGLAS

**df:** Grados de Libertad

**EGB:** Educación General Básica

**ENES:** Examen Nacional de Educación Superior

**E.S.O:** Educación Secundaria Obligatoria

**MLR:** Regresión Lineal Múltiple.

**MSE:** Error Cuadrático Medio.

**SVM:** Máquina de Vectores de Soporte.

**INEVAL:** Instituto Nacional de Evaluación Educativa

**ITS:** Instituto Tecnológico Superior

**V.I:** Variable Independiente

# TABLA DE CONTENIDO

CAPÍTULO 1 .....	1
1. INTRODUCCIÓN .....	1
1.1 Antecedentes.....	1
1.2. Descripción del problema.....	2
1.3. Justificación .....	2
1.4. Objetivos .....	3
1.5. Metodología.....	3
1.6. Alcance .....	4
1.7. Revisión bibliográfica .....	4
CAPÍTULO 2 .....	7
2. MARCO TEÓRICO .....	7
2.2. Estimación de los Coeficientes del Modelo .....	8
2.3. Tabla de Análisis de Varianza .....	9
2.4. Pruebas de hipótesis para los parámetros del modelo .....	12
2.5. R cuadrado .....	13
2.6. R <sup>2</sup> ajustado .....	13
2.7. Selección de variables.....	14
2.8. Método Backward.....	14
2.9. Método Forward.....	14
2.10. Herramientas de calidad para el análisis de problemas.....	14
CAPÍTULO 3 .....	18
3. RESULTADOS .....	18
3.1. Recolección de datos.....	18
3.2. Variables .....	18
3.3. Análisis descriptivos de las variables .....	20
3.4. Análisis del rendimiento académico de los estudiantes a través de un modelo de Regresión Lineal Múltiple.....	28
3.5. Análisis de causa-efecto para la mejora del rendimiento académico .....	33
CAPÍTULO 4 .....	35
4.1 DISCUSIÓN.....	35
CAPÍTULO 5 .....	37
5. CONCLUSIONES Y RECOMENDACIONES .....	37
5.1. Conclusiones.....	37

5.2. Recomendaciones..... 40  
BIBLIOGRAFÍA..... 42

## LISTADO DE FIGURAS

Figura 2.1. Pasos para la elaboración del Diagrama de Ishikawa (Fundibeq, 2012). .....	16
Figura 3.1. Boxplot del sexo del estudiante con respecto a la puntuación.....	20
Figura 3.2. Boxplot del estado civil del estudiante con respecto a la puntuación..	21
Figura 3.3. Boxplot lugar de residencia del estudiante con respecto a la puntuación.....	21
Figura 3.4. Boxplot tipo de colegio del estudiante con respecto a la puntuación.....	22
Figura 3.5. Boxplot especialidad/bachillerato del estudiante con respecto a la puntuación.....	23
Figura 3.6. Boxplot situación laboral del estudiante con respecto a la puntuación... .....	23
Figura 3.7. Boxplot asistencia a clases con respecto a la puntuación.....	24
Figura 3.8. Boxplot tiempo sin estudiar con respecto a la puntuación.....	24
Figura 3.9. Boxplot estado civil de los padres con respecto a la puntuación .....	25
Figura 3.10. Boxplot nivel académico del padre con respecto a la puntuación.....	26
Figura 3.11. Boxplot situación laboral del padre con respecto a la puntuación ...	26
Figura 3.12. Boxplot deseo de seguir posgrado con respecto a la puntuación.....	27
Figura 3.13. Boxplot uso de TICs con respecto a la puntuación .....	27
Figura 3.14. Boxplot Matriz de diagrama de dispersión múltiple.....	28
Figura 3.15. Selección de variables predictoras de acuerdo al método (BIC y adjr2).....	30
Figura 3.16. R <sup>2</sup> -ajustado vs número de predictores.....	30

Figura 3.17. Análisis de los residuos.....	32
Figura 3.18. Análisis de los residuos estudentizados.....	33
Figura 3.19. Diagrama de causa-efecto para mejora de rendimiento académico en tecnológico .....	34

## LISTADO DE TABLAS

Tabla 2.1. Formato de tabla ANOVA.....	10
Tabla 3.1. Coeficiente para un modelo con todos los predictores.....	29
Tabla A-1. Coeficientes de correlación entre las variables.....	46
Tabla A-2. Fórmula del modelo con todas las variables independientes.....	47
Tabla A-3. Método AIC para determinación de factores predictores.....	48
Tabla A-4. Coeficientes del modelo optimizado por AIC.....	49
Tabla A-5. Resumen de datos de las variables.....	49

# CAPÍTULO 1

## 1. INTRODUCCIÓN

### 1.1 Antecedentes

Es muy común en nuestro medio que asignaturas relacionadas con los números, como Matemáticas, Física, Contabilidad, presenten un mayor grado de dificultad para los estudiantes, ya sean estudiantes de educación media o centros de educación superior.

Si un estudiante presenta bajo rendimiento académico, este problema puede depender de muchos factores, por ejemplo, si el estudiante es el primero en su generación familiar que estudia, si en la familia existe algún problema emocional, o los padres no tienen trabajo, en ciertos casos un nivel socioeconómico bajo, entre otros.

Este problema no solo se presenta en nuestro medio, es un problema a nivel mundial, así lo menciona Osorio Sánchez en un foro académico realizado en España: El problema de aprender y enseñar Matemáticas (y en general Ciencias Básicas), ha sido objeto de investigación sistemática e institucional en los últimos años a nivel mundial. Las investigaciones han encontrado que son diversos los factores que inciden y derivado de ello se han planteado acciones para tratar de resolver tal problemática. (Osorio Sánchez, 2011)

En el 2013 se evidenció por medio de los resultados de exámenes del Instituto Nacional de Evaluación Educativa (INEVAL), que Matemáticas es la asignatura en la que mayores problemas presentan los estudiantes, así lo publica en un titular El Universo, que dice: “Matemáticas tiene el mayor déficit en las evaluaciones estudiantiles del 2013 en Ecuador”, destacando que “los niveles de desempeño en décimo de Educación General Básica (EGB) señalan que el 42,8% tiene el grado de insuficiente y el 45,9% alcanza el nivel elemental en Matemáticas, mientras que el 2,4% alcanza un promedio de excelente en esta asignatura.” (El Universo, 2014)

Como podemos darnos cuenta, la complejidad en el aprendizaje de las Matemáticas empieza desde la Educación General Básica, y este problema se va arrastrando hasta llegar a las instituciones de educación superior, donde los estudiantes se ven afectados drásticamente en el momento que luchan por un cupo para ingresar a la universidad.

## **1.2. Descripción del problema**

Llama la atención en los directivos de un instituto tecnológico de Guayaquil, que los estudiantes que ingresan llegan con un nivel muy bajo en el conocimiento de Matemáticas, teniendo consecuencias negativas en el estudiantado, ya que por ser una carrera tecnológica de educación superior la que ellos siguen, necesitan el uso del cálculo como base para complementar la parte teórica con la práctica. Desde que se formó el tecnológico, ya más de 10 años, esta situación se repite continuamente en cada semestre, generando también una deserción considerable en los estudiantes que ingresan.

La apreciación que se ha podido percibir en los alumnos que se deciden a estudiar en un tecnológico es que quieren algo práctico, no quieren algo que demande mucho estudio, por eso muchos no prestan atención a asignaturas de ciencia, como Matemáticas; pero en realidad si no dominan estos conocimientos, no van a poder ser eficientes en sus trabajos.

Dado estos antecedentes, los directivos de la institución consideran necesario analizar el rendimiento académico en Matemáticas de los estudiantes que ingresan al tecnológico, como un medio para tratar de determinar las posibles causas y proponer acciones de mejora.

## **1.3. Justificación**

Al realizar esta investigación es importante determinar el perfil de los estudiantes que ingresan a un tecnológico en la ciudad de Guayaquil, con el objetivo de conocer si es que existe alguna tendencia del estudiantado que escoge estudiar en un tecnológico superior, para que la institución identifique que actividades fomentar, para mejorar el desempeño del estudiante. Se necesita determinar factores que influyen en el rendimiento académico en Matemáticas, la cual no está

disponible en los centros educativos, por lo que será de mucha ayuda, para tomar decisiones a nivel general.

Se puede decir que los beneficiarios de este trabajo serán las autoridades de educación del tecnológico, ya que con esta investigación tendrán información que les permita el establecimiento de políticas educativas, así como también tendrán una investigación base que de paso a otras que profundicen más la problemática que existe en la educación a nivel general, y así buscar una solución al rendimiento académico de los estudiantes.

## **1.4. Objetivos**

### **1.4.1. Objetivo General**

Determinar las causas que influyen en el rendimiento académico en Matemáticas de los estudiantes que ingresan a un instituto tecnológico superior de Guayaquil.

### **1.4.2. Objetivos Específicos**

1. Analizar el rendimiento académico de los estudiantes a través de un modelo de Regresión Lineal Múltiple.
2. Elaborar el perfil de los estudiantes que ingresan al Instituto Tecnológico Superior.
3. Elaborar una propuesta de mejora para tratar el bajo rendimiento académico en matemáticas de los estudiantes.

## **1.5. Metodología**

Para efectos de analizar el rendimiento académico en Matemáticas de los estudiantes que ingresan a un instituto tecnológico superior de la ciudad, se empleará un modelo de Regresión Lineal Múltiple con el objetivo de determinar cuáles son las causas/factores que provocan variaciones en este, en el estudio se usarán variables como tipo de colegio, ambiente familiar, hábitos de estudio, edad, sexo etc., Para determinar el modelo óptimo se utilizaran los criterios backward y forward y se comparan los resultados a través del criterio de coeficiente de determinación ajustado y el AIC.

Por otro lado, se elaborará un perfil de los estudiantes del instituto tecnológico, a través del análisis descriptivo de datos y se utilizará técnicas de calidad para identificar causas/efectos y llegar al establecimiento de acciones de mejora en el instituto con respecto al rendimiento de los estudiantes.

### **1.6. Alcance**

El presente estudio analizará el rendimiento académico registrado en las pruebas de diagnóstico de matemáticas administradas a los estudiantes que ingresan a un instituto tecnológico superior de Guayaquil en el año 2017.

### **1.7. Revisión bibliográfica**

Para efectos de realizar el presente trabajo de titulación, se ha procedido a revisar varios artículos relacionados con el objetivo que se persigue en este estudio, entre ellos podemos citar:

(González, C., 2004), quien realizó un estudio cuyo objetivo era encontrar factores determinantes del bajo rendimiento académico en educación secundaria, en la investigación plantea un modelo de regresión lineal múltiple para medir el rendimiento en términos de ciertas variables como ingreso de los padres, nivel de educación de padres, sector de residencia del estudiante, estrategias de aprendizaje apropiadas, motivación orientada a la tarea, entre otras, teniendo como resultado que variables como factores intelectuales (conocimientos previos y aplicación de técnicas de estudio) y autoconcepto (pertenencia y actitud) influyen directamente en el rendimiento académico del estudiante.

De igual manera, los autores M. Pérez, G. Orlandoni, F. Aguilar y J. Ramoni en 2015, realizaron un estudio en el que aplican métodos multivariantes para evaluar los resultados obtenidos por los estudiantes que ingresan a la Universidad de Santander en el período 2015 en las pruebas del Instituto Colombiano para la Evaluación de la Educación (ICFES). Este diagnóstico propone un criterio de clasificación para identificar los estudiantes con altos niveles de riesgo de acuerdo a los puntajes de la prueba de cada área evaluada. Además de los puntajes obtenidos en las diferentes áreas de la prueba, se analizan otras variables como tipo de colegio, sexo, carrera a la que ingresa y región de procedencia, teniendo como

resultados que estudiantes de escuelas privadas están mejor preparados en inglés y factor numérico.

Existe también una publicación realizada en 2015 por los autores D. Martínez, M. Karanik, M. Giovannini y N. Pinto, en la cual, ante la necesidad de definir perfiles de estudiantes y determinar patrones que conducen al éxito o fracaso académico, proponen evaluar el rendimiento de los estudiantes mediante técnicas de minería de datos. La propuesta no se enfoca sólo en analizar el perfil del estudiante a través de sus calificaciones sino también estudiar el desempeño académico en base a otras variables tales como factores socioeconómicos, demográficos, culturales, entre otros, en base a lo cual elaboran diferentes perfiles de alumnos. El estudio se realiza para las asignaturas Algoritmos y Estructuras de Datos, en la carrera de Ingeniería de Sistemas de la Universidad Tecnológica Nacional Facultad Regional Resistencia (UTN-FRR).

Así mismo, (Lozano.,2016) analizó datos obtenidos de 1178 alumnos de cuatro institutos de Almería capital (España), mismos que cursaban la Educación Secundaria Obligatoria (E.S.O). Los datos se obtuvieron a partir de dos instrumentos de medida: adaptación del cuestionario de Test Autoevaluativo Multifactorial de Adaptación Infantil (TAMAI) y una medida del fracaso escolar. Las variables analizadas fueron en total 34 entre las que se encontraban las siguientes: nivel académico de los padres, el género, la motivación y las relaciones sociales en clases. Estas variables fueron sometidas a un análisis de fiabilidad de Alfa de Cronbach, mismo que dio un índice de 0.72 el cual es bastante aceptable. Posteriormente se llevó a cabo análisis factorial exploratorio, obteniéndose como resultado que existe una influencia directa de variables estudiadas.

Así mismo, Rosas *et al.*, 2006 utilizaron modelos de regresión lineal múltiple para predecir el rendimiento estudiantil. La investigación se realizó en el Instituto Universitario del Yacucuy, se utilizaron 28 variables explicativas y como variable dependiente se usó el índice de rendimiento académico al egresar del Instituto. Se realizó un modelo para cuatro especialidades estudiadas: Agrícola, Conservación de Recurso Naturales Renovables (C.R.N.R.), Alimentos y Pecuaria. Estos modelos explicaron respectivamente el 56.41%, 89.66%, 69.33% y el 73.10% de la

variabilidad total del rendimiento y las variables escogidas difirieron de acuerdo a la especialidad.

Gallego y Sierra (2012) realizaron un modelo de gestión educativo para de esta manera disminuir la separación existente entre lo que los estudiantes aprenden en la Facultad y las demandas exigidas en los centros/ instituciones/ empresas donde son acogidos. Las técnicas utilizadas en esta investigación fueron los Círculos de Calidad y los instrumentos de recogida de datos, brainstorming, diagrama de Ishikawa e histogramas para el análisis de los problemas relacionadas con el centro de prácticas y para la implementación de las soluciones propuestas.

Shinde, D. D., Ahirrao, S., & Prasad, R. (2018), utilizaron la técnica de diagrama de espina de pescado o también llamado diagrama de “causa y efecto” para identificar problemas en el sistema de educación en ingeniería en la India, se consideró a los estudiantes y docentes como la parte interesada. En este estudio se determinó que las principales causas para un bajo rendimiento son cuatro: Personales, universidad, recursos y académicos. Estos factores principales tienen sub factores que nos llevan a la causa de estos problemas. Los antecedentes familiares y la responsabilidad del estudiante, así también, los recursos adecuados que el instituto disponga son unos factores importantes en el rendimiento estudiantil. En conclusión, se recomendó implementar la educación basada en resultados.

# CAPÍTULO 2

## 2. MARCO TEÓRICO

### 2.1. Regresión Lineal Múltiple

La Regresión Lineal Múltiple permite generar un modelo lineal en el que el valor de la variable dependiente o respuesta (Y) se determina a partir de un conjunto de variables independientes llamadas predictores ( $X_1, X_2, X_3\dots$ ). Siendo una extensión de la regresión lineal simple (Zurita, 2008).

Los modelos de regresión múltiple pueden emplearse para predecir el valor de la variable dependiente o para evaluar la influencia que tienen los predictores sobre ella (Fahrmeir, L., & Tutz, G. 2013); la expresión matemática del modelo es la siguiente:

$$Y_i = (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_n X_{ni}) + e_i$$

Donde:

$\beta_0$ : es la ordenada en el origen, el valor de la variable dependiente “Y” cuando todos los predictores son cero, también definido como el intercepto.

$\beta_i$ : es el efecto promedio que tiene el incremento en una unidad de la variable predictora  $X_i$  sobre la variable dependiente Y, manteniéndose constantes el resto de las variables. Se conocen como coeficientes parciales de regresión.

$Y_i$ : Variable dependiente

$X_{1i}, \dots, X_{ni}$ : Son las variables de explicación o predictores, pueden ser las variables originales en el conjunto de datos o las transformaciones o combinaciones de ellas. La matriz de Diseño  $X$  del modelo es:

$$X = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1p-1} \\ 1 & X_{21} & X_{22} & \dots & X_{2p-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & X_{n1} & X_{n2} & \dots & X_{np-1} \end{bmatrix}$$

$e_i$ : es el residuo o error, la diferencia entre el valor observado y el estimado por el modelo. A menudo se debe al efecto de variables no incluidas o no medidas.

La adición de variables cualitativas requiere la introducción de variables ficticias o “dummy” (Olive, 2017).

Los supuestos del modelo de regresión son los siguientes (Gujarati, 2011):

1. Modelo de regresión lineal, o lineal en los parámetros.
2. Valores fijos de X o valores de X independientes del término de error. En este caso, esto significa que se requiere covarianza cero entre  $e_i$  y cada variable X.

$$\text{cov}(e_i, X_{2i}) = \text{cov}(e_i, X_{3i}) = 0$$

3. Valor medio de la perturbación  $e_i$  igual a cero.

$$E(e_i | X_{2i}, X_{3i}) = 0 \text{ por cada } i$$

4. Homocedasticidad o varianza constante de  $e_i$ .

$$\text{Var}(e_i) = \sigma^2$$

5. No autocorrelación, o correlación serial, entre las perturbaciones.

$$\text{Cov}(e_i, e_j) = 0 \quad i \neq j$$

6. El número de observaciones n debe ser mayor que el de parámetros por estimar.
7. Debe haber variación en los valores de las variables X.
8. No debe haber colinealidad exacta entre las variables X.
9. No hay sesgo de especificación.

## 2.2. Estimación de los Coeficientes del Modelo

El método de mínimos cuadrados se puede aplicar para estimar el vector de parámetros  $\hat{\beta}$ , el mismo que se define como:

$$\hat{\beta} = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{Y}$$

Donde:

$\mathbf{X}$  es una matriz de dimensión n x p, mientras que el Vector Aleatorio que estima  $\mathbf{Y}$  viene dado por (Zurita, 2008):

$$\hat{Y} = \hat{\beta}_0 \begin{bmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ 1 \end{bmatrix} + \hat{\beta}_1 \begin{bmatrix} x_{11} \\ x_{21} \\ \cdot \\ \cdot \\ x_{n1} \end{bmatrix} + \dots + \hat{\beta}_{p-1} \begin{bmatrix} x_{1\ p-1} \\ x_{2\ p-1} \\ \cdot \\ \cdot \\ x_{n\ p-1} \end{bmatrix}$$

$X^t$  se denomina a la matriz transpuesta de  $X$ .

Los coeficientes del modelo ( $\beta_i$ ) indican la intensidad y la dirección de la relación entre esa variable independiente (VI) y las variables dependientes (VD) (Olive, 2017):

1. Cuanto más se aleja de 0 más fuerte es la relación, si el valor de  $\beta_i$  es cero, esto indica que la variable  $X_i$  no está relacionada linealmente con la variable Y.
2. El signo indica la dirección (signo + indica que al aumentar los valores de la VI aumentan los valores de la VD; signo – indica que al aumentar los valores de la VI, los valores de la VD descienden).
3. Los valores de los coeficientes están condicionados al modelo de regresión múltiple, es decir que, si cambia el modelo, ya sea porque se eliminan variables regresoras o se añaden nuevas variables, el coeficiente  $\beta_i$  relacionada a la variable regresora  $X_i$  también cambiará.

### 2.3. Tabla de Análisis de Varianza

En la tabla ANOVA se descompone la variación de la variable dependiente Y en dos componentes: primero se evalúa la variación de Y alrededor de los valores predichos por la regresión y después con la variación de los valores predichos alrededor de la media. Ambos estimadores se pueden comparar con la prueba de la F y se determina si el modelo lineal es adecuado (Khuri, A. I. 2009).

La tabla ANOVA diseñada por Fisher proporciona una conveniente tabulación de sumas de cuadrados que miden las cantidades de variación asociadas con los diversos efectos en un modelo dado. La suma total de los cuadrados se determina mediante la suma de los cuadrados del tratamiento (SST) y la suma de los cuadrados del error residual (SSE). La SST es la variación atribuida a los tratamientos y la SSE es la variación atribuida al error (Khuri, A. I. 2009).

Usando razones de cuadrados medios (sumas de cuadrados divididas por sus correspondientes grados de libertad), es posible obtener estadísticos de prueba para ciertas hipótesis relativas a los efectos en consideración (Tabla 2.1).

**Tabla 2.1. Formato de tabla ANOVA**

	<b>Grado de Libertad</b>	<b>Suma de Cuadrados</b>	<b>Cuadrados Medios</b>	<b>F calculada</b>
<b>Modelo</b>	1	SC <sub>modelo</sub>	SC <sub>modelo</sub> /GI <sub>modelo</sub>	Cm <sub>modelo</sub> /Cm <sub>error</sub>
<b>Error</b>	n-2	SC <sub>error</sub>	SC <sub>error</sub> /GI <sub>error</sub>	
<b>Total</b>	n-1	SC <sub>total</sub>		

*Fuente: Khuri, A. I. 2009*

Una variable que sigue una distribución F se denota por  $F_{k_1, k_2}$ , donde los subíndices indican los grados de libertad asociados con las variables  $Z_1$  y  $Z_2$ , las cuales son variables ji cuadrada distribuidas en forma independiente, como se muestra en la siguiente ecuación (Gujarati, 2011):

$$F = \frac{Z_1/k_1}{Z_2/k_2}$$

La distribución F tiene las siguientes propiedades:

1. Al igual que la distribución ji cuadrada, la distribución F está sesgada hacia la derecha. También se puede notar que a medida que aumentan  $k_1$  y  $k_2$ , la distribución F se acerca a la distribución normal
2. El cuadrado de una variable con distribución t con k grados de libertad sigue una distribución F con 1 y k grados de libertad. Se expresa de la siguiente forma:

$$t_k^2 = F_{1, k}$$

3. El valor de la media de una variable con distribución F es  $\frac{k_2}{(k_2-2)}$ , él está definido para  $k_2 > 2$  y su varianza es:

$$\frac{2k_2^2(k_1 + k_2 - 2)}{k_1(k_2 - 2)^2(k_2 - 4)}$$

Definida para  $k_2 > 2$

La prueba F que se deriva de la tabla ANOVA prueba la hipótesis de que los efectos de los niveles de cada factor son iguales a cero, como se muestra a continuación (Fahrmeir, L., & Tutz, G. 2013). Además, permite verificar la validez de las restricciones lineales impuestas sobre uno o más parámetros del modelo de regresión lineal de k variables (Gujarati, 2011).

Por otro lado, al realizar el contraste de hipótesis en un modelo de regresión múltiple, permitirá realizar inferencia acerca de parámetros poblacionales utilizando datos provenientes de una muestra (Zurita, 2008).

En general, para realizar el contraste de hipótesis estadístico hay que realizar los siguientes pasos (Gujarati, 2011):

1. Establecer una hipótesis nula y una hipótesis alternativa relativos a los parámetros de la población

$$H_0 \equiv \beta_1 = \beta_2 = \dots = \beta_n = 0$$

La afirmación contraria sería:

$H_1$ : Al menos uno de los betas no es cero

2. Construir un estadístico para contrastar las hipótesis formuladas.
3. Definir una regla de decisión para determinar si la hipótesis nula debe ser, o no, rechazada en función del valor que tome el estadístico construido.

Es decir, en el caso de Regresión Lineal, si  $H_0$  es verdadero, ninguna de las variables regresaras influye en la variable de respuesta, es decir que el modelo no es válido.

Si el valor p es menor de 0.05, se rechaza  $H_0$  en favor de  $H_1$ , es decir que el modelo es estadísticamente significativo y por tanto las variables independientes

explican “algo” de la variable dependiente. Una medida de cuan bueno es el ajuste es el coeficiente de determinación  $R^2$  (Zurita, 2008).

p-valor del contraste es mayor al nivel de significación  $\alpha$  (0.05) se acepta  $H_0$ , es decir que el modelo de regresión no es influyente y debe buscarse un modelo alternativo. Si el p-valor es menor de 0,05 es que el modelo es estadísticamente significativo y por tanto las variables independientes explican “algo” de la variable dependiente, cuánto “algo” eso se determina mediante el R-cuadrado (Olive, 2017).

## 2.4. Pruebas de hipótesis para los parámetros del modelo

Conocidas las expresiones correspondientes a los intervalos de confianza, se deberá obtener los p-valor de cada una de las hipótesis que se puedan presentar en relación con los parámetros del modelo (Fahrmeir, L., & Tutz, G. 2013).

### 2.4.1. Intervalos de confianza

A los intervalos de confianza (IC) se les denomina también como estimaciones por intervalo, y esto se debe a que proporcionan un rango de valores verosímiles para  $\beta_j$  y no solamente una estimación puntual (Gujarati, 2011). El IC se construye de tal manera que el parámetro desconocido está contenido dentro del recorrido del IC con una probabilidad previamente especificada (Zurita, 2008).

### 2.4.2. Inferencias sobre los parámetros del modelo

Para los coeficientes de regresión, el intervalo de confianza viene dado por la siguiente expresión:

$$\beta_i = \hat{\beta}_i \pm t_{n-p, \frac{1+\gamma}{2}} \hat{\sigma}_i, i = 0, \dots, k$$

Siendo

$$\hat{\sigma}_i = \hat{\sigma} \sqrt{c_{i+1, i+1}}$$

$c_{i+1, i+1}$  = El elemento que ocupa la posición (i+1) en la diagonal principal de la matriz  $C = (X^T X)^{-1}$

Por otro lado, para la varianza del término de error

$$\sigma^2: \left( \frac{SSE}{X^2_{n-p, \frac{1+\gamma}{2}}}, \frac{SSE}{X^2_{n-p, \frac{1-\gamma}{2}}} \right)$$

## 2.5. R cuadrado

Este coeficiente es muy importante pues determina qué porcentaje de la varianza de la variable dependiente es explicado por el modelo de regresión. Cuanto mayor sea la R-cuadrado más explicativo y mejor es el modelo causal (Gujarati, 2011)

De este modo, se define el coeficiente de determinación  $R^2$  de la siguiente manera:

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} = 1 - \frac{\sum \hat{u}_i^2}{\sum \hat{y}_i^2}$$

Por ser cociente de sumas de cuadrados, este coeficiente será siempre positivo. Si todos los puntos están sobre la recta de regresión, la varianza no explicada será 0, y, por lo tanto:

$$R^2 = \frac{VE}{VT} = 1 - \frac{0}{VT} = 1$$

## 2.6. $R^2$ ajustado

Para comparar dos términos  $R^2$  se debe tener en cuenta el número de variables de X presentes en el modelo (Gujarati, 2011). Para determinar el valor de  $R^2$  ajustada se debe considerar un coeficiente de determinación alternativo, siendo:

$$\bar{R}^2 = 1 - \frac{\sum \hat{u}_i^2 / (n - k)}{\sum \hat{y}_i^2 / (n - 1)}$$

Donde:

K= es el número de parámetros en el modelo incluyendo el término de intercepto

$\bar{R}^2$ = Designada como  $R^2$  ajustada.

n= número de variables

El término “ajustado” significa que  $R^2$  es ajustado por los grados de libertad asociados a las sumas de cuadrados (Gujarati, 2011).

## **2.7. Selección de variables**

Cuando se tiene una gran cantidad de variables explicativas, se debe determinar cuáles de estas variables deben ser consideradas en el modelo de regresión. Para poder seleccionar estas variables podemos hacer uso de algoritmos como: método Backward, método Forward, métodos Stepwise, Best subset, entre otros (Olive, 2017).

## **2.8. Método Backward**

Denominado como “eliminación hacia atrás”, Partiendo de un modelo muy complejo, incorporando todas las variables que pueden influir en la respuesta, luego en las siguientes etapas se eliminarán las variables menos influyentes, hasta determinar las variables que causan un efecto importante en la respuesta (Olive, 2017).

## **2.9. Método Forward**

Se inicia de un modelo muy sencillo y se van agregando variables, conocido también como selección hacia adelante. En cada etapa se introduce la variable más significativa hasta que se cumple una cierta regla de parada (Khuri, 2009).

Para determinar la condición para suprimir o incluir un término dentro del modelo, se suelen considerar dos criterios: Criterios de significación del término y criterios de ajuste global (Gonzales, 2015).

Los métodos Forward y Backward utilizan criterios de significación. El criterio de significación puede ser la significación de cada coeficiente. Por otro lado, en vez de utilizar la significación de cada coeficiente, podemos aplicar un criterio global, de modo que se tenga en cuenta el ajuste y el exceso de parámetros. De este tipo de criterio se destacan: El Criterio de Información Akaike (AIC) y el criterio de Información de Bayes (BIC) (Gonzales, 2015).

## **2.10. Herramientas de calidad para el análisis de problemas**

El análisis de problemas puede ser realizado a partir de distintas herramientas de calidad, como: Lluvia de ideas (brainstorming), histogramas, análisis de Pareto, análisis de Causa-Efecto, etc. (Lemos, 2016).

Estas herramientas básicas son aplicadas en la solución de problemas con respecto a la calidad en las empresas, siendo adecuadas por su versatilidad y facilidad de aplicación (Izar y González, 2004).

### **2.10.1. Lluvia de ideas o brainstorming**

Es una técnica grupal que permite la generación de un gran número de ideas sobre un tema establecido, con el objetivo de buscar las causas de un problema, además también permite complementar las versiones de un problema (Izar y González, 2004).

### **2.10.2. Histogramas**

Permite visualizar en una tabla de datos las frecuencias y la distribución de una variable determinada (Izar y González, 2004).

### **2.10.3. Análisis de Pareto**

El análisis de Pareto clasifica de forma gráfica, utilizando barras, los problemas para dar prioridad a los más importantes por orden descendente, de izquierda a derecha. El principio de esta técnica se basa, en que el 80% de los resultados totales se originan en el 20% de los elementos, es decir que el 20% de las causas resuelven el 80% del problema y viceversa, y por tanto el 80% de las causas solo resuelven el 20% del problema (Izar y González, 2004).

### **2.10.4. Análisis de Causa-Efecto**

Mediante el diagrama de “espina de pez” de Ishikawa, se trata de identificar los factores (Causas) que originan un problema.

En este diagrama se representan cuatro tipos de factores: Personales, procedimentales, funcionales (tiempo-espacio) y materiales (Lozano, 2003).

Personales: Factores intrínsecos, referente a los individuos que realizan un proceso.

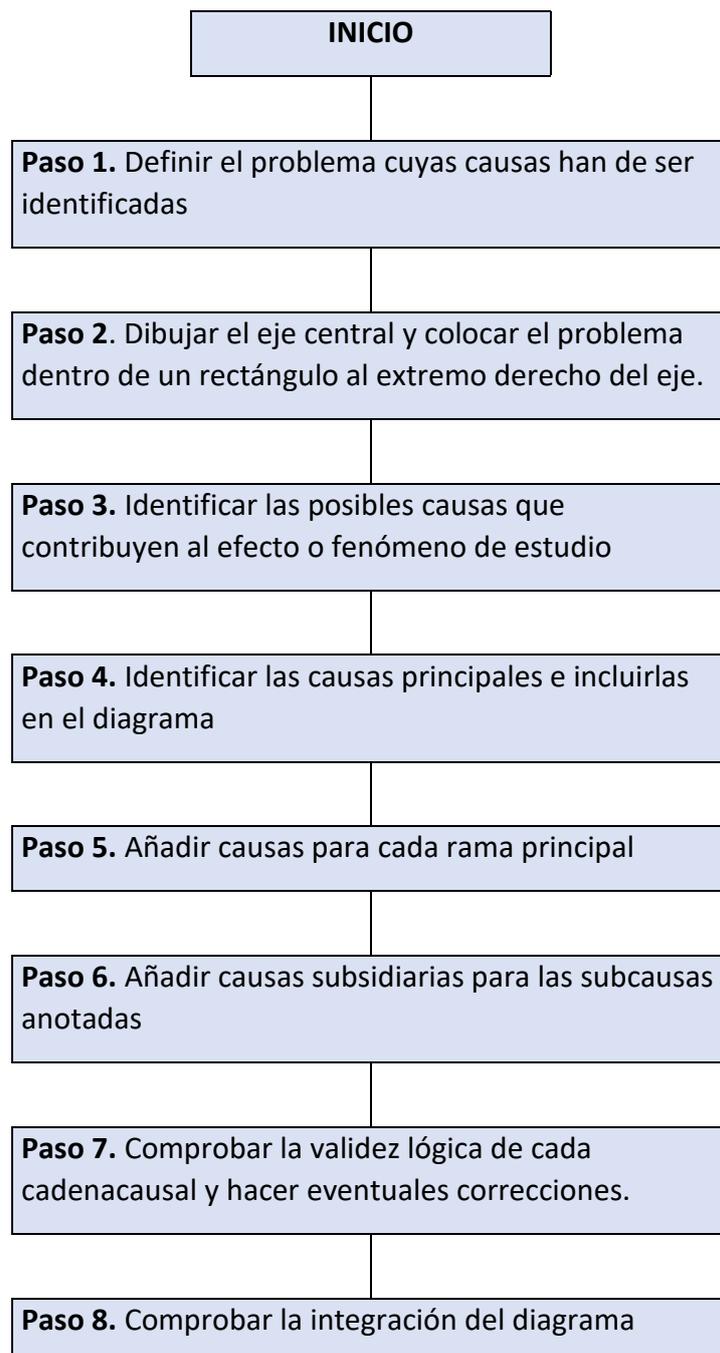
Procedimentales: La forma o el método como se realiza un proceso.

Funcionales: Factores extrínsecos

Materiales: Los recursos que se utilizan para realizar un proceso

Según Izar y González (2004), para evaluar e identificar las causas a través de la “espina de pez” de Ishikawa y del histograma se realizan los pasos del Diagrama de Ishikawa.

**Figura 2.1. Pasos para la elaboración del Diagrama de Ishikawa**



**Paso 9. Conclusión y resultado**

**FIN**

*Fuente: Fundibeq, 2012.*

# CAPÍTULO 3

## 3. RESULTADOS

### 3.1. Recolección de datos

Los datos fueron obtenidos de los registros administrativos de entrevistas y pruebas hechas a estudiantes del primer ciclo de un instituto tecnológico superior de Guayaquil.

### 3.2. Variables

#### 3.2.1. Variables dependientes

En el presente proyecto se consideró la variable cuantitativa “puntuación”, el cual es un indicador del rendimiento académico y mide el progreso del estudiante. La escala de valorización es del 1 al 20.

#### 3.2.2. Variables independientes

En este estudio se consideraron diferentes criterios para caracterizar los perfiles de alumnos exitosos. Además, para analizar las variables cualitativas, estas fueron convertidas a variables *dummy*. Las variables analizadas fueron las siguientes:

VI1: Sexo del estudiante, variable cualitativa, la cual tiene dos niveles: Femenino y Masculino; Variable dummy: sexo, 1: Masculino y 0: Femenino.

VI2: Edad de estudiante, variable cuantitativa.

VI3: Estado civil del estudiante, variable cualitativa, misma que tiene tres niveles: Soltero, casado y divorciado; Variable dummy: Soltero: 1 y Casado: 0

VI4: Lugar de residencia del estudiante, variable cualitativa, con dos niveles: Guayaquil y fuera de Guayaquil; Variable dummy: Residencia, 1: Guayaquil y 0: Fuera de Guayaquil.

VI5: Tipo de colegio del cual proviene el estudiante, variable cualitativa, con tres niveles: Particular, fiscal y distancia; variable dummy: Particular: 1 y fiscal: 0

VI6: Especialidad o bachillerato en el que se graduó el estudiante, variable cualitativa, con tres niveles: Ciencias, técnico y otros; variable dummy: Ciencias: 1 y técnico: 0

VI7: Situación laboral del estudiante, variable cualitativa, con dos variables: Estudia y trabaja o sólo estudia; variable dummy: Laboral, 1: Estudia y 0: Estudia y trabaja.

VI8: Asistencia a clases, variable cualitativa, con dos niveles: Asiste con regularidad y No asiste con regularidad; variable dummy: Asistencia, 1: Asiste con regularidad y 0: No asiste con regularidad.

VI9: El estudiante ha pasado más de un mes sin estudiar, variable cualitativa, con dos niveles: Si o NO; variable dummy: sin estudiar, 1: Si y 0: No

VI10: El estudiante proviene de familia de padres separados, variable cualitativa, con dos niveles: padres separados o padres no separados; variable dummy: Estado civil padres, 1: Padres separados y 0: Padres no separados.

VI11: Con cuantas personas vive el estudiante, variable cuantitativa.

VI12: Nivel académico de los padres, variable cualitativa, con cuatro niveles: Superior, secundaria, primaria y ninguno. Variables dummy: Superior, secundaria y primaria: 1, Ninguno: 0

VI13: Situación laboral de los padres, variable cualitativa, con dos niveles: Trabaja o No trabaja. Variable dummy: Padres trabajan, 1: No trabaja y 0: Trabaja.

VI14: Horas dedicadas al estudio en casa, variable cuantitativa.

VI15: El estudiante planea seguir estudios de posgrado, variable cualitativa, con dos niveles: Si o No. Variable dummy: Estudiar postgrado, 1: Si y 0: No

VI16: El estudiante utiliza TICs al estudiar, variable cualitativa, con dos niveles: Si o No. Variable dummy: tics, 1: Si y 0: No

VI17: Calidad de enseñanza en matemáticas recibidas en el colegio, variable cuantitativa.

VI18: Interés del estudiante por la asignatura de matemáticas, variable cuantitativa.

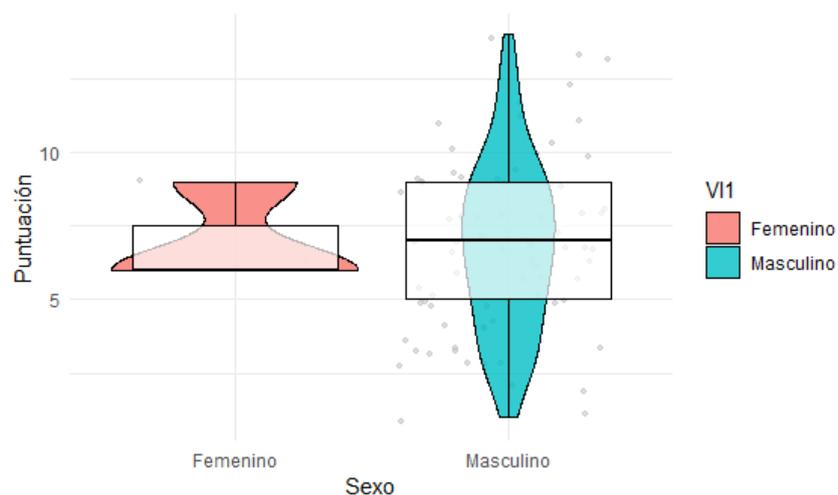
VI19: Grado de complejidad, variable cuantitativa.

### 3.3. Análisis descriptivos de las variables

Para la elaboración del perfil de los estudiantes se generaron diagramas combinados de cajas y bigotes (*boxplot*) y un diagrama de densidad, estas gráficas representan la distribución kernel y la densidad de probabilidad de cada uno de los predictores cualitativos para intuir su influencia en la variable dependiente, cabe recalcar que el objetivo del proyecto no es predecir la nota de puntuación del estudiante, si no identificar las causas en el rendimiento, razón por la cual no se busca mejorar el  $R^2$ . También se puede ver el resumen de los datos en el anexo Tabla A-5.

De un total de 81 alumnos, 78 son de género masculino y 3 de género femenino. En la figura 3.1 podemos observar que los hombres tienen un mejor rendimiento que las mujeres. Este resultado podría deberse a que existe una mayor cantidad de hombres que de mujeres, es decir que la muestra no es representativa, por lo tanto, no se podría decir que el sexo influye en el rendimiento. El p-value para la relación de estas variables fue de 0.9107, por lo que se puede inferir que no hay relación entre la puntuación y el género.

**Figura 3.1. Boxplot del sexo del estudiante con respecto a la puntuación**

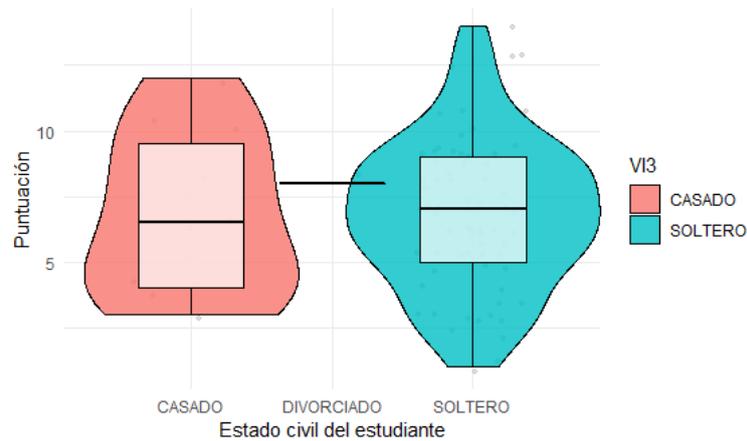


*Fuente: Espín, 2018.*

También se analizó el rendimiento con relación al estado civil del estudiante, de los cuales 70 son solteros, 10 casados y 1 divorciado. En la figura 3.2 Se

observa que al estar el estudiante soltero obtiene un mejor rendimiento en comparación con el estudiante que está casado. Con respecto al predictor divorciado, se lo podría considerar como un dato con poca validez porque no existe suficiente cantidad de observaciones.

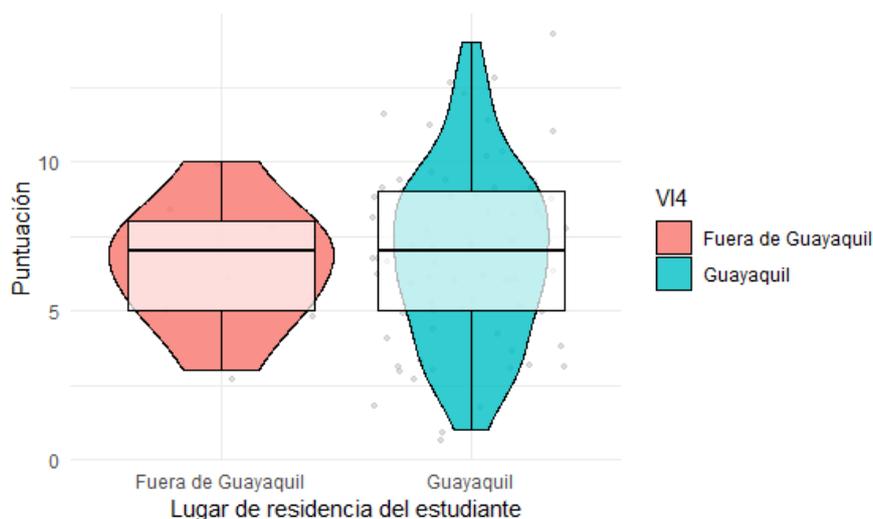
**Figura 3.2. Boxplot del estado civil del estudiante con respecto a la puntuación**



*Fuente: Espín, 2018.*

De los datos obtenidos existen 72 estudiantes que residen en Guayaquil y 9 que viven fuera de Guayaquil. En la figura 3.3 se observa que aparentemente no existe diferencia en el rendimiento con respecto al lugar de residencia del estudiante.

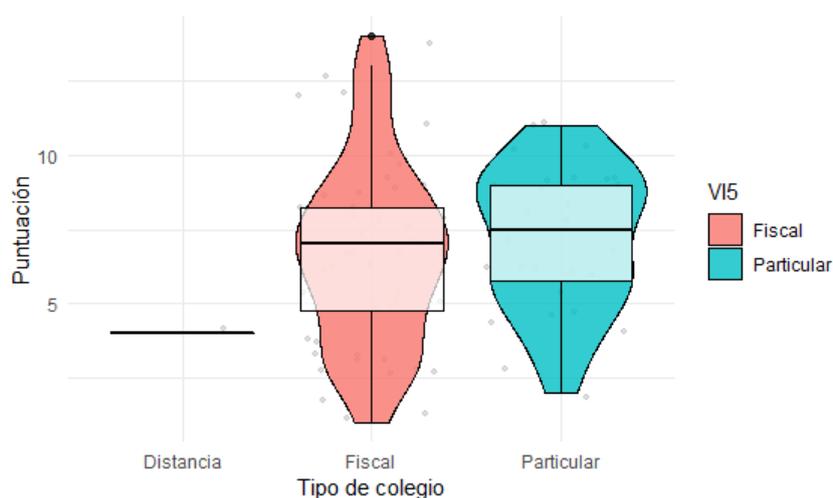
**Figura 3.3. Boxplot lugar de residencia del estudiante con respecto a la puntuación**



Fuente: Espín, 2018.

Al analizar el tipo de colegio de donde provienen los estudiantes se determinó que 28 de ellos vienen de colegio particular, 52 de colegio fiscal y 1 a distancia. En la figura 3.4 se puede observar que el estudiante obtiene un mejor rendimiento al provenir de un colegio particular pero que esta no se diferencia mucho en el caso de provenir de un colegio fiscal, pero si se observa una diferencia importante en el caso de provenir de un colegio a distancia. Aunque el número de observaciones de estudiantes que provienen de un colegio a distancia no es considerable para poder afirmarlo.

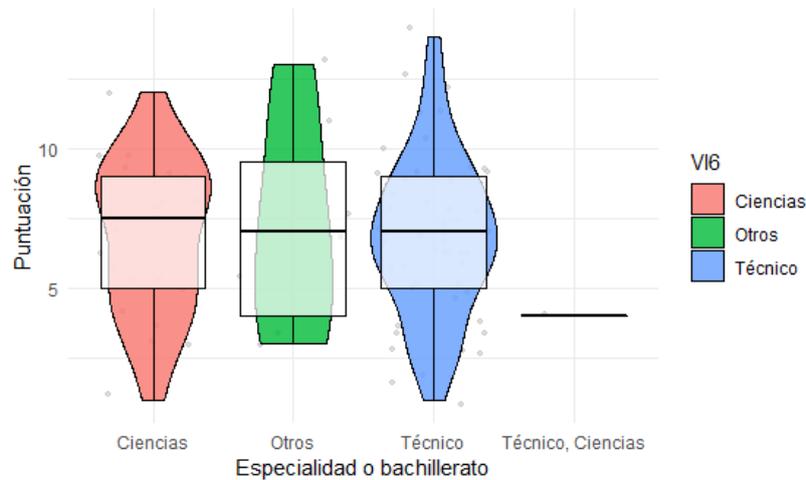
**Figura 3.4. Boxplot tipo de colegio del estudiante con respecto a la puntuación**



Fuente: Espín, 2018.

Del total de estudiantes 22 se graduaron en ciencias, 51 de técnicos, 7 otro tipo de especialidad y 1 realizó un bachillerato en ciencias y es técnico. En la figura 3.5 se observa que, al analizar la especialidad del estudiante en el bachillerato con el rendimiento académico en el tecnológico, las especialidades de ciencias en promedio obtienen un mejor desempeño en comparación con las otras especialidades. También se debe notar que existen estudiantes con especialidad técnica que alcanzan mejores puntajes que los estudiantes de las otras especialidades.

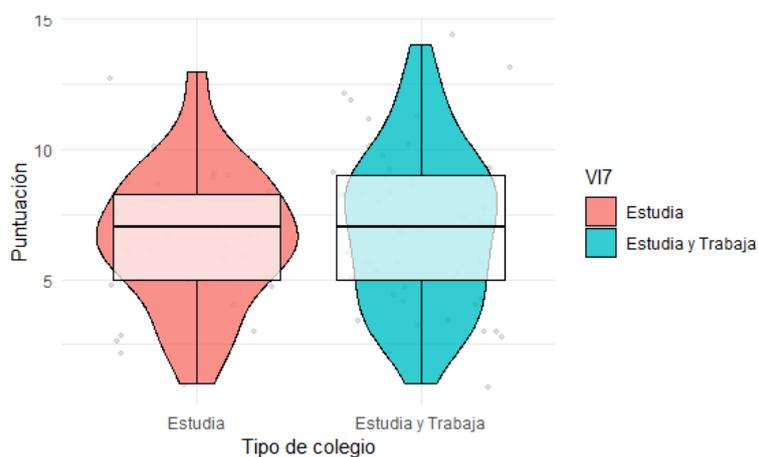
**Figura 3.5. Boxplot especialidad/bachillerato del estudiante con respecto a la puntuación**



*Fuente: Espín, 2018*

La cantidad de estudiantes que estudia y trabaja es de 49 y aquellos que sólo estudian es de 32. Al observar la figura 3.6, se concluye que la situación laboral del estudiante no influye en su rendimiento académico. Por otro lado, también se puede notar que los estudiantes que estudian y trabajan pueden alcanzar un mayor rendimiento que aquellos estudiantes que solo estudian.

**Figura 3.6. Boxplot situación laboral del estudiante con respecto a la puntuación**

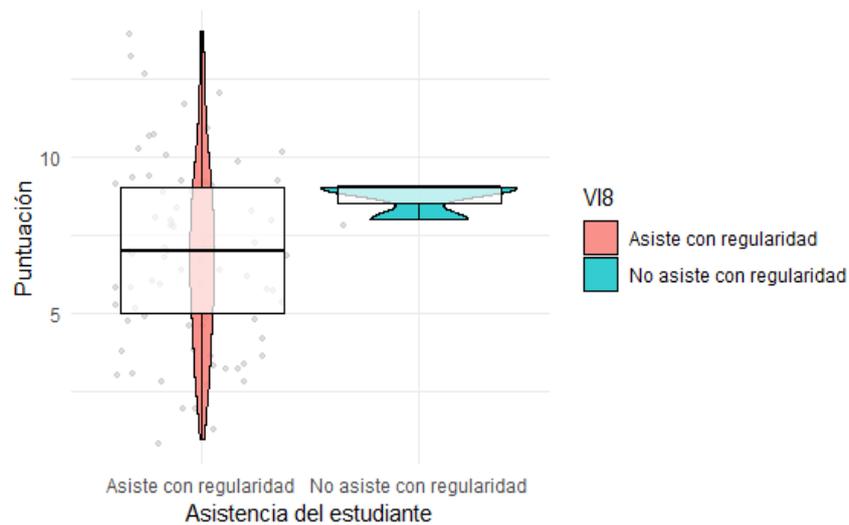


*Fuente: Espín, 2018.*

Los estudiantes que asisten con regularidad son 78 y los que no asisten con regularidad son 3. Según lo que se observa en la figura 3.7, los estudiantes que asisten con regularidad tienen mayor probabilidad de obtener mejores resultados,

aunque se observa que en promedio los alumnos que no asisten a clases tienen un mejor resultado, esto se debe a que la cantidad de casos no es representativa.

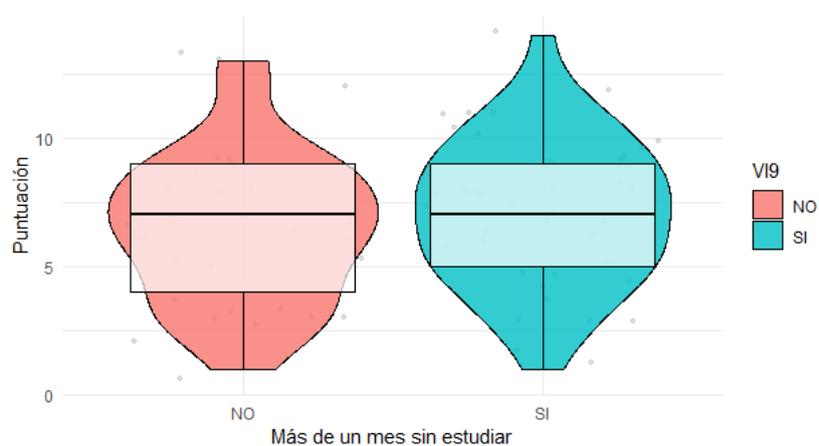
**Figura 3.7. Boxplot asistencia a clases con respecto a la puntuación**



*Fuente: Espin, 2018.*

De los datos obtenidos 44 estudiantes han dejado de estudiar por más de un mes y 37 estudiantes no han dejado de estudiar por más de un mes. En la figura 3.8 se observa que no hay diferencia en los valores promedios con respecto al rendimiento.

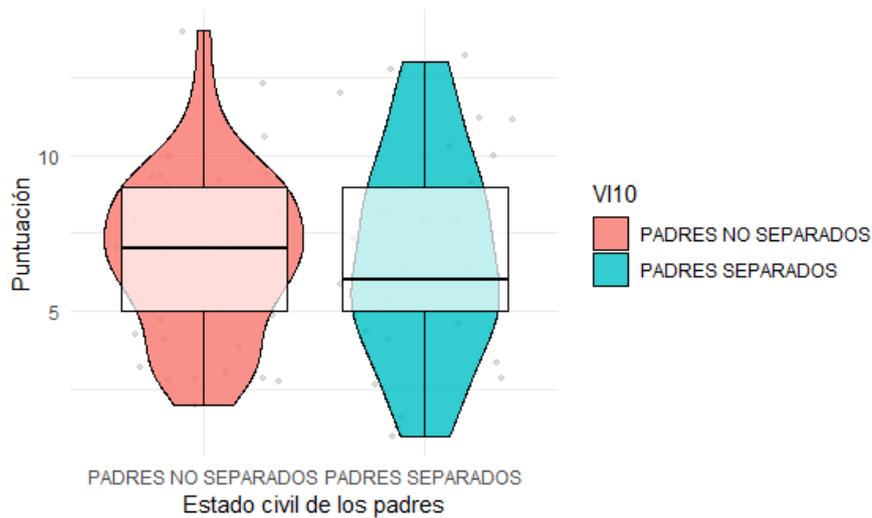
**Figura 3.8. Boxplot tiempo sin estudiar con respecto a la puntuación**



*Fuente: Espin, 2018.*

De lo estudiantes entrevistados 48 provienen de padres no separados y 33 proviene de padres separados. Se observa en la figura 3.9, los estudiantes que provienen de padres no separados tienen un mejor rendimiento.

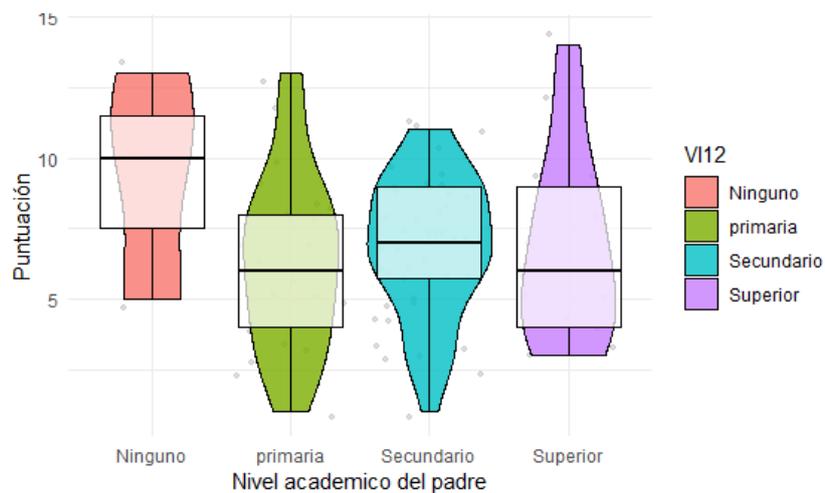
**Figura 3.9. Boxplot estado civil de los padres con respecto a la puntuación**



*Fuente: Espín, 2018.*

Se analizó también el nivel de estudio de los padres de los cuales se determinó que 21 estudiantes sus padres sólo llegaron hasta la primaria, 44 de ellos solo terminaron la secundaria, 13 tenían un título superior y 3 de ellos no tenían ninguna preparación. En la figura 3.10, se observa que en promedio los alumnos cuyos padres no tenían ninguna preparación lograron obtener mejores rendimientos que los otros tipos de alumnos cuyos padres tenían algún nivel de estudio.

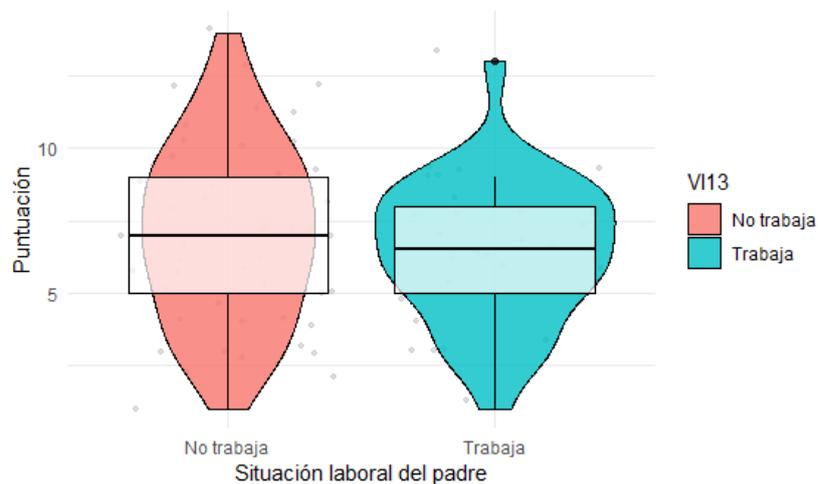
**Figura 3.10. Boxplot nivel académico del padre con respecto a la puntuación.**



*Fuente: Espín, 2018.*

Con respecto a la situación laboral del padre se determinó que 51 de ellos no trabajan y 30 de ellos trabajan. En la figura 3.11, se observa que en promedio esto no influyen de forma significativa en el rendimiento, pero por otro lado se observa que aquellos estudiantes que sus padres trabajan tienden a tener un mejor rendimiento.

**Figura 3.11. Boxplot situación laboral del padre con respecto a la puntuación**

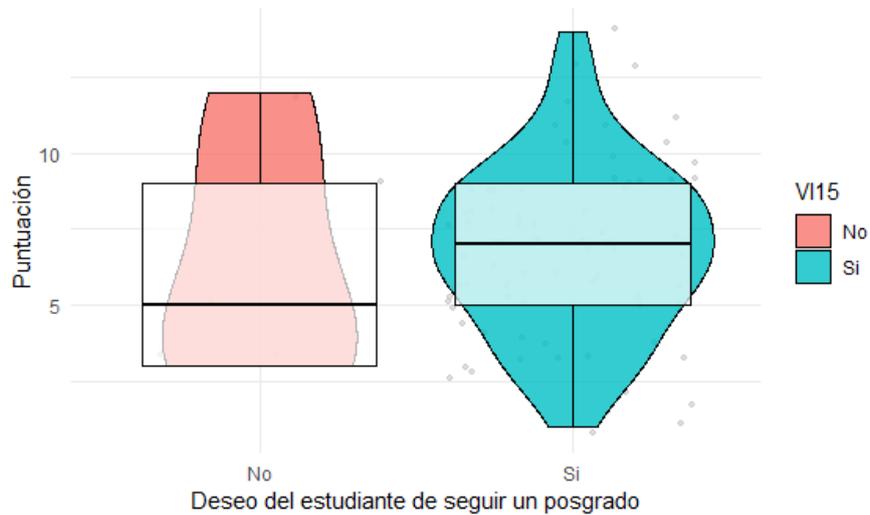


*Fuente: Espin, 2018.*

Se evaluó también el deseo de los estudiantes por seguir un posgrado de los cuales 76 de ellos contestaron afirmativamente y 5 dijeron que no. En la figura 3.12,

se observa que los estudiantes que desean seguir un posgrado tienen un mejor rendimiento.

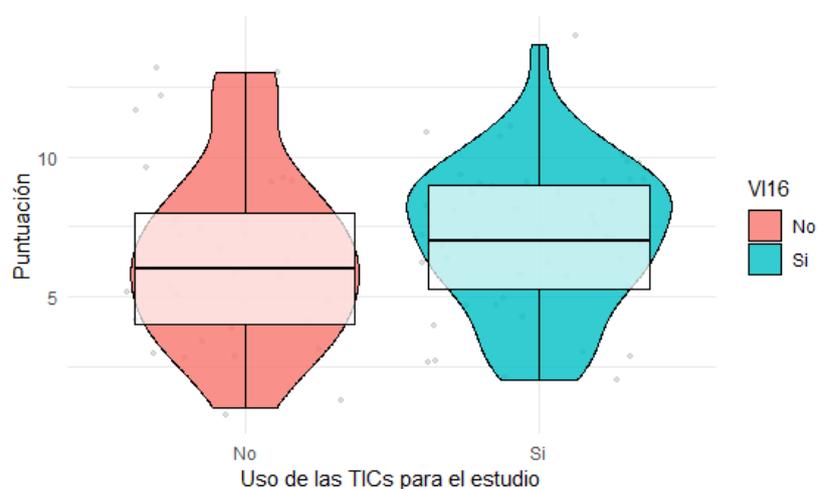
**Figura 3.12. Boxplot deseo de seguir posgrado con respecto a la puntuación**



*Fuente: Espín, 2018.*

Por último, se analizó el rendimiento de los estudiantes en base al uso de las TICs, 46 estudiantes declararon utilizar estas herramientas tecnológicas y 35 estudiantes dijeron que no. En la gráfica 4 se puede notar que los estudiantes que utilizan TICs poseen en promedio un mejor rendimiento.

**Figura 3.13. Boxplot uso de TICs con respecto a la puntuación**

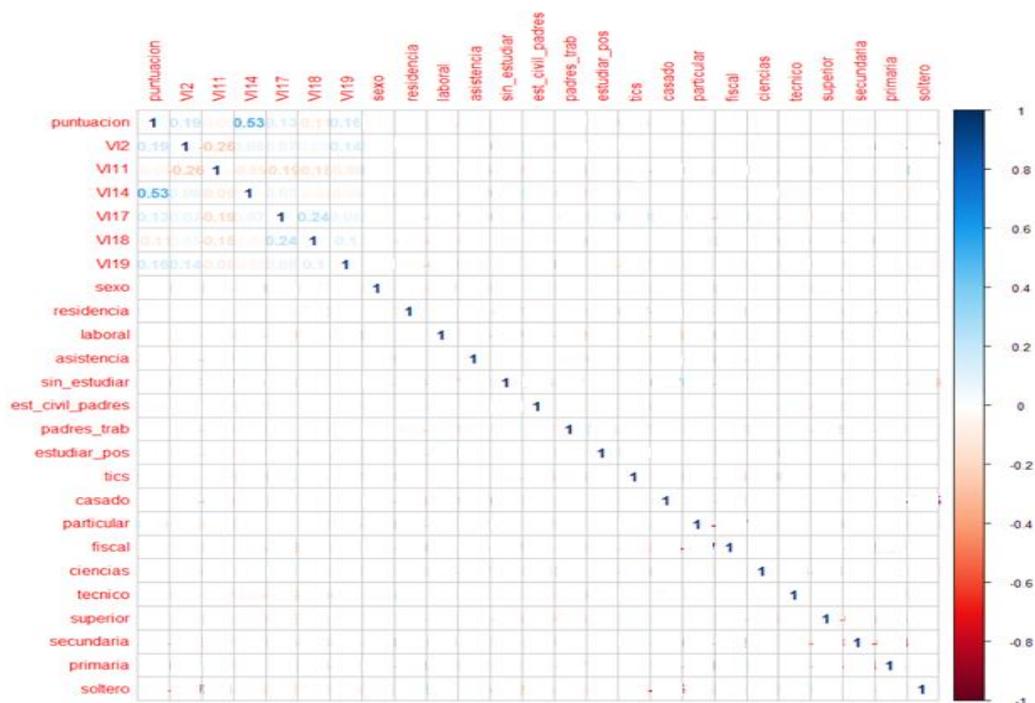


*Fuente: Espín, 2018.*

### 3.4. Análisis del rendimiento académico de los estudiantes a través de un modelo de Regresión Lineal Múltiple.

Las variables fueron analizadas mediante un diagrama de dispersión múltiple, para poder de determinar si existe una relación lineal o monótonica con la variable de respuesta. ver figura 3.14. Valores de correlación próximos a 1 o -1 indican una alta correlación entre variables.

**Figura 3.14. Matriz de diagrama de dispersión múltiple**



Fuente: Espín, 2018.

Los coeficientes de correlación se calcularon por el método de Pearson, en la gráfica 4.1. se observa cómo se correlacionan las variables cuantitativas (VI2: Edad, VI11: personas con la que vive el estudiante, VI14: Horas dedicadas al estudio en casa, VI17: Calidad de enseñanza en matemáticas recibidas en el colegio, VI18: Interés del estudiante por la asignatura de matemáticas, VI19: Grado de complejidad).

Los resultados indican que existe una correlación positiva fuerte de 0.53 entre la variable “puntuación” y la variable VI14 (Horas dedicadas al estudio).

Para la generación del modelo, primero se evaluó el modelo con todos los predictores (Ver Anexo: Tabla A-2). Los coeficientes obtenidos para cada uno de los predictores fueron los siguientes (Tabla 4.1):

**Tabla 3.1. Coeficiente para un modelo con todos los predictores**

(Intercept)	VI2	VI11	VI14	VI17
-6.175885037	0.161551833	-0.004395866	2.301620162	0.100387570
VI18	VI19	sexo	residencia	laboral
-0.182778082	0.209254590	1.028317601	1.020487966	-0.177638600
asistencia	sin_estudiar	est_civil_padres	padres_trab	estudiar_pos
-1.943439752	-0.189002678	-0.231231813	0.505022400	1.311974106
tics	casado	soltero	particular	fiscal
0.640209804	0.699165657	2.374480393	2.795861452	2.990450602
ciencias	tecnico	superior	secundaria	primaria
0.361520693	-0.069418505	-1.077411052	-1.248027700	-2.306902062

*Fuente: Espín, 2018.*

Para este modelo inicial los resultados determinaron que:

1. El modelo con todos los predictores es capaz de explicar el 47.65% de la varianza observada en el rendimiento de los estudiantes ( $R_{ajustado}^2 = 0.2521$ ).
2. El p-value del modelo del modelo es significativo (0.010), esto indica que el modelo es útil y que existe una relación entre los predictores y la variable de respuesta.
3. Los predictores que tiene significancia son VI2 y VI14.

### 3.4.1. BIC y ADJR2

Para seleccionar los predictores se utilizó el método Bayesian Point of View (BIC) y además también se realizó el método de  $R^2$ -ajustado. En la gráfica 3.15 se observa cuáles fueron las variables seleccionadas para el modelo de regresión lineal múltiple según el método empleado.



### 3.4.2. Método Backward

El modelo desarrollado por este método fue el siguiente:

(Intercept)	VI2	VI14	VI17	VI18	VI19	asistencia
-3.2780289	0.1435270	2.3282302	0.1750782	-0.1966050	0.2185999	-1.7002718
estudiar_pos	soltero	superior	secundaria			
1.5233444	2.0408749	1.2076717	1.0283018			

Con un valor  $R^2$ -adj de 0.3370337

### 3.4.3. Método Forward

El modelo obtenido por este método fue el siguiente:

(Intercept)	VI2	VI14	VI18	VI19	residencia	asistencia	padres_trab
-1.3053385	0.1126565	2.3192998	-0.1372163	0.2464647	0.8802436	-1.5946534	0.8265706
tics	soltero						
0.7496953	1.5929834						

El valor  $R^2$ -adj fue de 0.337281

### 3.4.4. Método stepwise mixto y AIC

Para determinar la calidad del modelo se utilizó una serie de procedimientos de selección automática de variables significativas, denominado el método del stepwise mixto y el criterio de Información de Akaike (AIC).

El modelo obtenido mediante esta metodología fue el siguiente:

$$Y = -2.5567 + 0.1655 * \text{Edad} + 2.4060 * \text{Horas\_dedicadas\_al\_estudio} + 0.8705 * \text{Padres\_trabajan} + 2.1570 * \text{Soltero} - 1.5410 * \text{Primaria}$$

El valor de  $R^2$ -adj fue de 0.3595 y p-value de  $2.39 \times 10^{-7}$

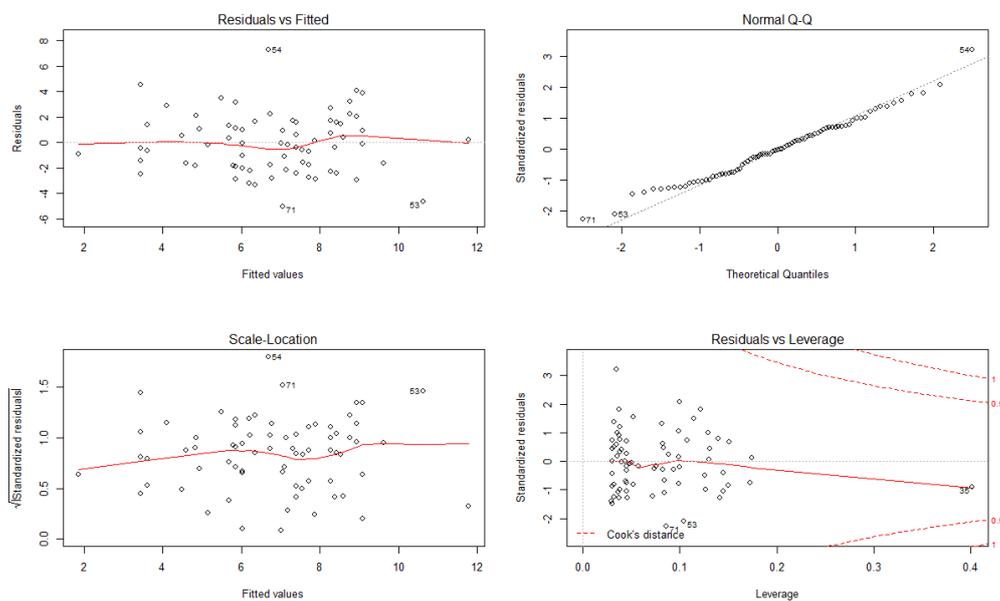
Según el principio de parsimonia, el modelo que se debe seleccionar como adecuado es el que contiene menor números de predictores.

Además, en la gráfica 4.2 se analizaron los residuos de este modelo:

1. Residual vs Filted, no hay evidencia que indique heterocedasticidad y multicolinealidad.
2. Normal Q-Q, se observa que los residuos se distribuyen de forma normal con excepción de los datos extremos 53, 54 y 71.

3. Scale-location, esta gráfica muestra si los residuos se reparten equitativamente a lo largo de los rangos de los predictores. Esta es la forma como se puede evaluar el supuesto de varianza igual. Se puede observar claramente que existe homocedasticidad.
4. Residual vs leverage, esta gráfica nos ayuda a encontrar casos influyentes. En la gráfica apenas se puede ver las líneas de distancia de Cook indicando que el dato 35,53,54 y 71 no son influyentes.

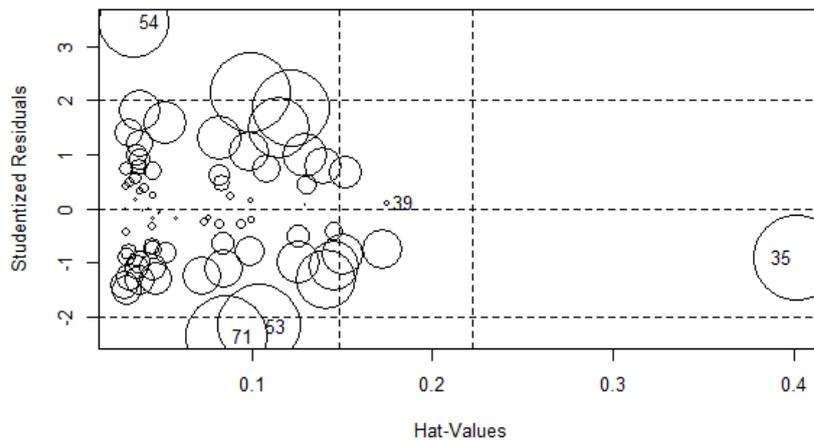
**Figura 3.17. Análisis de los residuos**



*Fuente: Espín, 2018.*

Se observa en la gráfica 3.18 varias observaciones influyentes y que los datos 35, 53, 54 y 71 salen fuera de los límites superior e inferior establecidos.

**Figura 3.18. Análisis de los residuos estudentizados**



*Fuente: Espin, 2018.*

### 3.5. Análisis de causa-efecto para la mejora del rendimiento académico

Finalmente se procedió a realizar un análisis de causa-efecto mediante un diagrama de espina de pescado. Las principales categorías de causas que se analizaron fueron:

1. **Personales:** Estado civil del estudiante, Estado civil de los padres y padres con trabajo.

En los diagramas de boxplot y los modelos de regresión lineal múltiple se pudo observar que el estudiante soltero que vive en una familia con ambos padres y cuyo padre trabaje presenta mayores probabilidades de obtener un mayor rendimiento.

2. **Socioeconómico:** Actitud del estudiante y recursos económicos.

Con respecto al factor socioeconómico es preferente que el alumno sea soltero y no se encuentre trabajando por lo menos al ingreso de la carrera.

3. **Académico:** Asistencia a clases, horas de estudio, tipo de bachillerato, deseo de seguir posgrado y utiliza TICs.

Es importante destacar que las horas que se dedique al estudio es uno de los predictores con mayor correlación con respecto al rendimiento. Además, el alumno debe presentar las siguientes actitudes: autoaprendizaje y deseos de superación. También es importante el nivel de bachillerato en el que haya graduado el alumno,

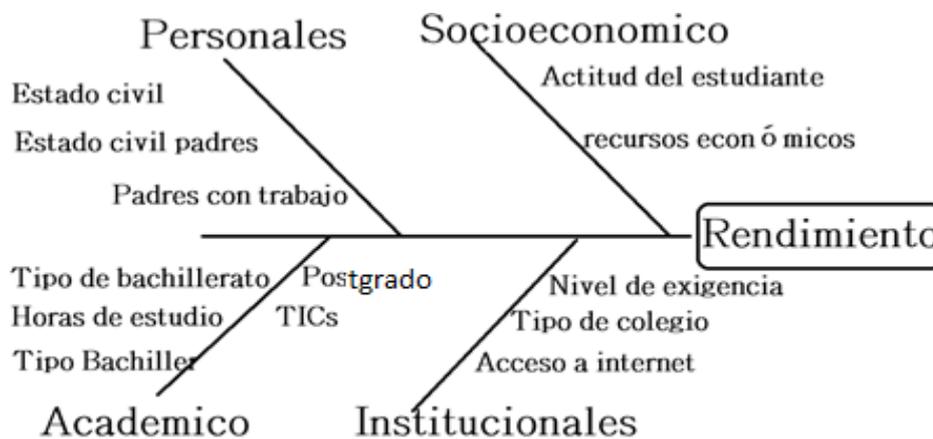
siendo el bachillerato en ciencias el que brinda una correlación significativa con respecto al rendimiento.

4. **Institucionales:** Nivel de exigencia, tipo de colegio y acceso a internet

El nivel de exigencia durante el bachillerato y el acceso a la información también es un factor importante para el rendimiento del alumno. Así, como el tipo de colegio del cual proviene.

En la gráfica 3.19, se puede observar la representación de las causas que producen un efecto en el rendimiento.

**Figura 3.19. Diagrama de causa-efecto para mejora de rendimiento académico en tecnológico**



*Fuente: Espín, 2018.*

# CAPÍTULO 4

## 4.1 DISCUSIÓN

El objetivo de las instituciones de educación superior como universidades y tecnológicos es el de preparar y generar en los estudiantes destrezas, habilidades y los conocimientos necesarios para desarrollarse en su futura profesión. En este contexto un indicador utilizado para determinar la eficacia y la calidad educativa es el rendimiento académico.

Por los resultados obtenidos en este estudio, se debe notar que el rendimiento académico no solo depende de factores intelectuales del estudiante, sino también de factores personales, institucionales y socioeconómicos; como lo dice (González, C., 2004), en el cual el objetivo de su estudio era encontrar factores que determinen el bajo rendimiento académico en educación secundaria, para lo cual semejante a este estudio, planteó un modelo de regresión lineal múltiple, para medir el rendimiento en términos de ciertas variables como ingreso de los padres, nivel de educación de padres, sector de residencia del estudiante, estrategias de aprendizaje apropiadas, motivación orientada a la tarea, entre otras, de las cuales, variables como nivel de educación de padres, y sector de residencia del estudiante, también se analizó en este estudio, haciendo factor a estas variables, y aunque para este análisis los resultados arrojan que el lugar de residencia del estudiante y el nivel de educación de padres no tienen mayor incidencia en el rendimiento académico del estudiante, en otros estudios como el mencionado (González, C., 2004), estas variables factor, si tienen incidencia significativamente en el rendimiento del estudiante.

En el presente estudio, una variable factor que incide en el rendimiento del estudiante es el estado civil, y si el padre trabaja o no trabaja.

En otro estudio hecho por los autores M. Pérez, G. Orlandoni, F. Aguilar y J. Ramoni en 2015, en el que aplicaron métodos multivariantes para evaluar los resultados obtenidos por los estudiantes que ingresan a la Universidad de Santander en el periodo 2015, los datos arrojaron que estudiantes que vienen de escuelas privadas están mejor preparados que estudiantes que vienen de escuelas

públicas, al igual que en el presente estudio, los resultados arrojaron que considerando la variable factor, colegio particular y colegio fiscal, los estudiantes que vienen de colegios particulares, es decir colegios privados, obtuvieron mejores notas que los estudiantes que vienen colegios fiscales.

Por lo tanto, para lograr una mejora en el rendimiento de los alumnos debemos conocer el estado real del estudiante, tanto los factores internos como los factores externos.

# CAPÍTULO 5

## 5. CONCLUSIONES Y RECOMENDACIONES

### 5.1. Conclusiones

1. El modelo basado en regresión lineal múltiple fué el obtenido mediante el método stepwise mixto AIC, el cual es capaz de explicar el 39.96% de la variabilidad de las variables explicativas, con un p-value de  $2.39 \times 10^{-7}$ , esto quiere decir que las variables independientes consideradas como: Edad, Horas dedicadas al estudio, padres trabajan, estado civil del estudiante, soltero, y padres estudiaron solo primaria; correlacionan de forma significativa con la variable respuesta.

Las variables más significativas que determinan las causas que influyen en el rendimiento académico en matemáticas de los estudiantes que ingresan a un instituto tecnológico superior son:

Horas dedicadas al estudio y estado civil soltero del estudiante, cuyas variables se detallan en la conclusión tres.

2. El perfil del estudiante que ingresa al tecnológico, de acuerdo a los datos obtenidos, está basado en las siguientes características:
  1. La mayor cantidad de estudiantes que ingresan al tecnológico son solteros.
  2. La mayor cantidad de estudiantes que ingresan al tecnológico viven en Guayaquil.
  3. La mayor cantidad de estudiantes que ingresan al tecnológico vienen de colegios fiscales.
  4. La mayor cantidad de estudiantes que ingresan al tecnológico vienen de colegios técnicos, no de especialidad ciencias como el común de los bachilleres, lo cual evidencia el bajo nivel en matemáticas.

5. Los padres de los estudiantes que ingresan al tecnológico, no tienen educación superior, la mayoría tienen educación primaria, secundaria y ninguna.
  6. La mayor cantidad de estudiantes que ingresan al tecnológico trabajan y estudian.
  7. La mayor cantidad de estudiantes que ingresan al tecnológico han dejado de estudiar por más de un mes desde que salieron del colegio, es decir están retomando los estudios.
  8. Los padres de la mayor cantidad de estudiantes que ingresan al tecnológico no trabajan.
3. La propuesta de mejora fue elaborada a partir del análisis de causa-efecto mediante el diagrama de espina de pescado. Lo que se propone es que el alumno le dedique un mayor tiempo y responsabilidad a su formación profesional, dado que, uno de los factores altamente significativos son las horas de estudio, es decir, que se debe crear conciencia en los alumnos acerca del tiempo de estudio que deben dedicar a su formación profesional o al estudio de la asignatura de matemáticas, puesto que el coeficiente que arroja el modelo para esta variable es de 2,4060; o sea, manteniendo fija las otras variables, por cada hora adicional de estudio, la nota del estudiante crecerá en 2,4060 puntos. El estado civil del estudiante también es un factor importante, los resultados demuestran que el estar soltero ayuda positivamente en el rendimiento del alumno, podríamos suponer que esto es debido a que solo está dedicado a estudiar y no a las cuestiones propias de un padre de familia. El coeficiente de esta variable es 2,1570; es decir, si el estudiante es soltero, la nota del estudiante crecerá en 2,1570 puntos. Así mismo, se determinó que el hecho de tener un padre que trabaje es un factor significativo, se puede suponer que esto es debido a que solo se dedicaría a estudiar. El coeficiente para esta variable es 0,8705; o sea, si los padres de un estudiante trabajan, aportará a la nota del estudiante en 0,8705 puntos.

En base a todos estos resultados se propone las siguientes mejoras:

1. Curso de nivelación de conocimiento antes de ingresar al instituto tecnológico para que los estudiantes de distintas especialidades puedan tener el mismo conocimiento en matemáticas.
2. Asignación de becas a los estudiantes con alto rendimiento. Esto ayudará a que el alumno se dedique a estudiar sin tener que trabajar. De esta forma podría aumentar el tiempo de estudio.
3. Asignar tutorías personalizadas para conocer las deficiencias en los temas tratados. Las tutorías deben ser realizadas en horarios fuera de clases.
4. Las clases deben fomentar el autoaprendizaje y la autoevaluación, es decir que el alumno aprenda a conocer sus debilidades y fortalezas para que pueda tomar las acciones correctivas necesarias.

Cabe mencionar que la mejora 1 y la mejora 3 ya han sido aplicados a algunas instituciones tecnológicas de educación superior, dando resultados satisfactorios. Las mismas autoridades de ciertas instituciones indican que en el año 2014, se daba curso de nivelación en Matemáticas a los nuevos estudiantes, y la mayoría lograba terminar la carrera escogida, además que el estudiante entendía mejor las asignaturas relacionadas a la carrera, algo que actualmente no ocurre, como por ejemplo, un estudiante que seguía tecnología en Electrónica, dominaba mejor materias como electrotecnia, instalaciones eléctricas; que son materias que están estrechamente ligadas con las matemáticas, pero que lamentablemente se dejó de dar cursos de nivelación, por cambios en las políticas públicas de educación. Docentes de institutos tecnológicos tienen la apreciación y aseguran de forma innata que los cursos de nivelación si daban resultados.

## 5.2. Recomendaciones

Con las características del perfil de estudiantes, mencionadas en la conclusión dos, podemos deducir que los estudiantes que ingresan al tecnológico, en su gran mayoría pertenecen a un sector informal de la sociedad, por lo que valdría la pena que el Estado otorgue becas a dichos estudiantes, para que financien sus estudios, y reciban cursos de nivelación, para que tengan un sistema de educación de calidad y puedan superarse para el bien común de Ellos y del país.

Se recomienda que el modelo sea evaluado por otros métodos y modelos como, por ejemplo: Análisis de componentes principales, regresión logarítmica o polinomiales.

Se podría mejorar el modelo si se aumenta el número de observaciones para que los coeficientes obtenidos sean más representativos.

Se deberían aumentar más variables como habilidades sociales, organización, autoconcepto, interés por los estudios y además también se deberían evaluar factores inherentes al profesor.

## BIBLIOGRAFÍA

- Albarracin, S., Irigoyen, S., & Papel, G. (2010). El rendimiento académico de los alumnos según su perfil familiar.
- Castaño Castrillón, J. J. (2014). Correlación entre Criterios de Admisión, y Desempeño Académico, en Estudiantes de la Facultad de Medicina de la Universidad de Manizales.
- Chaparro Caso López, A. A., González Barbera, C., & Caso Niebla, J. (2016). Familia y rendimiento académico: configuración de perfiles estudiantiles en secundaria. *Revista electrónica de investigación educativa*, 18(1), 53-68.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*.
- Fahrmeir, L., & Tutz, G. (2013). *Multivariate statistical modelling based on generalized linear models*. Springer Science & Business Media.
- Fundación Iberoamericana para la gestión de calidad (2012). Diagrama causa-efecto. Disponible en: [www.fundibeq.org](http://www.fundibeq.org)
- Gallego, M. R. R., & Sierra, R. O. (2012). Modelo de gestión para la calidad en las prácticas de pedagogía. *Profesorado. Revista de Currículum y Formación de Profesorado*, 16(3), 357-372.
- González Barbera, C. (2004). Factores determinantes del bajo rendimiento académico en educación secundaria. Universidad Complutense de Madrid, Servicio de Publicaciones.
- Gonzales, A. (2015). Selección de variables: Una revisión de métodos existentes. Universidad da Coruña.
- Gujarati, D. N., & Porter, D. C. (2011). *Econometria Básica-5*. Amgh Editora.

- Huang, S., & Fang, N. (2013). Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models. *Computers & Education*, 61, 133-145.
- Izar, L. J., & González, O. J. H. (2004). Las 7 herramientas básicas de la calidad. *Universitaria potosina*.
- Khuri, A. I. (2009). *Linear model methodology*. Chapman and Hall/CRC.
- La Red Martínez, D. L., Acosta, J. C., Cutro, L. A., Uribe, V. E., & Rambo, A. R. (2010). Data warehouse y data mining aplicados al estudio del rendimiento académico y de perfiles de alumnos. In XII Workshop de Investigadores en Ciencias de la Computación.
- Lemos, P. L. (2016). *Herramientas para la mejora de la calidad*. FEMETAL.
- Lozano Díaz, A. (2003). Factores personales, familiares y académicos que afectan al fracaso escolar en la Educación Secundaria. *Electronic journal of research in educational psychology*, 1(1).
- Martínez, D. L. L. R., Karanik, M., Giovanni (Chaparro Caso López, 2016)
- ni, M., & Pinto, N. (2015). Perfiles de rendimiento académico: un modelo basado en minería de datos. *Campus Virtuales*, 4(1), 12-30.
- Mativo, J. M., & Huang, S. (2014, October). Prediction of students' academic performance: Adapt a methodology of predictive modeling for a small sample size. In *Frontiers in Education Conference (FIE)*, 2014 IEEE (pp. 1-3). IEEE.
- Olive, D. J. (2017). Multiple linear regression. In *Linear Regression* (pp. 17-83). Springer, Cham.
- Rosas, M., Chacín, F., García, J., Ascanio, M., & Cobo, M. (2006). Modelos de regresión lineal múltiple en presencia de variables cuantitativas y

cualitativas para predecir el rendimiento estudiantil. Revista de la Facultad de Agronomía, 23(2), 197-214.

Shinde, D. D., Ahirrao, S., & Prasad, R. (2018). Fishbone Diagram: Application to Identify the Root Causes of Student–Staff Problems in Technical Education. *Wireless personal communications*, 100(2), 653-664.

Talsma, K., Schüz, B., y Norris, K. (2018). Calibración de la autoeficacia y el rendimiento académico: Autoeficacia  $\neq$  profecía autocumplida. *Aprendizaje y diferencias individuales*. doi: 10.1016 / j.lindif.2018.11.002.

Yang, S. J., Lu, O. H., Huang, A. Y., Huang, J. C., Ogata, H., & Lin, A. J. (2018). Predicting Students' Academic Performance Using Multiple Linear Regression and Principal Component Analysis. *Journal of Information Processing*, 26, 170-176.

Wilkinson, G. and C. Rogers (1973). Symbolic description of factorial models for the analysis of variance. *Applied Statistics* 22, 392–399

Zurita, G. (2008). *Probabilidad y Estadística Fundamentos y Aplicaciones*. ICM-ESPOL, Guayaquil, Ecuador.

## Apéndices y anexos

Tabla A-1. Coeficientes de correlación entre las variables

	puntuacion	VI2	VI11	VI14	VI17	VI18	VI19	sexo
puntuacion	1.000	0.187	-0.054	0.526	0.129	-0.111	0.156	
VI2	0.187	1.000	-0.258	0.075	0.066	0.048	0.135	
VI11	-0.054	-0.258	1.000	-0.094	-0.193	-0.153	-0.084	
VI14	0.526	0.075	-0.094	1.000	0.067	-0.040	-0.048	
VI17	0.129	0.066	-0.193	0.067	1.000	0.237	0.065	
VI18	-0.111	0.048	-0.153	-0.040	0.237	1.000	0.104	
VI19	0.156	0.135	-0.084	-0.048	0.065	0.104	1.000	
sexo								1.000
residencia								
laboral								
asistencia								
sin_estudiar								
est_civil_padres								
padres_trab								
estudiar_pos								
tics								
casado								
particular								
fiscal								
ciencias								
tecnico								
superior								
secundaria								
primaria								
soltero								

Fuente: Espín, 2018.

**Tabla A-2 Fórmula del Modelo con todas las variables independientes**

```

Call:
lm(formula = puntuacion ~ VI2 + VI11 + VI14 + VI17 + VI18 + VI19 +
    sexo + residencia + laboral + asistencia + sin_estudiar +
    est_civil_padres + padres_trab + estudiar_pos + tics + casado +
    soltero + particular + fiscal + ciencias + tecnico + superior +
    secundaria + primaria, data = datostesis)

Residuals:
    Min       1Q   Median       3Q      Max
-4.846 -1.567 -0.059  1.423  6.824

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -6.175885   6.365523  -0.970   0.336
VI2           0.161552   0.067450   2.395   0.020 *
VI11         -0.004396   0.239599  -0.018   0.985
VI14          2.301620   0.457284   5.033 5.32e-06 ***
VI17          0.100388   0.165374   0.607   0.546
VI18         -0.182778   0.171263  -1.067   0.290
VI19          0.209255   0.176661   1.185   0.241
sexo          1.028318   1.660953   0.619   0.538
residencia    1.020488   1.060043   0.963   0.340
laboral      -0.177639   0.731464  -0.243   0.809
asistencia   -1.943440   1.640382  -1.185   0.241
sin_estudiar -0.189003   0.677652  -0.279   0.781
est_civil_padres -0.231232  0.620096  -0.373   0.711
padres_trab   0.505022   0.700410   0.721   0.474
estudiar_pos  1.311974   1.351834   0.971   0.336
tics          0.640210   0.664680   0.963   0.340
casado        0.699166   2.882896   0.243   0.809
soltero       2.374480   2.867391   0.828   0.411
particular    2.795861   3.219163   0.869   0.389
fiscal        2.990451   3.134135   0.954   0.344
ciencias      0.361521   1.311525   0.276   0.784
tecnico      -0.069419   1.159095  -0.060   0.952
superior     -1.077411   1.912234  -0.563   0.575
secundaria   -1.248028   1.806848  -0.691   0.493
primaria     -2.306902   1.818278  -1.269   0.210
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.496 on 56 degrees of freedom
Multiple R-squared:  0.4765,    Adjusted R-squared:  0.2521
F-statistic: 2.124 on 24 and 56 DF,  p-value: 0.01063

```

Fuente: Espín, 2018.

**Tabla A-3. Método AIC para determinación de factores predictores**

```

Step: AIC=141.36
puntuacion ~ VI2 + VI14 + padres_trab + soltero + primaria

              Df Sum of Sq    RSS    AIC
<none>                400.02 141.36
- padres_trab         1   13.729 413.74 142.09
+ asistencia          1    5.272 394.74 142.29
+ residencia          1    4.949 395.07 142.35
+ VI19                1    4.526 395.49 142.44
+ VI17                1    4.198 395.82 142.51
+ tics                1    4.093 395.92 142.53
+ est_civil_padres    1    2.779 397.24 142.80
+ estudiar_pos        1    2.431 397.58 142.87
+ sexo                1    2.068 397.95 142.94
+ sin_estudiar        1    1.063 398.95 143.15
+ VI18                1    1.003 399.01 143.16
+ laboral             1    0.683 399.33 143.22
+ VI11                1    0.682 399.33 143.22
+ tecnico             1    0.538 399.48 143.25
+ particular          1    0.389 399.63 143.28
+ secundaria          1    0.187 399.83 143.32
+ ciencias            1    0.152 399.86 143.33
+ superior            1    0.050 399.97 143.35
+ fiscal              1    0.030 399.99 143.35
+ casado              1    0.018 400.00 143.36
- soltero             1   28.608 428.62 144.96
- primaria            1   30.025 430.04 145.22
- VI2                 1   50.840 450.86 149.05
- VI14                1  202.114 602.13 172.49

Call:
lm(formula = puntuacion ~ VI2 + VI14 + padres_trab + soltero +
    primaria, data = datostesis)

Coefficients:
(Intercept)          VI2          VI14  padres_trab    soltero    primaria
-2.5567         0.1655         2.4060         0.8705         2.1570        -1.5410

```

*Fuente: Espín, 2018.*

**Tabla A-4. Coeficientes del modelo optimizado por AIC**

```

Call:
lm(formula = puntuacion ~ VI2 + VI14 + padres_trab + soltero +
    primaria, data = datostesis)

Residuals:
    Min       1Q   Median       3Q      Max
-5.0526 -1.7231 -0.0903  1.5768  7.3157

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.55670    1.91089  -1.338  0.18495
VI2          0.16555    0.05362   3.087  0.00283 **
VI14         2.40597    0.39084   6.156 3.38e-08 ***
padres_trab  0.87052    0.54258   1.604  0.11283
soltero      2.15696    0.93133   2.316  0.02329 *
primaria    -1.54102    0.64950  -2.373  0.02023 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.309 on 75 degrees of freedom
Multiple R-squared:  0.3996,    Adjusted R-squared:  0.3595
F-statistic: 9.982 on 5 and 75 DF,  p-value: 2.39e-07
    
```

*Fuente: Espín, 2018.*

**Tabla A-5. Resumen de datos de las variables**

puntuacion	VI2	VI11	VI14	VI17	VI18
Min. : 1.000	Min. :17.00	Min. :2.000	Min. :0.000	Min. : 2.000	Min. : 1.00
1st Qu.: 5.000	1st Qu.:18.00	1st Qu.:2.000	1st Qu.:1.000	1st Qu.: 5.000	1st Qu.: 6.00
Median : 7.000	Median :20.00	Median :4.000	Median :2.000	Median : 7.000	Median : 8.00
Mean : 6.815	Mean :22.38	Mean :3.852	Mean :1.519	Mean : 6.506	Mean : 7.37
3rd Qu.: 9.000	3rd Qu.:23.00	3rd Qu.:5.000	3rd Qu.:2.000	3rd Qu.: 8.000	3rd Qu.: 9.00
Max. :14.000	Max. :50.00	Max. :7.000	Max. :3.000	Max. :10.000	Max. :10.00
VI19					
Min. : 1.00					
1st Qu.: 5.00					
Median : 6.00					
Mean : 6.16					
3rd Qu.: 7.00					
Max. :10.00					

*Fuente: Espín, 2018.*