Estudio estadístico sobre la influencia del uso del celular en el rendimiento académico de matemáticas de los estudiantes del Colegio Venecia IED

Carlos Andrés Pedraza Huertas

Andrés Felipe Vargas Ortegón

Universidad Pedagógica Nacional

Facultad de Ciencia y Tecnología

Departamento de Matemáticas

Licenciatura en Matemáticas

Bogotá D.C., 2025

Estudio estadístico sobre la influencia del uso del celular en el rendimiento académico de matemáticas de los estudiantes del Colegio Venecia IED

Trabajo de grado como requisito parcial para obtener el título de Licenciados en Matemáticas

Andrés Felipe Vargas Ortegón

Código: 2019240055

Carlos Andrés Pedraza Huertas

Código: 2019240040

Asesor

César Guillermo Rendón Mayorga

Profesor del Departamento de Matemáticas

Universidad Pedagógica Nacional

Facultad de Ciencia y Tecnología

Departamento de Matemáticas

Licenciatura en Matemáticas

Bogotá D.C., 2025

Quiero dedicar este logro a mi querida madre Hilda, a mi padre Orlando, a mis hermanos Mario, Luis y Zulma, y a cada una de las personas que me acompañaron en este proceso.

-Andres Vargas-

Dedico este trabajo de grado a mi familia, en especial a mi madre Lilia, a mi padre Alberto y a mi hermana Paola, que con su apoyo me han motivado a lograr grandes metas.

-Carlos Pedraza-

Agradecimientos

Agradezco a la Universidad Pedagógica por la formación recibida a lo largo de mi proceso académico, y al profesor César por haber aceptado ser nuestro asesor. Su dedicación, el tiempo brindado más allá de sus obligaciones, y el compromiso con la realización de este trabajo de grado fueron fundamentales para su desarrollo. También extiendo mi agradecimiento a cada uno de los profesores, compañeros y amigos que, con sus enseñanzas y apoyo, han contribuido significativamente a mi crecimiento personal y profesional.

De manera muy especial, agradezco a mis padres por todo el tiempo, el trabajo, el amor y el esfuerzo que han dedicado para que cada día pueda ser mejor y logre cumplir mis sueños. Su apoyo incondicional siempre ha sido mi más grande impulso.

-Andres Vargas-

Expreso mi gratitud a mi familia por su apoyo durante mi carrera, con sus palabras de ánimo he alcanzado metas que parecían muy difíciles. También doy las gracias a la Universidad Pedagógica y a los profesores de la institución por enseñarme sobre las matemáticas y sobre la vida.

Adicionalmente, aprecio el apoyo y la guía dada por el asesor del trabajo de grado, el profesor Cesar Rendón, cuyas orientaciones y conocimientos fueron fundamentales en la creación de la presente tesis.

Finalmente quiero agradecer a la profesora Yeimmy Riaño, y los compañeros que nos ayudaron con los aspectos logísticos, relacionados con la obtención de los datos, con su ayuda se obtuvo una valiosa información que permitió la creación de esta tesis.

-Carlos Pedraza-

Tabla de Contenido

Índice de Figuras

Índice de Tablas

1 CAPÍTULO I. PRELIMINARES

En este capítulo se presenta una introducción al trabajo de grado, con la cual se busca contextualizar la problemática abordada, así como justificar su pertinencia en el contexto educativo actual. Asimismo, se exponen el objetivo general y los objetivos específicos que orientan el desarrollo del estudio, los cuales permiten delimitar el alcance de la investigación y guiar el análisis de estadística multivariada que realizamos.

1.1 Introducción

El uso del celular ha aumentado considerablemente en los últimos años, convirtiéndose en un dispositivo indispensable para personas de todas las edades debido a su versatilidad. Su capacidad para facilitar múltiples actividades como chatear, jugar, ver películas, estudiar, entre otras, lo hace altamente atractivo y funcional. Según el Departamento Administrativo Nacional de Estadística [DANE] (DANE, 2023), el 90,4 % de las personas mayores de cinco años utilizan celular en Colombia, y en la ciudad de Bogotá la cifra asciende al 93,5 %. Estos datos, sumados a lo que se observa cotidianamente, permiten afirmar que actualmente casi todas las personas cuentan con un teléfono móvil.

En el caso de los niños y jóvenes, la familiaridad con estos dispositivos es aún mayor. Tal como lo señala Prensky (2001), los estudiantes actuales son considerados “nativos digitales”, ya que han crecido rodeados de tecnología, lo que les otorga una gran agilidad en su uso. Esto mismo ocurre con el celular, el cual está presente desde las primeras etapas de vida de los estudiantes y se convierte en un elemento central en su cotidianidad. Esta presencia constante del celular transforma sus dinámicas sociales, incluyendo el ámbito educativo, en el cual se ha estudiado su influencia en el rendimiento académico.

Por otro lado, el rendimiento escolar de los estudiantes en la asignatura de matemáticas ha sido motivo de preocupación constante en el ámbito educativo, dada su relevancia en el desarrollo de competencias lógico-matemáticas y en la formación integral de los estudiantes. En este sentido, resulta pertinente analizar en el contexto actual si el uso frecuente y prolongado de los teléfonos celulares puede estar asociado a un bajo desempeño en matemáticas, teniendo en cuenta que el tiempo invertido en redes sociales, juegos y otras aplicaciones podría desplazar las horas dedicadas al estudio y afectar procesos cognitivos clave como la concentración, la memoria y la resolución de problemas.

El presente trabajo tiene como propósito estudiar la relación estadística entre el tiempo de uso del celular y el rendimiento académico en matemáticas de estudiantes de grados octavo y décimo del Colegio Venecia IED. A partir de un enfoque cuantitativo, se recopilaron datos sobre el uso del celular mediante una aplicación y se contrastaron con las calificaciones finales de matemáticas, buscando establecer si existe alguna correlación importante entre estas variables. Adicionalmente, se consideraron otras variables, como las actitudes hacia las matemáticas, dado que los aspectos afectivos y motivacionales también influyen en el aprendizaje y en el rendimiento académico.

En cuanto a la estructura de este documento, el trabajo se organiza en cinco capítulos. En el primer capítulo se presentan los objetivos generales y específicos que orientan el estudio. El segundo capítulo corresponde al marco teórico, en el cual se realiza una revisión conceptual sobre el rendimiento académico, las actitudes hacia las matemáticas y el impacto del uso del celular, además de un marco matemático que expone las pruebas estadísticas y los métodos de análisis que se implementaron durante el estudio.

En el tercer capítulo se describe la metodología empleada, incluyendo la caracterización del estudio, la descripción de las variables, el tipo de muestreo, las técnicas e instrumentos de recolección de datos y las fases que guiaron el desarrollo del trabajo de grado. El cuarto capítulo presenta la implementación del modelo estadístico, el análisis de los resultados y la verificación de los supuestos del modelo. Finalmente, en el quinto capítulo se exponen las conclusiones derivadas del estudio, las cuales recogen las principales reflexiones y aportes obtenidos a lo largo del trabajo de grado.

1.2 Objetivo general

Indagar por posibles relaciones estadísticas entre el rendimiento académico en la asignatura de matemáticas y el uso del celular en estudiantes de grado octavo y grado décimo del Colegio Venecia IED, con el objetivo de evaluar el impacto del celular mediante una técnica de estadística multivariada.

1.3 Objetivos específicos

- Configurar un marco de referencia sobre el impacto que tiene el uso del celular en el rendimiento académico escolar a partir de una revisión documental.
- Recolectar, depurar y organizar datos asociados al tiempo de uso del celular por parte de los estudiantes y a su rendimiento académico en la clase de matemáticas utilizando software estadístico (p. ej. R, Excel).
- Determinar, estudiar e implementar la técnica estadística más apropiada para el análisis.
- Identificar las relaciones estadísticas entre el desempeño en matemáticas, el uso del celular y otras posibles variables estadísticas a través de la implementación de una técnica estadística adecuada.

2 CAPÍTULO II. MARCO TEÓRICO

En el presente capítulo se describen los conceptos matemáticos abordados durante el desarrollo del presente trabajo de grado, entre estos están los coeficientes de correlación de Pearson y Spearman, las pruebas de bondad de ajuste como Shapiro-Wilk y Kolmogórov-Smirnov para determinar la normalidad de las variables, una descripción del modelo de regresión, sus estadísticos y las pruebas que verifican la significancia del modelo. Para facilitar la comprensión de los conceptos abordados en este trabajo de grado, se sugieren algunos textos introductorios como Estadística para Dummies (Rumsey, 2013), Introducción a la Probabilidad y la Estadística (Mendenhall et al., 2009) y Estadística (Triola, 2009).

Adicionalmente, en este capítulo se encuentran apartados en los que se presentan los resultados de una revisión documental de algunos conceptos importantes para el desarrollo de este estudio. Los conceptos que son revisados son el rendimiento académico, el impacto del uso del celular y las actitudes de los estudiantes hacia las matemáticas.

2.1 Coeficiente de correlación

Como lo menciona Lahura (2003) “El coeficiente de correlación es un estadístico que proporciona información sobre la relación lineal existente entre dos variables cualesquiera. Básicamente, esta información se refiere a dos características de la relación lineal: la dirección o sentido y la cercanía o fuerza” (p. 6). El autor enfatiza que este coeficiente resulta útil para medir relaciones lineales, pero no describe adecuadamente aquellas de naturaleza no lineal. Si se utiliza en circunstancias donde la relación no sea estrictamente lineal, el resultado únicamente reflejaría la ausencia de relación lineal, sin descartar la posibilidad de que exista otro tipo de asociación.

Los coeficientes de correlación solo miden relaciones lineales de variables continuas con distribución normal bivariada (Pearson) o monótonas con variables ordinales organizadas en rangos o jerarquías (Spearman), las cuales tienden a cambiar al mismo tiempo, pero no necesariamente a un ritmo constante. (Mendivelso y Rodríguez, 2021, p. 42)

Si bien existen diversos coeficientes de correlación, los mencionados por Mendivelso y Rodríguez (2021), el coeficiente de Pearson y el de Spearman son los más empleados en la práctica, dadas las particularidades de sus supuestos y ámbitos de aplicación. A continuación, se describe cada uno en detalle.

Coeficiente de correlación de Pearson

El coeficiente de correlación de Pearson es una medida estadística que cuantifica la fuerza[1] y la dirección de la relación lineal entre dos variables continuas. Es decir, nos indica qué tan estrechamente relacionadas están dos variables y si esa relación es directa (a mayor valor de una, mayor valor de la otra) o inversamente lineal (a mayor valor de una, menor valor de la otra). En nuestro estudio, y como se verá más adelante, utilizamos el coeficiente de Pearson para evaluar si existe una relación lineal entre el tiempo de uso del celular y el rendimiento académico en matemáticas, y así determinar si a mayor tiempo dedicado al celular, menor es el desempeño en esta asignatura.

Según Hernández et al. (2018), el coeficiente de correlación de Pearson ha demostrado ser una herramienta estadística invaluable en una amplia gama de disciplinas. Desde la ingeniería, donde se emplea para evaluar fenómenos como la deformación del viento y la eficiencia de turbinas; hasta la medicina, donde se utiliza para explorar relaciones entre enfermedades, tratamientos y factores de riesgo. Esta medida ha sido fundamental para comprender y cuantificar

[1] Valor numérico el cual establece que tan bien se ajusta el modelo lineal al conjunto de datos

las relaciones entre variables. Su aplicación se extiende también a campos como la psicología, la ciencia de datos, la economía, las finanzas y muchos más.

En psicología, por ejemplo, se utiliza para analizar propiedades psicométricas de escalas y para investigar la relación entre variables como el estrés y el agotamiento laboral. En la ciencia de datos es una herramienta esencial para desarrollar modelos predictivos y analizar grandes conjuntos de datos. La economía y las finanzas se benefician del coeficiente de Pearson para estudiar la relación entre variables económicas y financieras, como la inversión extranjera y el mercado bursátil. Estos son tan solo algunos ejemplos de la gran polivalencia que se le ha dado al uso de este método estadístico. (Hernández et al., 2018)

> Lo interesante es que, a pesar de que el uso del coeficiente de correlación de Pearson es extensísimo, también lo son las incorrecciones y omisiones al momento de emplearlo, las imprecisiones cuando de verificar sus supuestos se trata, o incluso, la confusión y desinformación a la hora de interpretar los resultados. (Hernández et al., 2018, p. 589)

Debido a su gran acogida y versatilidad en diversos campos, son muchos los estudiantes, profesores, investigadores e incluso profesionales de este campo, que incurren en errores en su uso. Algunos de los errores más comunes son: la falta de claridad en nociones de términos como regresión y correlación, asumir que no existe alguna relación entre las variables si la correlación es nula, asumir erróneamente que la correlación indica causalidad, entre otros.

Al estudiar dos o más variables de forma conjunta, lo que se busca es determinar si el comportamiento de una se modifica de manera sistemática cuando varía la otra, o bien establecer que no existe tal patrón. Desde el punto de vista estadístico, este fenómeno se mide a través de la covarianza.

La covarianza entre dos variables aleatorias X y Y se define mediante:

$$\mathrm{Cov}\ (\mathrm{X}, \mathrm{Y}) \equiv \sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] \tag{1}$$

donde E es el valor esperado, μ_X y μ_Y son las medias poblaciones de X y Y, respectivamente; y las expresiones Cov (X, Y) y σ_{XY} son equivalentes. Si se tiene que Y es igual a X, en la ecuación (1) se obtiene:

$$\mathrm{Cov}\ (\mathrm{X}, \mathrm{X}) \equiv \sigma_{Xx} = E[(X - \mu_X)(X - \mu_x)] = E(X - \mu_X)^2 \tag{2}$$

A partir de esto, se concluye que la covarianza de una variable consigo misma es su varianza. La expresión (1) se puede reescribir de distintas formas dependiendo de si X y Y son continuas o discretas. Adicionalmente, las ecuaciones (3) y (4) aplican en el caso en que X y Y no tienen la misma probabilidad de ocurrir:

$$\mathrm{Cov}\ (\mathrm{X}, \mathrm{Y}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (X - \mu_X)(Y - \mu_Y) f_{XY}\, dxdy \tag{3}$$

$$\mathrm{Cov}\ (\mathrm{X}, \mathrm{Y}) = \sum_{\boldsymbol{x}} \sum_{\boldsymbol{y}} (X - \mu_X)(Y - \mu_Y) f_{XY}, \tag{4}$$

Donde f_{XY} es la función de densidad conjunta; la ecuación (3) funciona para variables continuas, mientras que la (4) lo hace para las discretas. Lo fundamental de este hecho radica en que, cuando todos los valores de X y Y se ubican por encima o por debajo de sus medias poblacionales de manera simultánea en el producto $(X - \mu_X)(Y - \mu_Y)$, la covarianza resulta positiva. En cambio, si todos los valores de X se sitúan por encima o por debajo de μ_X mientras los de Y se comportan de forma opuesta, la covarianza se torna negativa. Asimismo, la intensidad (fuerza) del resultado dependerá de que tan grandes sean las diferencias en $(X-\mu_X)(Y-\mu_Y)$.

Dado que X y Y pueden encontrarse en unidades de medición distintas, el valor resultante de la covarianza puede carecer de interpretación. Precisamente, esta limitación impulsó el desarrollo del coeficiente de correlación de Pearson, que no es más que la versión estandarizada de la covarianza.

Se define ρ_{XY} como el coeficiente de correlación lineal de Pearson entre dos variables aleatorias numéricas X y Y, tal que:

$$\rho_{XY} = \frac{\mathrm{Cov\ (X, Y)}}{\sqrt{Var(X)}\sqrt{Var(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}, \tag{5}$$

donde σ_X y σ_Y representan las desviaciones estándares de X y Y; mientras que $Var(X)$ u $Var(Y)$ son las varianzas de X y Y respectivamente. En la ecuación (5) las unidades del numerador se cancelarán con las del denominador, lo que conlleva a que sea adimensional y facilite su interpretación y uso. Enseguida se presenta la expresión para casos muestrales.

$$r_{XY} = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{[\sum(X-\bar{X})^2\sum(Y-\bar{Y})^2]^{1/2}} = \frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}}, \tag{6}$$

En la expresión (6), las barras indican las medias muéstrales de X y Y, mientras que S_{XX}, S_{YY} y S_{XY} corresponden a las sumas de cuadrados corregidas para X y Y y el producto cruzado XY.

Para ilustrar diferentes escenarios de correlación entre dos variables mediante el coeficiente de correlación de Pearson se presenta la Ilustración 1.

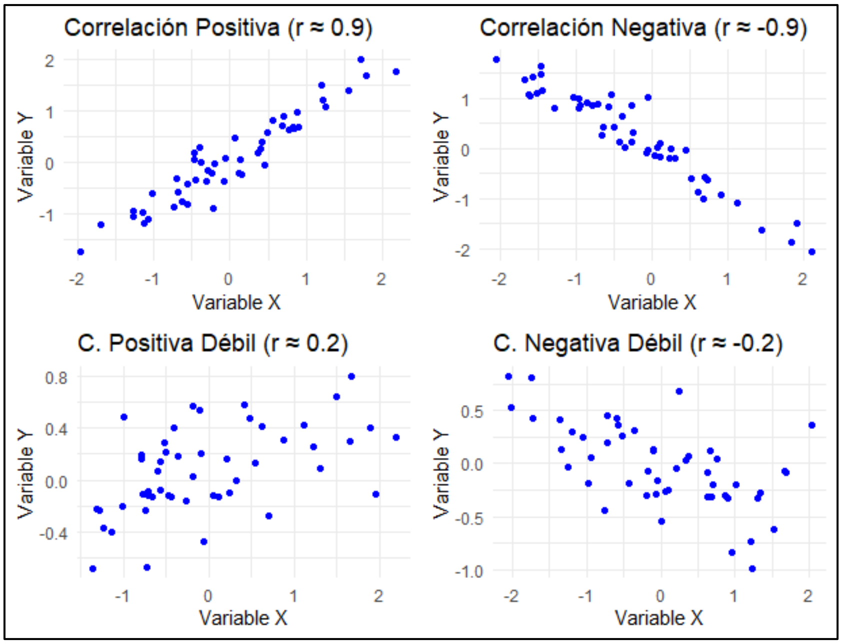

Ilustración 1- Coeficiente de correlación de Pearson

En la parte superior izquierda de la Ilustración I, se observa una correlación positiva fuerte ($r_{XY} \approx 0.9$), donde los puntos siguen un patrón ascendente, indicando que a medida que aumenta la variable X, también lo hace Y. En contraste, la parte superior derecha muestra una correlación negativa fuerte ($r_{XY} \approx -0.9$), donde el patrón es descendente, reflejando que cuando X aumenta, Y disminuye. En la parte inferior, se presentan ejemplos de correlaciones débiles: a la izquierda, una correlación positiva débil ($r \approx 0.2$), con un patrón poco definido, pero ligeramente ascendente, y a la derecha, una correlación negativa débil ($r \approx -0.2$), donde el patrón es igualmente disperso, pero ligeramente descendente.

El coeficiente de correlación de Pearson posee una serie de características distintivas que lo convierten en una herramienta estadística fundamental en diversas disciplinas. Entre sus propiedades más relevantes destacan:

I. Adimensionalidad: El valor del coeficiente no depende de las unidades de medida de las variables, lo que permite comparar resultados entre diferentes escalas. Esta característica facilita la interpretación y generalización de los hallazgos.

II. Rango definido: El coeficiente se encuentra limitado a un rango de valores entre -1 y 1, donde los extremos representan una correlación perfecta (positiva o negativa) y un valor cercano a cero indica una ausencia de relación lineal.

La Tabla 1 presenta una de las interpretaciones más utilizadas del coeficiente de correlación de Pearson, propuesta por Cohen. En esta, los valores de r se categorizan en rangos que van desde correlación nula ($0.00 \leq |r_{XY}| < 0.10$) hasta correlación fuerte ($0.50 \leq |r_{XY}| \leq 1.00$). Sin embargo, es importante señalar que esta no es una norma universal y su aplicabilidad puede variar significativamente según el área de estudio o contexto en el que se utilice. Por ejemplo, en disciplinas como psicología o ciencias sociales, valores de r considerados débiles podrían ser significativos, mientras que en ingeniería o física podrían requerirse valores más altos para interpretaciones similares. Por lo tanto, se debe tener cuidado al interpretar r, considerando las particularidades de cada campo.

Tabla 1 – Interpretación del coeficiente de correlación

Rango de valores	**Interpretación**
$0.00 \leq \lvert r_{XY}\rvert < 0.10$	Correlación nula
$0.10 \leq \lvert r_{XY}\rvert < 0.30$	Correlación débil
$0.30 \leq \lvert r_{XY}\rvert < 0.50$	Correlación moderada
$0.50 \leq \lvert r_{XY}\rvert \leq 1.00$	Correlación fuerte

III. Relación lineal: Es crucial comprender que el coeficiente de Pearson mide únicamente la fuerza y dirección de una relación lineal. Un valor nulo no implica necesariamente independencia entre las variables, sino simplemente la ausencia de una asociación lineal. Podrían existir otros tipos de relaciones no lineales entre las variables.

IV. Simetría: El valor del coeficiente no se altera al intercambiar las variables $r_{XY} = r_{YX}$, lo que significa que ninguna de las variables puede considerarse causalmente dependiente de la otra.

V. Invariancia ante transformaciones lineales: El coeficiente de Pearson es insensible a cambios en el origen o la escala de las variables, lo que lo hace robusto ante transformaciones lineales como sumar o multiplicar por una constante.

Supuestos del coeficiente de correlación de Pearson:

Como se mencionó anteriormente es común usar de manera incorrecta este instrumento, para hacerlo de manera adecuada es esencial el cumplimiento de los siguientes supuestos:

I. Nivel de medición: Las dos variables deben ser de intervalo[2] o de razón[3], aunque no es necesario que ambas tengan el mismo nivel de medición (no puede ser dicotómica).

II. Datos Pareados: para que el cálculo de esta medida pueda realizarse, se necesitará que los casos en cuestión tengan la misma cantidad de datos en cada variable. Si faltan datos en una u otra, estos registros no se podrán incluir en los cálculos.

III. Normalidad bivariada: lo común es que se pruebe la normalidad marginal de X y Y; sin embargo, para el uso del coeficiente de correlación de Pearson se hace necesario comprobar

[2] Las variables tienen escala de intervalo si pueden usarse para jerarquizar y además los intervalos (las diferencias) entre observaciones se expresan en términos de una unidad de medición fija.

[3] Las variables tienen escala de razón, si los valores, tienen las propiedades de los datos de intervalo y el cociente (o razón) entre dos medidas tiene sentido, además de que los datos de razón tienen que ser numéricos. Cabe mencionar que esta escala requiere que se tenga el valor cero para indicar que en este punto no existe la variable.

que la función de probabilidad conjunta sea normal, en otras palabras, validar el supuesto de normalidad bivariada.

IV. Ausencia de datos atípicos a nivel bivariado: Como pasa con la normalidad, los datos atípicos deben identificarse a nivel bivariado, resulta un poco más complicado que con una sola variable. Una alternativa ampliamente empleada para identificar valores atípicos en datos con múltiples variables consiste en calcular la distancia de Mahalanobis (García y Uribe, 2013).

V. Linealidad: Aunque es lo que se pretende deducir al usar este coeficiente, previamente se puede observar este supuesto mediante diagramas de dispersión para hacerse una idea informal.

VI. Independencia de observaciones: establece que cada medición debe provenir de sujetos o unidades de análisis autónomas, sin que exista influencia o dependencia entre ellas.

Según Hernández et al. (2018), se debe tener presente la importancia de no limitarse a verificar la normalidad de cada variable por separado. Rencher (2002), en su obra *Methods of Multivariate Analysis*, explica que al trabajar con múltiples variables, evaluar únicamente la normalidad univariada no es suficiente, ya que las correlaciones entre las variables pueden influir y la normalidad marginal no asegura la normalidad conjunta. Esto implica que los datos pueden no cumplir con la normalidad bivariada, incluso si se confirma la normalidad univariada. Del mismo modo, como señala Timm, si una de las variables no sigue una distribución normal, el vector multivariado también será considerado no normal.

Coeficiente de correlación lineal de Spearman

Como señala Canavos (1998): "En algunos procedimientos inferenciales se necesita especificar la distribución de la población de interés, para realizar estimaciones con respecto a los parámetros (media, varianza, desviación estándar, etc.) de la población, por ello este tipo de inferencias se denominan métodos paramétricos" (p. 572). Sin embargo, existen procedimientos inferenciales que no se encuentran sujetos a la forma de la distribución de la población de interés. Estos procedimientos reciben el nombre de métodos no paramétricos o métodos independientes de la distribución. En este sentido estos métodos requieren pocas suposiciones, y en ocasiones son más fáciles de aplicar que los métodos paramétricos, adicionalmente pueden aplicarse a las observaciones que se definen en una escala de intervalo y, en algunos casos, sobre escalas nominales.

Para determinar si es preferible un método paramétrico o uno no paramétrico, se tiene en cuenta la escala de medición empleada para generar los datos, como menciona Anderson (2008) todos los datos son generados por una de las cuatro escalas de medición: nominal, ordinal, de intervalo o de razón, las cuales se pueden definir como:

1. Escala nominal. Una escala de medición es nominal si los datos son etiquetas o categorías que se usan para definir un atributo de un elemento. Los datos nominales pueden ser numéricos o no numéricos.
2. Escala ordinal. Una escala de medición es ordinal si los datos pueden usarse para jerarquizar u ordenar las observaciones. Los datos en escala ordinal pueden ser numéricos o no numéricos.

3. Escala de intervalo. Una escala de medición es de intervalo si los datos tienen las propiedades de los datos de escala ordinal y los intervalos entre observaciones se expresan en términos de una unidad de medición fija. Los datos de intervalo tienen que ser numéricos.
4. Escala de razón. Una escala de medición es de razón si los datos tienen las propiedades de los datos de la escala intervalo y el cociente (o razón) entre dos medidas tiene sentido. Los datos de razón tienen que ser numéricos.

Con datos nominales y ordinales no es apropiado calcular medias, varianzas o desviaciones estándar; por lo tanto, no se emplean los métodos paramétricos, y se hace necesario el uso de métodos no paramétricos. En este sentido para que un método estadístico se clasifique como no paramétrico, este debe satisfacer, por lo menos, una de las siguientes condiciones:

a) Ser un método que pueda ser usado con datos nominales.
b) Ser un método que pueda ser usado con datos ordinales.
c) Ser un método que pueda ser usado con datos de escala intervalo o razón cuando no sea posible hacer suposiciones acerca de la forma de la distribución de la población.

Una medida no paramétrica de asociación cuando se emplean rangos y se aplica entre dos variables para el caso de datos ordinales, se define a partir del coeficiente de correlación de Pearson entre los rangos $R_i(x)$ y $R_i(y)$, en la ecuación (6) se reemplaza X por $R_i(x)$ y Y por $R_i(y)$:

$$r_s = \frac{\sum_{i=1}^{n}[(R_i(x) - \bar{R}(x))(R_i(y) - \bar{R}(y))]}{[(\sum_{i=1}^{n} R_i(x) - \bar{R}(x))^2(\sum_{i=1}^{n} R_i(y) - \bar{R}(y))^2]^{1/2}}$$

Si se tiene que $R_i(x) = R_i(y)$, esto es igual a:

$$\frac{\sum_{i=1}^{n} \frac{R_i(x)}{n}}{n} = \frac{1 + 2 + 3 + 4 + \cdots + n}{n} = \frac{n+1}{2}$$

Además, se tiene que por la sumatoria de términos al cuadrado,

$$\sum_{i=1}^{n} {R_i}^2(x) = \sum_{i=1}^{n} {R_i}^2(y) = \frac{n(n+1)(2n+1)}{6}$$

$$\sum_{i=1}^{n} [(R_i(x) - \bar{R}(x)]^2 = \sum_{i=1}^{n} {R_i}^2(x) - n\bar{R}^2(x) = \frac{n(n+1)(2n+1)}{6} - n\left(\frac{n+1}{2}\right)^2 = \frac{n^3+n}{12}$$

De la misma manera,

$$\sum_{i=1}^{n} [(R_i(y) - \bar{R}(y)]^2 = \frac{n^3+n}{12}$$

Se define la diferencia de rangos como $d_i = R_i(x) - R_i(y)$, luego:

$$\sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} [R_i(x) - R_i(y)]^2 = \sum_{i=1}^{n} \left([(R_i(x) - \bar{R}(x)] - [(R_i(y) - \bar{R}(y)]\right)^2$$

Desarrollando el producto,

$$= \sum_{i=1}^{n} [(R_i(x) - \bar{R}(x)]^2 + \sum_{i=1}^{n} [(R_i(y) - \bar{R}(y)]^2 - 2\sum_{i=1}^{n} [(R_i(x) - \bar{R}(x)]\,[(R_i(y) - \bar{R}(y)]$$

$$= \frac{\sum_{i=1}^{n}[(R_i(x) - \bar{R}(x)]^2 + \sum_{i=1}^{n}[(R_i(y) - \bar{R}(y)]^2 - \sum_{i=1}^{n} d_i^2}{2} = \frac{n^3+n}{12} - \sum_{i=1}^{n} \frac{d_i^2}{2}$$

Finalmente, remplazando las expresiones obtenidas en la expresión del coeficiente de correlación inicial, se tiene que,

$$r_s = \frac{\frac{n^3+n}{12} - \sum_{i=1}^{n} \frac{d_i^2}{2}}{\left[\left(\frac{n^3+n}{12}\right)\left(\frac{n^3+n}{12}\right)\right]^{1/2}} = 1 - 6\frac{\sum_{i=1}^{n} d_i^2}{n^3 - n}$$

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \tag{7}$$

En donde:

n es el número de elementos o individuos a los que se les va a asignar un rango.

$R_i(x)$ es el rango de la observación i respecto de la variable X.

$R_i(y)$, es el rango de la observación i respecto de la variable Y.

$\bar{R}(y)$ y $\bar{R}(x)$, son las medias de los rangos y es la misma para cada variable.

Con este procedimiento se logra deducir la fórmula (7) y a su vez se evidencia que el coeficiente de correlación de rangos de Spearman es igual al coeficiente de correlación de rangos de Pearson aplicado a los rangos de los valores de las variables analizadas.

Cabe mencionar que al igual que el coeficiente de correlación de Pearson r, el coeficiente de correlación de rangos r_s se define en el intervalo $-1 \leq r_s \leq 1$ y mide el grado de asociación lineal entre los rangos de X y Y. Adicionalmente al emplear los rangos, r_s mide la tendencia de las variables X y Y de relacionarse en forma monótona, esto quiere decir que si el valor de r_s es cercano a 1 o -1, se puede estimar una asociación monótona creciente o decreciente respectivamente.

El coeficiente de correlación de Spearman se interpreta tomando en cuenta el objetivo planteado en la investigación y la relevancia de estas relaciones para el fenómeno que se estudia. No debe basarse únicamente en la cifra matemática resultante, sino apoyarse también en la evidencia científica previa para no confundir la casualidad (Martínez et al., 2009). De igual manera que el r de Pearson, el coeficiente de Spearman cuantifica la intensidad de la relación lineal entre dos variables sin implicar causa-efecto. Por ello, la interpretación depende, en gran medida, de los detalles de la investigación y del conocimiento del tema.

2.2 Pruebas de bondad de ajuste

Las pruebas de bondad de ajuste son herramientas estadísticas que se utilizan para determinar si los datos de una muestra pueden considerarse provenientes de una distribución o modelo de probabilidad específico como la normal, binomial, de Poisson, exponencial, entre otros. Estas pruebas resultan fundamentales para identificar la distribución que mejor describe los datos y, en consecuencia, decidir si se pueden aplicar pruebas paramétricas o no paramétricas en los análisis estadísticos. En esencia, permiten verificar la adecuación de un modelo probabilístico a los datos observados.

Se dispone de diferentes técnicas para evaluar la normalidad de un conjunto de datos, algunas pueden ser gráficas como la creación de histogramas junto a su curva de densidad, de esta manera se tiene una visualizacion de la distribucion de los valores de la variable; sin embargo esta técnica gráfica puede ser subjetiva, por ello se suele utilizar un contraste de hipotesis el cual se fundamenta en un analisis numerico, que permite establecer si la distribucion de los valores de la variable se aproxima a una distribucion.

Algunas técnicas de contraste de hipótesis para verificar que los datos siguen una distribucion normal son la de Shapiro-Wilk y la de Kolmogorov-Smirnov. A continuación, se presentan las pruebas de hipótesis, sus estadísticos y la manera en la que se implementan en un conjunto de datos utilizando el lenguaje de programación R.

Test Shapiro-Wilk

Este test se utiliza para verificar la normalidad en muestras con menos de 50 observaciones y, en el caso de muestras grandes, ofrece resultados equivalentes al test de Kolmogórov-Smirnov

(Novales, 2010). La prueba de Shapiro-Wilk (SW), como prueba de normalidad, se introduce considerando que el gráfico de probabilidad normal, empleado para observar la adecuación de los datos a la distribución normal, se asemeja a una regresión lineal, donde la línea diagonal representa el ajuste perfecto. Sin embargo, esta línea se asimila a los residuos del modelo de regresión. Al analizar la magnitud de dicha variación mediante un análisis de varianza se puede examinar el nivel de ajuste de manera más precisa (Carmona y Carrión, 2015).

La prueba de Shapiro-Wilk es una prueba que contrasta las siguientes hipótesis:

H_0: La muestra sigue una distribucion normal

H_1: La muestra no sigue una distribución normal

Para rechazar o no la hipótesis nula se puede contrastar el p-valor obtenido de la prueba para saber si es mayor que el nivel de significancia, y si es así entonces no se rechaza H_0.

Este p-valor se calcula con base en el estadístico de prueba W, definido por Shapiro y Wilk (1965):

$$w = \frac{\left(\sum_{i=1}^{n} a_i x_i\right)^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2} \tag{8}$$

Tal que:

x_i: Es la i-esima estadística de orden[4].

$\overline{x}$: Es la Media muestral

[4] En el caso de variables aleatorias, si se tiene una secuencia de n variables aleatorias, $x_1, \dots, x_n$, los estadísticos de orden $X_{(1)}, \dots, X_{(n)}$ también son variables aleatorias, que se definen ordenando las variables en orden ascendente, es decir: $X_{(1)} = \min\{x_1, \dots, x_n\} < X_{(1)} < \cdots < X_{(n-1)} < X_{(n)} = \max\{x_1, \dots, x_n\}$

$$a_i = (a_1, \dots, a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{\frac{1}{2}}}$$

$m = (m_1, \dots, m_n)^T$: Son los valores esperados de las estadísticas de orden.

V: Es la matriz de covarianza de las estadísticas de orden.

Cabe mencionar que en esta prueba la decisión de no rechazar la hipótesis nula H_0 no se toma directamente con base en el valor del estadístico W, ya que este solamente da una idea del ajuste de los datos a una distribución normal.

Un valor W alto (cercano a 1) genera un p-valor alto, lo que sugiere que los datos son normales, por otro lado, cuando el valor de W es bajo (lejano a 1) se produce un p-valor bajo, lo que indica que los datos pueden no seguir una distribución normal.

Esto se logra evidenciar mejor al realizar la prueba a algunos datos utilizando lenguajes de programación como R. Enseguida se presenta un ejemplo trivial para ilustrar el funcionamiento del test en un entorno computacional.

Sean los datos:

```
> data <- c(1.1, 1.2, 1.3, 1.3, 1.3, 1.4, 1.5, 1.8, 1.8, 1.8, 1.8, 2.0, 2.0, 2.1, 2.1, 2.1, 2.3, 2.2)
```

Ilustración 2- Datos de distribución normal

Al aplicar la prueba se obtiene lo siguiente:

```
> shapiro.test(data)

        Shapiro-Wilk normality test

data:  data
W = 0.91662, p-value = 0.1126
```

Ilustración 3- Shapiro.test con datos normales

Se observa que el valor W es de 0.91662 y el p-valor es 0.1126, el cual es mayor a un nivel de significancia del 0.05, por lo tanto, no hay evidencia estadística para rechazar la normalidad de los datos.

Por otra parte, si los datos son:

```
> data <- c(1.1, 1.2, 1.3, 1.3, 1.3, 1.4, 1.5, 1.8, 1.8, 1.8, 1.8, 2.0, 2.0, 2.1, 2.1, 2.1, 2.3, 2.2, 4)
```

Ilustración 4- Datos de distribución no normal

Se obtiene que:

```
        Shapiro-Wilk normality test

data:  data
W = 0.79325, p-value = 0.0009134
```

Ilustración 5- Shapiro.test con datos no normales

Con estos datos (son los valores anteriores con un valor adicional) el W es de 0.79325, y el p-valor es de 0.0009134, obteniendo un cambio significativo en este último valor, y para estos datos la evidencia indicaría que no siguen una distribución normal.

Test Kolmogórov-Smirnov

El test de Kolmogorov-Smirnov (KS) es una prueba no paramétrica utilizada para comparar una distribución empírica con una distribución teórica (test univariante) o para comparar dos distribuciones empíricas entre sí. Este test evalúa si los datos de una muestra pueden considerarse provenientes de una distribución específica o si existen diferencias significativas entre dos conjuntos de datos.

El test KS es especialmente útil en estudios donde se requiere verificar la adecuación de los datos a una distribución teórica, como la normal, exponencial, uniforme, o cualquier otra distribución conocida. También se emplea para comparar dos muestras independientes y determinar si ambas provienen de la misma distribución. Su flexibilidad para trabajar sin asumir previamente una forma paramétrica lo convierte en una herramienta valiosa en estadística descriptiva e inferencial.

Para utilizar el test KS de manera adecuada, se deben considerar los siguientes supuestos:

1. Independencia de las observaciones: Los datos dentro de cada muestra deben ser independientes entre sí.
2. Escala de medición: Los datos deben medirse en una escala ordinal, de intervalo o de razón.
3. Distribución continua: Este test está diseñado para distribuciones continuas. Aunque existen versiones adaptadas para distribuciones discretas, su uso estándar asume continuidad.

Como lo menciona Canavos (1998), la prueba de KS no necesita que los datos se encuentren agrupados y es aplicable a preferiblemente a muestras de tamaño grande, es decir, superiores a 50 observaciones. Esta se basa en una comparación entre las funciones de distribución acumulativa que se observa en la muestra ordenada y la distribución acumulativa que se observa en la muestra ordenada y la distribución propuesta en la hipótesis nula. Si esta comparación revela una diferencia suficientemente grande entre la distribución muestra y propuesta, entonces la hipótesis nula se rechaza.

H_0: los datos siguen la distribución especificada (o las dos muestras provienen de la misma distribución)

H_a:Los datos no siguen la distribución especificada (o las dos muestras no provienen de la misma distribución).

Cuando la prueba Kolmogorov-Smirnov se aplica para contrastar la hipótesis de normalidad de la población, el estadístico de prueba es la máxima diferencia:

$$D = máx|F_n(x) - F_0(x)|$$

Siendo $F_n(x)$ la función de distribución maestral y $F_0(x)$ la función teórica. La hipótesis H_0 se rechaza si para algún valor x observado el valor de D se encuentra dentro de la región crítica de tamaño α. La estadística D mide la mayor discrepancia entre las funciones acumulativas, D se compara con un valor crítico.

En el caso de dos muestras, se comparan $F_{n1}(x)$ y $F_{n2}(x)$, las funciones de distribución acumulativa empíricas de ambas muestras y el estadístico es:

$$D = máx|F_{n1}(x) - F_{n2}(x)| \tag{9}$$

De nuevo, presentando un ejemplo trivial con fines ilustrativos. Sean los datos:

```
> data <- c(1.1, 1.2, 1.3, 1.3, 1.3, 1.4, 1.5, 1.8,
1.8, 1.8, 1.8, 2.0, 2.0, 2.1, 2.1, 2.1, 2.3, 2.2)
```

Ilustración 6 - Datos de distribución normal

Se aplica la prueba, de la siguiente manera:

```
> data_norm <- scale(data)
>
> ks.test(data_norm, 'pnorm')

        Asymptotic one-sample Kolmogorov-Smirnov
        test

data:  data_norm
D = 0.18557, p-value = 0.565
alternative hypothesis: two-sided

Aviso:
In ks.test.default(data_norm, "pnorm") :
  ties should not be present for the one-sample Kolmogorov-Smirnov test
```

Ilustración 7- ks.test

Como se evidencia existen empates (valores repetidos en el conjunto de datos), por ello el aviso que nos proporciona R, para solucionarlo se puede agregar un poco de ruido (sumarle un pequeño valor distinto a cada dato) para evitar los empates de la siguiente manera:

```
> data_noise <- data + rnorm(length(data), mean = 0, sd = 0.001)
>
> data_noise
 [1] 1.100357 1.200512 1.300502 1.301013 1.300549
 [6] 1.398176 1.499651 1.800533 1.800236 1.799583
[11] 1.800213 2.000469 2.001750 2.100504 2.099105
[16] 2.101349 2.300240 2.199496
```

Ilustración 8 - Datos con "ruido normal"

Verificamos que los datos no son repetidos con la función anyDuplicated(), si obtenemos cero es que ningún valor esta repetido:

```
> anyDuplicated(data_noise)
[1] 0
```

Ilustración 9 - anyDuplicated

Finalmente volvemos a realizar la prueba de Kolmogorov-Smirnov:

```
> ks.test(scale(data_noise), 'pnorm')

        Exact one-sample Kolmogorov-Smirnov test

data:  scale(data_noise)
D = 0.1849, p-value = 0.5112
alternative hypothesis: two-sided
```

Ilustración 10 – ks.test

De esta manera ya no tenemos el aviso de R y evidenciamos además que los valores no cambian mucho comparándolos con la prueba realizada sin agregar el ruido, sin embargo, en ambos casos (con ruido o sin ruido en los datos), el valor p es mayor que el valor nivel de significancia de significancia de 0.05, por lo tanto, no se rechaza la normalidad de los datos.

2.3 Modelo de regresión lineal múltiple

La técnica de análisis de regresión múltiple permite estudiar la relación lineal de una variable dependiente con dos o más variables independientes. Esta técnica es la que se utiliza para reconocer si el tiempo que los estudiantes pasaban en el celular y su actitud hacia las matemáticas tenían alguna influencia en su calificación final del año en la materia de matemáticas; la razón de utilizar esta técnica se describirá en la Fase 5 del trabajo (Capítulo 3, sección 3.6).

Como menciona Anderson (2008), esta técnica estadística parte de una ecuación lineal la cual describe cómo se relaciona la variable dependiente y con las independientes (x_1, x_2, ..., x_p). Esta ecuación es conocida como el modelo de regresión múltiple, el cual tiene la siguiente forma:

$$y = \beta_0 + \beta_1 x_1 + \beta_2\, x_2 + \cdots + \beta_p x_p + \epsilon$$

En este modelo $\beta_0, \beta_1, \beta_2, \dots, \beta_p$, son parámetros y el término del error se denota por ϵ.

Al examinar este modelo se evidencia que y es una función lineal de las variables independientes x_1, x_2, ..., x_p más el término del error ϵ, este último corresponde a la variabilidad en y que no se puede explicar por el efecto lineal de las p variables independientes. Adicionalmente este término de error permite analizar los supuestos del modelo de regresión múltiple como se verá más adelante.

Cabe mencionar que si se pudieran conocer los valores de los parámetros $\beta_0, \beta_1, \beta_2, \dots, \beta_p$, se podría usar la ecuación tal y como se mostró anteriormente, sin embargo, los valores de estos parámetros no suelen conocerse, y por lo tanto es necesario estimarlos con base en los datos muéstrales. En este caso se calculan los estadísticos muéstrales $b_1, b_2, \dots, b_p$, los cuales se usan como estimadores puntuales de los parámetros $\beta_0, \beta_1, \beta_2, \dots, \beta_p$, obteniendo así la ecuación de regresión múltiple estimada:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \cdots + b_px_p$$

En la cual $b_1, b_2, \dots, b_p$ solo las estimaciones de $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ y $\hat{y}$ es el valor estimado de la variable dependiente.

Para obtener la ecuación de regresión múltiple estimada el método de los mínimos cuadrados es uno de los más usuales. Este método permite obtener la ecuación que mejor aproxima la relación lineal entre las variables dependientes y la independiente.

El método de los mínimos cuadrados se puede calcular empleando la siguiente expresión:

$$\min \sum (y_i - \hat{y}_i)^2,$$

en la cual y_i es el valor observado en la variable dependiente para la observación i, y $\hat{y}_i$ es el valor estimado para la variable dependiente en la observación i.

Con ello se hace que la suma de los cuadrados de los residuales (la diferencia entre los valores observados y los valores estimados de la variable dependiente $(y_i - \hat{y}_i)$) sea un mínimo, de esta manera se logra estimar los coeficientes de la regresión.

Una vez calculados los coeficientes estos se pueden interpretar de la siguiente manera: cada b_i representa la estimación del cambio en y debido a un cambio en una unidad en x_i mientras todas las demás variables independientes permanecen constantes. Por ejemplo, si se tiene la ecuación, $\hat{y} = b_0 + b_1x_1 + b_2x_2$, en la cual la variable dependiente es un tiempo en horas, el valor del coeficiente b_1 indica el aumento (o disminución) esperado en el tiempo correspondiente al incremento en una unidad de x_1, cuando x_2 permanece constante, de manera análoga el valor de b_2 es una estimación del cambio esperado en el tiempo, cuando x_1 es constante.

Por otra parte, cuando se obtiene la ecuación estimada de regresión, se busca medir cómo se ajusta está a los datos; en este caso se utiliza el coeficiente de determinación múltiple (R^2), el cual es la medida de la bondad de ajuste de la ecuación de regresión múltiple estimada. Este coeficiente se puede interpretar como la proporción en la variabilidad de la variable dependiente que se explica por la ecuación de regresión estimada, el coeficiente se calcula de la siguiente manera:

$$R^2 = \frac{SCR}{STC}$$

Tal que:

$$STC = SCR + SCE$$

Donde:

STC es igual a la suma total de cuadrados, $\sum(y_i - \overline{y})^2$

SCR es la suma de cuadrados debida a la regresión, $\sum(\hat{y}_i - \overline{y})^2$

SCE es igual a la suma de cuadrados debida al error, $\sum(y_i - \hat{y}_i)^2$

Cuando se multiplica R^2 por 100, el valor obtenido se interpreta como el porcentaje de la variabilidad en y que es explicada por la ecuación de regresión estimada.

Cabe mencionar que agregar variables independientes al modelo hace que el valor de R^2 aumente, por lo tanto, se opta por ajustar al coeficiente al número de variables independientes para evitar sobreestimar la variabilidad explicada por la ecuación de regresión estimada, de esta manera para n número de observaciones y p variables independientes el coeficiente de determinación ajustado se calcula de la siguiente manera:

$$R_a^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1} \tag{10}$$

Adicionalmente para el modelo de regresión se cuenta con pruebas de significancia, las cuales son las pruebas F y t. Con la prueba F (también llamada prueba de significancia global) se busca determinar si existe una relación significativa entre la variable de respuesta (dependiente) y el conjunto de las variables explicativas.

Si la prueba F indica que hay significancia global, se utiliza la prueba t (conocida como prueba de significancia individual) para ver si cada una de las variables individuales es significativa, es decir se realiza una prueba t para cada una de las variables independientes del modelo.

Prueba *F*

Sea el modelo de regresión múltiple $y = \beta_0 + \beta_1 x_1 + \beta x_2 + \cdots + \beta_p x_p + \epsilon$, la prueba contrasta las siguientes hipótesis:

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_p = 0$$

$$H_a: \beta_\mathrm{i} \neq 0; para\ algún\ i\ tal\ que\ \ 1 \leq i \leq p$$

Si se rechaza la hipótesis nula, la prueba proporciona suficiente evidencia estadística para concluir que uno o más parámetros no es igual a cero y por lo tanto la relación global entre y y el conjunto de variables explicativas es significativa; por el contrario, si no se puede rechazar H_0 no se tiene evidencia para concluir que exista una relación significativa.

El estadístico de esta prueba es el siguiente:

$$F = \frac{CMR}{CME}$$

En donde CMR es el cuadrado medio debido a la regresión, el cual se calcula como:

$$CMR = \frac{SCR}{p}$$

En la cual SCR es la suma de cuadrados debida a la regresión y p indica los grados de libertad.

Por su parte CME indica el cuadrado medio debido al error, el cual se calcula como:

$$CME = \frac{SCE}{n - p - 1}$$

Cabe mencionar que este proporciona una estimación insesgada de la varianza del término del error ϵ.

Con base en esto se establece la regla de rechazo de la hipótesis nula:

- Rechazar H_0 si el $valor\ p \leq \alpha$, o,
- Rechazar H_0 si $F \geq F_\alpha$, donde F_α pertenece a la distribución F con p grados de libertad.

Ahora bien, si la prueba F indica que la relación de regresión múltiple es significativa, se puede realizar una prueba t para determinar la significancia de cada uno de los parámetros.

Prueba t

Para cualquier parámetro β_i se tiene que:

$$H_0: \beta_i = 0$$

$$H_a: \beta_i \neq 0$$

Cuyo estadístico de prueba es:

$$t = \frac{b_i}{s_{b_i}}$$

En donde

s_{b_i} es la estimación de la desviación estándar de b_i

De esta manera se tiene la regla de rechazo siguiente:

- Rechazar H_0 si el $valor\ p \leq \alpha$, o,

- Rechazar H_0 si $t \leq -t_{\alpha/2}$ o si $t \geq t_{\alpha/2}$, donde $t_{\alpha/2}$ es un valor de la distribución t con $n - p - 1$ grados de libertad.

2.4 Supuestos del modelo de regresión lineal múltiple

Una vez revisados los conceptos principales del modelo de regresión múltiple es importante ahondar en los supuestos de este, por ello, a continuación, se enuncian estos supuestos, los cuales se expresan sobre el término del error ϵ. En este sentido los supuestos del modelo según Anderson (2008) son:

1. El vector de errores ϵ sigue una distribución normal multivariada, cuyo vector de medias o valor esperado es cero, es decir $E(\epsilon) = 0$
2. La varianza de ϵ denotada como σ^2 es constante, es decir es la misma para todos los valores de las variables independientes x_1, x_2, ..., x_p.
3. Los valores de ϵ son independientes entre sí, por lo tanto, el valor de ϵ para un determinado conjunto de valores de las variables independientes no está relacionado con el valor de ϵ de ningún otro conjunto de valores.
4. Las variables independientes se relacionan de manera lineal con la variable dependiente.
5. Las variables explicativas (independientes) son linealmente independientes entre sí.

Para verificar los supuestos anteriormente mencionados, se requieren de algunas pruebas y análisis gráficos para validar el modelo de regresión lineal múltiple, por ello a continuación se muestran dichas pruebas y métodos gráficos que permiten evaluar si cada uno de los supuestos se cumple.

Test Durbin–Watson (Independencia)

Como menciona Anderson (2008) en la regresión lineal existe la posibilidad de que los datos estén correlacionados a lo largo del tiempo, esto sucede cuando el valor de y en el periodo t, denotado como y_t, está relacionado con el valor de y en un periodo anterior. En estos casos existe una autocorrelación en los datos.

Cuando existe la autocorrelación se incumple una de las suposiciones del modelo de regresión: la independencia de los términos del error. Por esta razón es importante verificar que no exista autocorrelación, con el objetivo de que el modelo de regresión sea estadísticamente confiable.

Se pueden presentar dos casos de autocorrelación de primer orden, es decir cuando el valor de y en el periodo t esta relacionado con su valor en el periodo $t-1$, estos dos casos se denominan de autocorrelación positiva y de autocorrelación negativa, en la autocorrelación positiva el residual positivo de un periodo va seguido de un residual positivo correspondiente al siguiente periodo, y si se tiene un residual negativo, el residual del siguiente periodo también será negativo.

Por su parte en la autocorrelación negativa si se tiene un residual positivo, el residual del siguiente periodo será negativo, luego el siguiente residual será positivo y se seguirá esta tendencia de residuales negativos y positivos durante el tiempo t.

Cuando los valores de los errores (los residuales) denotados como ϵ no son independientes, estos estarán relacionados de la siguiente manera:

$$\epsilon_t = \rho\epsilon_{t-1} + z_t$$

En donde:

ρ es un parámetro cuyo valor absoluto es menor que 1

z_t es una variable aleatoria distribuida normal e independiente, con media cero y varianza σ^2

Por lo tanto, si $\rho = 0$, los términos del error no estarán relacionados y cada uno tendrá media cero y varianza σ^2, de esta manera no existe autocorrelación, y se cumple con el supuesto de independencia.

Sin embargo, si $\rho > 0$, existe autocorrelación positiva y si $\rho < 0$, existe autocorrelación negativa, en cualquiera de estas dos situaciones se incumple el supuesto de independencia de los residuales.

Para detectar la autocorrelación, se puede utilizar la prueba de Durbin–Watson la cual permite detectar autocorrelaciones de primer orden. En esta prueba se utilizan los residuales para determinar si $\rho = 0$. Para ello se utiliza el estadístico de Durbin–Watson, el cual se calcula de la siguiente manera:

$$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \tag{11}$$

Donde:

e_t es el residual t, el cual se obtiene de la diferencia entre el valor y_i real y el valor $\hat{y}_i$ calculado, es decir: $e_t = y_i - \hat{y}_i$

Entonces, si los valores sucesivos de los residuales se encuentran cercanos, se tiene autocorrelación positiva y el valor del estadístico de la prueba será pequeño. Por otra parte, si los valores sucesivos de los residuales se encuentran alejados unos de otros, se tiene autocorrelación negativa y el valor del estadístico de la prueba será grande. El estadístico de la prueba va de cero a cuatro y si su valor es dos, esto indica que no existe autocorrelación.

Para esta prueba se realiza un contraste de hipótesis, por lo tanto, la hipótesis nula a contrastar es que no existe autocorrelación.

$$H_0: \rho = 0$$

Y la hipótesis alternativa es que existe autocorrelación:

$$H_a: \rho \neq 0$$

En esta prueba también se tiene los valores críticos que determinan las zonas de rechazo de H_0, las zonas de no rechazo de H_0 y las zonas en las que la prueba es no concluyente, esto lo podemos ver con el siguiente diagrama

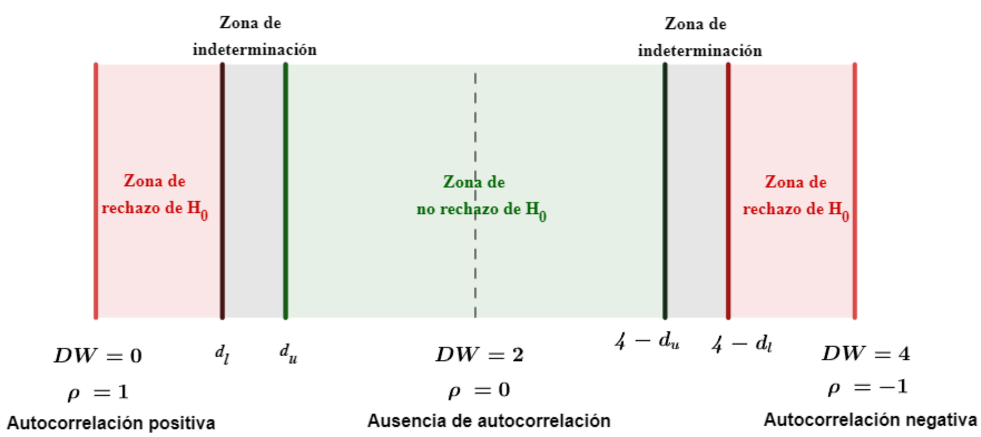

Ilustración 11- Zonas de rechazo y aceptación del Test Durbin–Watson

Donde d_l es el valor critico inferior y d_u es el valor critico superior.

De esta manera se puede contrastar el estadístico de la prueba y si:

$d < d_l$, entonces de rechaza H_0 y se concluye que existe autocorrelación positiva, también se rechaza H_0, si: $d > 4 - d_l$, teniendo así autocorrelación negativa.

Por otra parte, si $d_u < d < d_l$ no se rechaza H_0 y se puede concluir que no hay evidencia estadística de autocorrelación, por lo tanto, se puede afirmar que los residuales son independientes.

Cuando d cae en las zonas de indeterminación, se afirma que no se tiene suficiente evidencia estadística para concluir que exista o no una autocorrelación en los residuales.

Test Breusch-Pagan (homocedasticidad)

El test de Breusch-Pagan es una prueba estadística utilizada para detectar la presencia de heterocedasticidad en un modelo de regresión. La heterocedasticidad ocurre cuando la varianza de los errores no es constante a lo largo de las observaciones, lo que puede afectar la validez de las inferencias estadísticas basadas en la regresión (Wooldridge, 2010).

Este test evalúa si la varianza de los errores depende de los valores de las variables explicativas en el modelo. Su aplicación es común en análisis econométricos y en estudios donde la estabilidad de la varianza del error es un supuesto clave para obtener estimaciones eficientes.

Para utilizar el test de Breusch-Pagan (BP) de manera adecuada, se deben considerar los siguientes supuestos:

1. Independencia de los errores: Los errores deben ser independientes entre sí.
2. Modelo de regresión lineal: La prueba parte de un modelo de regresión estimado previamente.

3. Normalidad de los errores: Aunque la prueba puede ser robusta en algunos casos, se recomienda que los errores sigan una distribución normal para obtener resultados más confiables.

Como lo mencionan Breusch y Pagan (1979), la prueba se basa en una regresión auxiliar de los residuos al cuadrado sobre las variables explicativas originales. Si la varianza del error depende sistemáticamente de estas variables, entonces la hipótesis nula se rechaza, indicando la presencia de heterocedasticidad.

H_0: los errores tienen varianza constante (homocedasticidad)

H_a: los errores no tienen varianza constante (heterocedasticidad)

El estadístico de prueba se calcula como:

$$LM = nR^2 \tag{12}$$

Donde n es el número de observaciones y R^2 es el coeficiente de determinación de la regresión auxiliar. Este estadístico sigue aproximadamente una distribución χ^2 con grados de libertad iguales al número de variables explicativas. Según Wooldridge (2010) el criterio para rechazar H_0 es que el p-valor sea menor que el nivel de significancia α elegido, en tal caso se acepta H_a afirmando que los errores no tienen varianza constante.

Multicolinealidad y el test VIF

En la regresión múltiple se cuenta con variables independientes, las cuales se usan para explicar o predecir la variable dependiente. No obstante, estas variables puede que no sean independientes entre ellas, ya que, en cierta medida, estas variables pueden estar correlacionadas unas con otras, lo cual puede traer problemas de sobreajuste al modelo.

En este sentido la multicolinealidad se refiere a la correlación entre las variables independientes, por lo tanto, como menciona Rendon (2023) “Cuando los predictores del modelo tienen una alta correlación lineal entre sí (es decir que una variable del modelo puede ser explicada por otras que también forman parte del modelo o que una variable pueda expresarse como combinación lineal de otras), se dice que hay multicolinealidad” (p. 21). Por ello cuando hay multicolinealidad, algunos teóricos sugieren eliminar las variables que la están generando.

Para determinar qué variables están generando la multicolinealidad, se puede utilizar la prueba VIF (*variance inflation factor*), cuya formulación para una variable x_j es:

$$VIF_j = \frac{1}{1 - R_j^2} \tag{13}$$

En donde:

R_j^2 es el coeficiente de correlación múltiple del modelo de regresión.

Como mencionan Chatterjee y Hadi (2006) “si x_j tiene una fuerte relación linear con otras variables predictoras entonces R_j^2 será cercano a 1 y por lo tanto el valor del VIF_j será grande, entonces, cuando el factor de inflación de la varianza sea mayor a 10 esto suele ser una señal de que los datos tienen un problema de colinealidad” (p. 236).

2.5 Método de Stepwise

Hay algoritmos que funcionan comparando distintos modelos lineales construidos a partir de diversas combinaciones de un mismo conjunto de variables, esto con el fin de seleccionar la mejor combinación de ellas y así elegir el mejor modelo (Gonzales, 2022); esto asumiendo que el mejor modelo no necesariamente es aquel que integra la mayor cantidad de variables.

El método de Stepwise es una técnica de selección de variables en modelos multivariados, utilizada para identificar las variables más relevantes en la predicción de una variable dependiente. Su objetivo es construir un modelo parsimonioso, es decir, un modelo que incluya únicamente las variables explicativas más significativas, eliminando aquellas que no aportan información relevante.

Este método se basa en la incorporación y eliminación sucesiva de variables en función de su impacto en el modelo, evaluado mediante criterios estadísticos como el p-valor, el coeficiente de determinación ajustado (R^2 ajustado), el criterio de información de Akaike (AIC) o el criterio de información bayesiano (BIC).

Existen tres variantes principales del método *stepwise*: selección hacia adelante (*Forward stepwise*), eliminación hacia atrás (*Backward stepwise*) y regresión paso a paso (*Stepwise*).

La selección hacia adelante comienza con un modelo nulo, es decir, sin variables explicativas. A partir de ahí, se van incorporando variables una por una en función de su significancia estadística. En cada paso, se elige la variable que más mejora el ajuste del modelo y el proceso continúa hasta que ninguna variable adicional contribuye significativamente a mejorar la precisión del modelo (Gonzales, 2022).

Por otro lado, la eliminación hacia atrás parte de un modelo que incluye todas las variables candidatas. A medida que avanza el proceso, se eliminan, una por una, las variables menos significativas. En cada iteración, se descarta la variable cuyo p-valor es mayor que el umbral predefinido (Gonzales, 2022). El procedimiento finaliza cuando todas las variables restantes en el modelo son estadísticamente significativas.

Finalmente, la regresión paso a paso combina ambos enfoques. Se inicia con un modelo nulo o con un conjunto de variables iniciales, añadiendo variables significativas y eliminando aquellas que dejan de serlo tras la incorporación de nuevas variables. Este proceso continúa iterativamente hasta que no haya más cambios en el conjunto de variables seleccionadas.

Sin embargo, como menciona (Gonzales, 2022): "Un primer inconveniente evidente en este tipo de métodos es que, al aumentar el número de variables, el número de modelos que tendremos para comparar es cada vez mayor" (p. 12). Como lo menciona el autor este tipo de métodos estima demasiados modelos, incluso para programas como Excel, por esta razón para su implementación es recomendable usar software pensado para análisis de datos profesional como R, SPSS o Python.

2.6 Rendimiento académico

En esta sección se abordan diversas definiciones relacionadas con el rendimiento académico escolar, un tema que recubre una gran importancia en el ámbito educativo y que es tratado con precisión en este estudio. El propósito de este análisis es profundizar en la comprensión de un concepto clave, ya que el rendimiento académico constituye una de las principales variables de interés que guiarán el desarrollo de este análisis.

De acuerdo con Pizarro (1985, citado en Lamas, 2015, p. 315), el rendimiento académico puede entenderse como un indicador estimativo de las capacidades demostradas, que refleja de manera aproximada la adquisición[5] de aprendizajes obtenidos a lo largo de un proceso de instrucción o formación. En esta misma línea Martínez - Otero (1996, citado en Rodríguez, 2014, p. 15), definen el rendimiento académico como un "producto que rinde o da el alumnado en el ámbito de los centros oficiales de enseñanza". En consonancia con estas posturas, Monsalve (2016) señala que:

> El rendimiento académico hace referencia a la evaluación del conocimiento adquirido en ámbito escolar. Un estudiante con buen rendimiento académico es aquel que obtiene calificaciones positivas en los exámenes que debe rendir a lo largo de un curso. En otras palabras, el rendimiento académico es una medida de las capacidades del alumno, que expresa lo que éste ha aprendido a lo largo del proceso formativo. (p. 4)

De esta manera, y en concordancia con lo planteado por Monsalve (2016), para Caballero et al., (2007), el rendimiento académico se define como el cumplimiento de los objetivos, logros y metas establecidas en el programa o asignatura que cursa el estudiante, reflejándose a través de calificaciones que resultan de un proceso de evaluación. Dicho proceso determina si el estudiante aprueba o no las diversas pruebas, materias o cursos, evidenciando así su nivel de desempeño[6].

En síntesis, a partir de las definiciones citadas, se puede concluir que el rendimiento académico se concibe como un indicador que refleja el nivel de adquisición de aprendizajes a lo largo de un proceso formativo. Dicho indicador se manifiesta mediante el cumplimiento de los objetivos y metas establecidos en los programas educativos, así como en las calificaciones que el estudiante obtiene en evaluaciones formales. En este sentido, el rendimiento académico se

---

[5] En este contexto, el término *adquisición* hace referencia al proceso mediante el cual los estudiantes incorporan, asimilan y consolidan conocimientos, habilidades o actitudes como resultado de experiencias de aprendizaje.

[6] El *nivel de desempeño* hace referencia al grado en que un estudiante demuestra dominio de los conocimientos, habilidades y competencias esperadas en relación con los objetivos de aprendizaje propuestos

configura como una medida de las capacidades demostradas en el ámbito escolar y una evidencia del grado de apropiación de los contenidos impartidos en cada asignatura o curso.

Tal como destacan estos mismos autores, las calificaciones se configuran como uno de los principales indicadores del rendimiento académico, pues permiten evidenciar de forma relativamente objetiva el nivel de logro alcanzado por el estudiante en las diferentes asignaturas o cursos. En el presente estudio, por tanto, se tomará en consideración la calificación en la clase de matemáticas con el fin de analizar el desempeño académico de los estudiantes en esta asignatura, tomando en cuenta su valor como medida cuantificable, representativa del grado de comprensión y apropiación de los contenidos impartidos. Desde luego, reconocemos que el rendimiento atiende a distintos tipos de factores que influencian el desempeño de los estudiantes; sin embargo, para efectos del modelo de regresión a diseñar, serán las calificaciones el elemento que se analice.

Escala de valoración nacional

En el sistema educativo colombiano, los lineamientos para la cuantificación del rendimiento académico los establece el Decreto No. 1290 del 2009 (Ministerio de Educación Nacional, 2009), expedido por el Ministerio de Educación Nacional. El Artículo 5 habla sobre la escala de valoración nacional, señalando que cada institución educativa tiene la autonomía para definir su propia escala. Sin embargo, deberá garantizar su equivalencia con los niveles de desempeño nacional: desempeño bajo, desempeño básico, desempeño alto y desempeño superior.

Escala de valoración institucional

En consonancia con la normativa nacional, el colegio Venecia IED de Bogotá, ha definido la escala de calificación en su Sistema de Evaluación Institucional (SIE) que se presenta en la Tabla 2:

Tabla 2 – Escala de calificación del colegio Venecia IED

ESCALA CUALITATIVA	DESCRIPCIÓN	ESCALA CUANTITATIVA
Desempeño Superior	Presenta dominio de los logros propuestos y profundiza en el conocimiento, con un excelente nivel de desempeño.	4.6 a 5.0
Desempeño Alto	Demuestra alcance de los logros propuestos y calidad de trabajo escolar, con un buen nivel de desempeño.	4.0 a 4.5
Desempeño Básico	Cumple con los logros básicos y presenta dificultades en su trabajo, con un aceptable nivel de desempeño.	3.0 a 3.9
Desempeño Bajo	No supera los logros propuestos y muestra poco interés por el aprendizaje.	1.0 a 2.9

2.7 Impacto del uso del celular

El uso del celular ha incrementado de manera significativa en las últimas décadas, convirtiéndose en una herramienta indispensable para la vida diaria de muchas personas. Según un informe del DANE (2023), el 85,8% de los habitantes de Bogotá mayores de cinco años utilizan teléfonos celulares, lo que demuestra el alto nivel de penetración de estos dispositivos en la sociedad colombiana. Este fenómeno no es exclusivo de los adultos, ya que niños y jóvenes también muestran una adopción considerable de esta tecnología.

El celular, en sus inicios, fue concebido principalmente como una herramienta para la comunicación móvil. Según Naranjo et al. (2016):

> La tecnología móvil, al igual que el internet comparten un origen similar: las necesidades y requerimientos de comunicaciones de las fuerzas militares de las naciones poderosas de la época. Durante la segunda guerra mundial se evidenció esta gran necesidad de

comunicaciones y movilidad, por lo que se construyeron los primeros medios inalámbricos de telecomunicaciones. (p. 4)

El celular ha trascendido su función original de comunicación para convertirse en un dispositivo multifuncional. Pedrero et al. (2012) destacan que, además de permitir la comunicación, los teléfonos móviles facilitan el acceso a internet, redes sociales y aplicaciones educativas, entre otras funciones. Este acceso constante a información y entretenimiento plantea desafíos y oportunidades en el ámbito educativo.

El uso del teléfono celular conlleva tanto ventajas como desventajas. De acuerdo con Besoli et al. (2018), las repercusiones perjudiciales prevalecen y abarcan los aspectos físicos, psicológicos y sociales, tales como afectaciones de salud, amenazas a la seguridad vial, intimidación y acoso. De igual modo, un uso desmedido del teléfono móvil puede incidir en la percepción de seguridad, la construcción de la identidad y la integración en grupos sociales.

Impacto social del uso del celular

En medio de un entorno cada vez más marcado por el uso del celular, el diálogo familiar se ha reducido al mínimo. La interacción entre padres e hijos se ha tornado fría, distante y poco expresiva, mientras que los jóvenes han reemplazado la comunicación directa por mensajes de texto (Bermello,2016). Este fenómeno contribuye a un creciente aislamiento y a la sensación de soledad, debido a la falta de relación cercana entre las personas. La virtualidad, en muchos casos, ha desplazado el contacto presencial, debilitando los lazos familiares y sociales.

Por otra parte, el uso excesivo del teléfono celular puede tener un impacto directo en la salud física, afectando diversos aspectos del bienestar. Tal como señalan Cadena et al. (2016, p. 63): "Respecto a los problemas de salud físicos, destacan tres principales: el malestar en los ojos, seguido de dolor de cabeza y finalmente el dolor en las articulaciones de las manos, lo que, según la literatura, se debe a la posición que se adopta al utilizar el teléfono celular". También, los teléfonos celulares emiten campos electromagnéticos que, al permanecer en contacto cercano con el cerebro, podrían afectarlo de manera significativa; incluso, en casos extremos, se ha asociado su uso con la aparición de tumores cerebrales (Bermel1o, 2016). Asimismo, pueden presentarse enfermedades como el síndrome del túnel carpiano, enfermedad que afecta los brazos, además de problemas en la columna, en los nervios cervicales y trastornos del sueño (Cadena et al., 2016).

Impacto psicológico del uso del celular

En cuanto al ámbito psicológico, Cadena et al. (2016) destacan que los principales síntomas incluyen pensar de manera constante en el teléfono, experimentar incomodidad cuando no se responde de inmediato una llamada o mensaje y manifestar preocupación ante la ausencia del dispositivo. Asimismo, se ha observado que el uso excesivo del celular puede incrementar la ansiedad, el estrés y la dependencia, afectando la salud mental y el rendimiento en actividades cotidianas, tales como el rendimiento académico.

En línea con lo mencionado, Kuss y Griffiths (2017), indican que el uso excesivo de redes sociales puede convertirse en una forma de afrontar el estrés y los problemas emocionales de la vida diaria, tales como la soledad o la depresión, llegando a facilitar patrones de conducta adictivos. Además, la dependencia emocional del celular puede generar sentimientos de inquietud o irritabilidad cuando el dispositivo no está disponible, lo que se traduce en una menor predisposición para participar en actividades de aprendizaje y en la disminución de la calidad de la interacción en el aula. Asimismo, se resalta que el fenómeno conocido como *fear of missing out* (FOMO[7])

[7] Es un término que se traduce como "miedo a quedarse por fuera" o "miedo a perderse algo". Se refiere a una ansiedad social que surge cuando una persona siente que otros están teniendo experiencias gratificantes o importantes de las que ella no forma parte.

contribuye a incrementar la ansiedad y a intensificar la necesidad de mantenerse conectado, generando un círculo vicioso de dependencia.

Impacto cognitivo del uso del celular

El uso excesivo del teléfono celular puede interferir significativamente en los procesos cognitivos implicados en el aprendizaje. Diversos estudios han evidenciado que la constante exposición a estímulos digitales y la multitarea inducida por el celular pueden afectar la capacidad de concentración y la consolidación de la memoria. Por ejemplo, Rosen et al. (2011) sostienen que la interrupción frecuente de la atención, causada por notificaciones y mensajes, reduce la eficacia de la memoria de trabajo, lo cual dificulta la retención y el procesamiento de la información en contextos académicos. Este fenómeno se traduce en una menor capacidad para resolver problemas complejos y en la disminución del rendimiento en actividades que requieren un alto grado de concentración, lo que incide directamente en el aprendizaje.

Issa e Isaias (2016) y Baron (2016) también coinciden al afirmar que el uso de dispositivos móviles e internet puede ejercer un efecto perjudicial en habilidades cognitivas básicas, particularmente en la memoria. Esto se debe a que el cerebro se ajusta a los estímulos que lo rodean, generando conexiones mentales poco deseables y confiando en la tecnología para tareas que requieren recordar datos clave.

No obstante, es importante destacar que el celular, cuando se utiliza de manera controlada y con fines pedagógicos, puede convertirse en un recurso valioso para potenciar el aprendizaje. Chen y Yan (2016) indican que la integración estratégica de dispositivos móviles en el proceso educativo puede favorecer el acceso a recursos digitales, la realización de actividades colaborativas y el fomento de la participación activa de los estudiantes. En concordancia, Castro et al. (2007)

mencionan que su integración adecuada en los procesos de enseñanza puede enriquecer la experiencia de aprendizaje, permitiendo el acceso a recursos digitales y promoviendo la innovación en la educación.

El uso del teléfono celular representa un reto de gran complejidad que involucra tanto beneficios como desafíos en los ámbitos social, físico, psicológico y cognitivo. Si bien, el uso excesivo puede ocasionar problemas de salud, dependencia, aislamiento y menor desempeño académico, la implementación de estrategias de uso responsable puede contrarrestar dichos efectos negativos. Al integrarlo de manera apropiada en procesos de enseñanza y aprendizaje, el celular puede convertirse en un instrumento que favorezca el trabajo en grupo, el acceso ágil a información y la participación activa de los estudiantes, contribuyendo así al fortalecimiento de sus competencias y al mejoramiento de su rendimiento académico.

Impacto del celular en el rendimiento académico

En el estudio de Tsitsika et al. (2014), realizado en seis países europeos con una muestra de 10,930 adolescentes, se encontró lo siguiente: "*Adolescents who engaged heavily in SNS scored significantly lower in offline Activities and Academic performance compared with adolescents who used SNS moderately* (both $p < .001$)."[8] (p. 144). Aunque esta investigación se centró en las redes sociales, Kuss y Griffiths (2017) destacan que estas constituyen la principal actividad a la que los jóvenes dedican su tiempo en los teléfonos móviles. Además, el estudio muestra que el 40% de los adolescentes pasa más de dos horas diarias en redes sociales, evidenciando una relación inversa

[8] Los adolescentes que empleaban de manera intensiva las redes sociales obtuvieron puntuaciones significativamente más bajas en las actividades fuera de línea y en su rendimiento académico, en comparación con quienes las usaban de forma moderada (ambos $p < .001$).

entre el uso excesivo de estas plataformas y el rendimiento académico, pues el tiempo en línea desplaza las horas de estudio.

Investigaciones como las de Hilt (2019) han explorado la dependencia al celular y cómo esta afecta aspectos fundamentales como los hábitos de lectura y el rendimiento académico. Como lo muestra la Ilustración 12 con $r = -0.207$, $p = 0.020$, esto indica que, a mayor adicción del celular, menor es el promedio académico, aunque sea una correlación débil (ver Tabla 1).

Correlaciones

		Hab_Lect	TDM	Act_Lect	Promedio
Hab_Lect	r de Pearson	1	-.184*	.536**	.037
	P		.039	.000	.682
TDM	r de Pearson	-.184*	1	-.230**	-.207*
	P	.039		.009	.020
Act_Lect	r de Pearson	.536**	-.230**	1	.213*
	P	.000	.009		.017
Promedio	r de Pearson	.037	-.207*	.213*	1
	P	.682	.020	.017	

Ilustración 12- Hilt 2019 Correlaciones

De igual forma, el mismo estudio revela que a mayor uso del celular, menores son los hábitos de lectura. La lectura desempeña un papel fundamental en el proceso educativo, puesto que de su dominio depende en gran medida el acceso a los saberes de cualquier disciplina y, en consecuencia, el logro académico. (Ochoa et al., 2017)

"Los escolares que pasaban más tiempo al día frente a una pantalla […] se percibían más lentos en la solución de problemas matemáticos, mantenían menos la atención en clases y resolvían con mayor dificultad tareas complejas en el centro educativo" (Martínez et al., 2021, p. 573). En este estudio se logró evidenciar que el tiempo que los estudiantes pasan en las pantallas; en redes sociales, juegos, navegando en internet, etc., se asocia con un bajo rendimiento escolar, teniendo incidencias directas en su aprendizaje.

Los estudios de Tsitsika et al. (2014), Hilt (2019) y Martínez et al. (2021) coinciden al evidenciar una correlación negativa entre el uso excesivo de dispositivos móviles y el rendimiento académico. Esta tendencia se manifiesta en la reducción del tiempo de estudio, la disminución de los hábitos de lectura, las dificultades para sostener la atención y resolver problemas, afectando directamente el rendimiento académico.

2.8 Actitudes hacia las matemáticas

En el último tiempo, el estudio de las actitudes hacia las matemáticas ha cobrado una gran relevancia en el ámbito educativo, ya que se ha reconocido que el aprendizaje de esta disciplina no depende únicamente de factores cognitivos como la memoria o la capacidad lógica, sino también de aspectos afectivos, emocionales y motivacionales. En una de sus obras, McLeod (1988) marca un punto de partida en la preocupación por las emociones y los sentimientos en matemáticas, estableciendo una distinción ya clásica entre actitudes, creencias y emociones como componentes del llamado dominio afectivo matemático. De entre estos componentes, las actitudes han ocupado un papel preponderante en la educación matemática, tanto por su impacto directo en el rendimiento como por la cantidad de investigaciones que han generado (Palacios et al., 2014). Así, comprender y trabajar en el desarrollo de actitudes positivas hacia las matemáticas se ha convertido en un objetivo clave para fomentar una relación más cercana y significativa con esta ciencia.

Según Gómez (2000), las actitudes son una predisposición evaluativa con carácter positivo o negativo, que determina las intenciones individuales e influye sobre la conducta, estando compuesto por elementos cognitivos y afectivos. Por su parte, Auzmendi (1992; como se cita en Méndez y Macía, 2007) menciona que las actitudes son "aspectos no directamente observables sino inferidos, compuestos tanto por las creencias como por los sentimientos y las predisposiciones comportamentales hacia el objeto al que se dirigen" (p. 338). El autor indica que las actitudes, aunque no visibles directamente, se manifiestan en las creencias, sentimientos y acciones. Esto explica por qué distintas personas reaccionan de formas diversas ante una misma situación.

Siguiendo la misma línea está Palacios et al. (2014), quienes señalan que "La actitud hacia las matemáticas tendría que ver con la valoración, el aprecio y el gusto por esta disciplina subrayando más la vertiente afectiva que la cognitiva" (p. 68). Es importante aclarar la diferencia que hay entre actitud hacia las matemáticas y actitud matemática, esta última va ligada al modo en que se usan ciertas capacidades matemáticas, sujetas a la parte cognitiva.

Como señalan Zan y Martino (2007), la actitud hacia las matemáticas es un constructo complejo que comprende creencias, emociones y comportamientos. Cuando una persona se acerca a las matemáticas con una disposición positiva y confianza en sus habilidades, es más probable que asuma retos con curiosidad y perseverancia, lo que incrementa sus posibilidades de éxito.

Por el contrario, una actitud negativa o el temor al fracaso puede obstaculizar el proceso de aprendizaje y debilitar la confianza en la propia capacidad. En la misma línea Martínez (2008), menciona que:

> El éxito o el fracaso, en el aprendizaje de los contenidos matemáticos tiene más de un responsable y, en el caso del aprendiz, suele atribuirse no sólo a la configuración cognitiva del sujeto sino, también, al capital afectivo, pues, muchas de sus reacciones evaluativas y predisposiciones de actuar, de los sujetos ante los objetos, suelen depender de sus creencias, emociones o sentimientos […] dado que los niveles de satisfacción, interés, frustración, alegría, gusto, repugnancia, apego, incertidumbre, miedo, aversión, desánimo, resistencia o preocupación suelen condicionar sus actuaciones. (pp. 251-252)

Las emociones, creencias y actitudes están involucradas con el éxito o con el fracaso de los estudiantes en el desarrollo de sus tareas, y en la construcción de saberes matemáticos. Por ello consideramos que la variable actitud frente a la clase de matemáticas es digna de ser analizada.

En este sentido, Bandura (1997) destaca que "*Perceived self-efficacy is defined as people's beliefs about their capabilities to produce designated levels of performance that exercise influence over events that affect their lives. Self-efficacy beliefs determine how people feel, think, motivate themselves and behave*"[9] (p. 1). El autor presenta cómo la autoconfianza y la percepción sobre la propia eficacia influyen directamente en el aprendizaje. Cuando una persona cree en sus capacidades, enfrenta mejor los retos, aumentando sus probabilidades de éxito; en cambio, la falta de confianza limita su desempeño académico.

Para abordar el estudio de las actitudes de forma sistemática, se ha adoptado un modelo ampliamente aceptado por diversos investigadores en el campo: el modelo tripartito o tridimensional. "Este modelo, propuesto en 1960 por Rosenberg y Hovland, considera que las actitudes están formadas por tres componentes: afectivo, cognitivo y conductual" (Ursini y Sánchez, 2019, p. 17). El componente afectivo se refiere a las emociones que una persona experimenta hacia las matemáticas, como agrado, ansiedad o rechazo. El cognitivo, por su parte, implica las creencias y conocimientos que se tienen sobre esta disciplina, como su utilidad o dificultad. Finalmente, el componente conductual —también conocido como conativo— hace referencia a las disposiciones o intenciones de actuar frente a las matemáticas, como el esfuerzo por participar en clase o, por el contrario, evitarlas.

---

[9] La autoeficacia percibida se define como las creencias de las personas acerca de sus capacidades para alcanzar ciertos niveles de desempeño que influyen sobre los eventos que afectan sus vidas. Las creencias de autoeficacia determinan cómo se sienten, piensan, se motivan y se comportan las personas.

En relación con este tema, diversos estudios (*v. g*., Auzmendi, 1992; Bazán, 2002) han analizado la conexión entre el rendimiento académico y la actitud hacia las Matemáticas, evidenciando que, por lo general, las posturas tienden a ser negativas y se asocian con un desempeño insatisfactorio. En la misma línea, Ling y Walberg (1983) trabajaron con una muestra de 2368 estudiantes de 13 años, a quienes aplicaron diversas pruebas para evaluar sus actitudes y desempeño en matemáticas. Tras examinar los resultados, constataron que los grupos con posturas más positivas obtenían mejores calificaciones en la asignatura. Asimismo, observaron que quienes pertenecían a los grupos de mayor rendimiento presentaban actitudes más favorables hacia esta área.

Como mencionan Ursini y Sanchez. (2019), existen diversas formas de identificar las actitudes hacia las matemáticas, entre ellas: diarios de clase, observaciones directas, escalas y entrevistas, entre otras. El método más utilizado para cuantificar las actitudes los estudiantes hacia las matemáticas son las escalas, instrumentos que presentan una serie de afirmaciones ante las cuales el encuestado selecciona su nivel de acuerdo o desacuerdo. Entre los diferentes tipos de escalas, las de tipo Likert son las más comunes. Como ejemplos están la *Mathematics Attitude Scale* (MAS), el Cuestionario de actitudes hacia las matemáticas de Aiken y Dreger, el *Attitudes Toward Mathematics Inventory* (ATMI), la encuesta "Las matemáticas y yo" y la Escala de actitudes hacia las matemáticas de Guarín.

3 CAPITULO III. METODOLOGÍA

En este capítulo se exponen los aspectos metodológicos que sustentan el estudio, se describe el tipo de estudio, las técnicas e instrumentos empleados para la recolección de datos y la descripción de las variables. Asimismo, se comentan las distintas fases que componen el proceso del estudio, desde la etapa de planeación y la recolección de información, hasta el análisis descriptivo de los resultados.

3.1 Pregunta de exploración

Para el desarrollo del trabajo, se considera la siguiente pregunta para ser explorada a través de una regresión lineal múltiple: ¿Cuál es la relación entre el uso del celular y el rendimiento académico en matemáticas de los estudiantes del Colegio IED Venecia de grados octavo y décimo?

Al respecto, la hipótesis frente a la pregunta es que, a mayor uso del celular, menor el rendimiento académico en matemáticas.

3.2 Tipo de estudio

El trabajo se ciñe a los planteamientos de un estudio correlacional, dado que en este trabajo se busca determinar cómo las variables de estudio están correlacionadas de manera positiva, negativa o sin correlación. Este tipo de estudio, como lo indica Monje (2011), es el indicado en situaciones complejas en las que importa relacionar variables, midiéndolas e interrelacionándolas simultáneamente en situaciones de observación, identificando asociaciones entre variables, sin conducir directamente a identificar relaciones causa – efecto, pero sí a conjeturarlas. Además, adopta un diseño cuantitativo y no experimental, esto significa que no manipulamos las variables, sino que observamos y analizamos los datos como ocurren de manera natural.

3.3 Población y Muestra

La población de este estudio la conforman los estudiantes del Colegio IED Venecia. Por su parte, la muestra está compuesta por 40 estudiantes de octavo y décimo grado (14 y 26 estudiantes respectivamente). Cabe destacar que se incluyeron únicamente a quienes contaban con la autorización previa de sus padres (Anexo 1), tuvieron la voluntad de participar y disponían de un teléfono celular en óptimas condiciones[10]. Esto significa que el muestreo realizado no fue probabilístico sino por conveniencia, en consecuencia, señalamos el cuidado que se debería tener al extrapolar los resultados que de aquí se deriven.

3.4 Técnicas e instrumentos de recolección de datos

Como se mencionó antes, se optó por un muestreo no probabilístico por conveniencia, ya que se seleccionaron algunos cursos de octavo grado y décimo grado. De estos cursos se incluyó únicamente a los estudiantes que contaban con un teléfono celular en óptimas condiciones y cuyos padres otorgaron su consentimiento para la participación. Esta estrategia se empleó debido a que la accesibilidad a todos los estudiantes era limitada por normativas institucionales y legales, descartándose un proceso de selección aleatoria.

Para recolectar datos sobre el tiempo de uso del celular, se utilizó la aplicación StayFree[11], disponible de manera gratuita en las tiendas Play Store y App Store. Esta herramienta registra la duración de la actividad diaria del dispositivo y, entre otras funciones, muestra cuántas horas, minutos y segundos se dedican a cada aplicación. Previamente se descargó la aplicación en nuestros celulares para familiarizarnos y poder ver que en realidad los datos que arroja son muy precisos.

---

[10] Cuando se hace referencia a un celular en óptimas condiciones, se entiende un dispositivo que sea un teléfono inteligente (smartphone), que se encuentre en estado funcional y que cuente con suficiente capacidad de almacenamiento para instalar y ejecutar la aplicación requerida sin inconvenientes.

[11] Si quiere saber un poco más de la aplicación puede dirigirse a la siguiente página web: https://stayfreeapps.com/

Por su parte, el registro de calificaciones se obtuvo al finalizar el año escolar 2024. Los profesores del área de matemáticas facilitaron el consolidado de notas finales de cada periodo y la calificación definitiva de la asignatura.

Por otra parte, para las variables de control y la medición de la variable "actitudes hacia las matemáticas" se utilizó la escala AMMEC (Actitudes hacia las Matemáticas y las Matemáticas Enseñadas con Computadores), diseñada por Ursini, Sánchez y Orendain (2004). Esta escala, enfocada en estudiantes de secundaria, consta de 29 afirmaciones organizadas en un formato Likert de cinco puntos, y presenta un elevado nivel de confiabilidad (alfa de Cronbach = 0.795) (Ursini y Sánchez, 2021). Tal como señalan los autores, no es obligatorio aplicar la escala completa. Dado que el interés de este estudio no abarca las actitudes frente a las matemáticas enseñadas con tecnología, se emplearon dos subescalas: la Subescala 1 (AM), que mide las actitudes hacia las matemáticas, y la Subescala 3 (ACM), que evalúa la autoconfianza para el trabajo matemático (Anexo 2).

3.5 Descripción de variables

Variable dependiente

La variable dependiente es la denominada Rendimiento académico (nota final de matemáticas del año 2024): como se mencionó previamente en el marco teórico, una manera de cuantificar el rendimiento académico es a través de las calificaciones (notas) obtenidas a lo largo del curso. No está a nuestro alcance un método de recolección de información diferente. Por esta razón, esta variable se mide con base en las calificaciones de los estudiantes durante el año escolar 2024. Esta variable es de tipo cuantitativa continua y tiene un rango $0 \leq y \leq 5$.

Variables independientes

Se consideran las siguientes variables independientes:

- Tiempo de uso promedio del celular en días escolares: esta variable tiene en cuenta el tiempo del uso del celular, en horas, solo de los días en que los estudiantes asisten a clase. Es de tipo cuantitativa continua y su rango aproximado es $0 \leq x_1 \leq 15$. Su medición se hizo a través de la aplicación StayFree que los estudiantes descargaron en sus celulares.
- Tiempo promedio de uso del celular en días no escolares: esta variable tiene en cuenta el tiempo del uso del celular los días sábados, domingos y festivos. Es de tipo cuantitativa continua y su rango aproximado es $0 \leq x_2 \leq 15$. Su medición se hizo a través de la aplicación StayFree que los estudiantes descargaron en sus celulares.
- Actitudes hacia las matemáticas: esta variable se evalúa con la escala AMMEC (una descripción más detallada de esta escala se encuentra más adelante en la Fase 4 de este mismo capítulo), cuyas afirmaciones se puntúan en un rango de 0 a 4 y, posteriormente a estas puntuaciones, se les calcula el promedio. Es de tipo cuantitativa continua y su rango es $0 \leq x_3 \leq 4$.
- Variables de control: la variable género que es de tipo cualitativa nominal, edad que la registramos de tipo cuantitativa discreta y curso que es de tipo cualitativa ordinal. Estas variables solo se tuvieron en cuenta para el análisis descriptivo de los datos. No se tuvieron en cuenta al momento de realizar el análisis de regresión lineal múltiple porque no eran de interés en este estudio.

- Calificación último periodo: esta variable se mide con base en la calificación en matemáticas de los estudiantes del cuarto periodo del año escolar 2024. Esta variable es de tipo cuantitativa continua, tiene un rango $0 \leq y \leq 5$.
- Grado: esta variable es de tipo cualitativa nominal, es binaria y es de tipo *dummy* (es decir solo toma valores 0 o 1), las únicas dos opciones son grado octavo (8°) o grado (10°).

3.6 Fases del proceso de elaboración del trabajo de grado

El desarrollo del presente trabajo de grado se estructuró en distintas fases que permitieron organizar y ejecutar de manera sistemática cada una de las actividades requeridas para alcanzar los objetivos propuestos. Estas etapas presentan un paso a paso de cómo se desarrolló este estudio estadístico, desde la planeación y revisión teórica, pasando por el trabajo de campo y terminando con un análisis descriptivo de los datos.

Fase 1 - Planeación

Lo primero que se hizo fue buscar referentes teóricos que presentaran experiencias y resultados de estudios similares a los que se pretendían realizar, con el fin de guiar el desarrollo del trabajo. Entre los principales textos consultados se encuentran "Dependencia del celular, hábitos y actitudes hacia la lectura y su relación con el rendimiento académico*"* (Hilt, 2019), "Aumento de horas de pantalla se asocia con un bajo rendimiento escolar*"* (Lamana et al., 2021) y "*Online Social Networking in Adolescence: Patterns of Use in Six European Countries and Links With Psychosocial Functioning*" (Tsitsika et al., 2013).

Una vez consolidada la idea, fue necesario encontrar una institución educativa en la que se pudiera llevar a cabo el estudio. Se decidió realizar la investigación en un colegio vinculado con el Departamento de Matemáticas de la Universidad Pedagógica Nacional. La primera opción fue el

Colegio IED Venecia, ya que uno de los autores de este trabajo de grado estaba realizando sus prácticas allí, entonces la gestión sería más fácil y en efecto, así fue.

El primer paso fue conversar con algunos profesores del área de matemáticas para exponerles la idea de proyecto. Posteriormente, se organizó una reunión con la coordinadora de la jornada de la mañana y una profesora. En dicha reunión se presentó el proyecto y se establecieron varios acuerdos, algunos fueron presentar una carta expedida por el Departamento de Matemáticas (Anexo 3), asistir a la reunión de padres de familia para informar del proyecto, enviar el estudio antes de su presentación, entre otras disposiciones. Con estos requisitos cumplidos, el Colegio IED Venecia otorgó la aprobación para iniciar el estudio.

Fase 2 – Presentación del proyecto

Se llevó a cabo una presentación oficial del proyecto ante estudiantes y padres de familia. En primer lugar, se realizó una exposición del proyecto en cada uno de los cursos participantes, explicando sus objetivos y el propósito de la investigación. Se informó a los estudiantes sobre la necesidad de descargar la aplicación StayFree en sus celulares, asegurándoles que esta no sería invasiva ni afectaría su privacidad. Además, se les indicó que debían completar una escala diseñada para recopilar información sobre sus actitudes hacia las matemáticas.

En segundo lugar, asistimos a la reunión de entrega de boletines del tercer período, donde se presentaron los objetivos, el propósito y la metodología del estudio a los padres de familia. El objetivo de esta reunión fue garantizar total transparencia en el proceso y resolver cualquier inquietud.

Durante este encuentro, se entregó el consentimiento informado (Anexo 1) a los padres de familia y estudiantes participantes. En este documento se explica detalladamente el propósito del estudio, la metodología utilizada y las medidas adoptadas para garantizar la confidencialidad de la información recolectada. Los consentimientos diligenciados por los acudientes fueron recopilados antes de iniciar cualquier actividad relacionada con la investigación.

Fase 3 – Instalación de la aplicación y aplicación del cuestionario

Durante una semana se asistió diariamente a la institución para brindar apoyo a los estudiantes en la descarga e instalación de la aplicación en sus celulares. Se les explicó nuevamente la importancia de mantener la aplicación activa durante un tiempo, ya que los datos recolectados serían fundamentales para el estudio. En total, más de noventa estudiantes lograron descargar e instalar la aplicación con éxito. Sin embargo, algunos dispositivos presentaron limitaciones de almacenamiento, lo que impidió la instalación de la aplicación en algunos casos.

Además, se solicitó a los estudiantes que respondieran de manera individual y autónoma la escala sobre actitudes (Anexo 2). Antes de completar el instrumento, se les proporcionaron instrucciones detalladas sobre cómo debían responder y se enfatizó que la escala no tendría ninguna calificación ni repercusión académica. El propósito era que contestaran de manera honesta, reflejando sus verdaderas percepciones y sentimientos respecto a cada afirmación.

Un mes después de la instalación de la aplicación, se visitó la institución para proceder con la descarga de los archivos que contenían el registro del tiempo de uso del celular por parte de los estudiantes. Durante este proceso, surgieron algunas dificultades: algunos estudiantes habían desinstalado la aplicación antes de completar el período de recolección de datos, mientras que otros habían extraviado sus dispositivos. Finalmente, el tiempo que consideramos para la recolección de los datos sobre el uso del celular fue del 18 de octubre al 13 de noviembre.

A pesar de estos inconvenientes, se logró recopilar un total de 44 archivos con información válida para el estudio. Los datos se descargaron en formato Excel, organizados de acuerdo con la estructura que se presenta en la Ilustración 13. Sin embargo, luego de revisar, cuatro archivos venían con los datos incompletos, no recogieron la información que se necesitaba, por tal motivo fueron descartados.

Aplicaciones	11 de noviembre	12 de noviembre	13 de noviembre	Uso total
Telegram	0s	0s	0s	58m 9s
TeraBox	0s	0s	0s	7m 53s
Threads	0s	0s	0s	3m
TikTok	2h 7m 11s	3h 1m 22s	16m 41s	68h 19m 27s
Traductor	0s	10s	0s	26m 48s
WhatsApp	27m 2s	1h 10m 38s	1h 48m 23s	39h 55m 58s
Word	0s	0s	0s	2m 12s
YouTube	1h 49m 52s	1h 5m 47s	3m 14s	28h 41m 33s
Uso total	11h 7m 19s	12h 16m 15s	5h 53m 46s	256h 41m 10s

Ilustración 13 – Ejemplo del Excel generado por la app StayFree

Para la recolección de las calificaciones de los estudiantes, necesarias para cuantificar el rendimiento académico, fue preciso esperar a que los docentes finalizaran sus cursos y realizaran cierre de notas. Una vez sucedido esto, se obtuvo un archivo en formato Excel que contenía las calificaciones de los estudiantes por período, así como la nota definitiva en la asignatura de matemáticas.

Fase 4- Tratamiento de los datos

Una vez aplicada la escala para identificar las actitudes de los estudiantes hacia las matemáticas, fue necesario procesar las respuestas obtenidas. Para ello se siguió el procedimiento descrito por Ursini y Sánchez (2021).

Dado que la escala utilizada es de tipo Likert con cinco opciones de respuesta (mucho, sí, indeciso, poco y no), cada afirmación debía recibir un puntaje numérico entre 0 y 4. Sin embargo, al asignar estos puntajes era fundamental distinguir entre afirmaciones formuladas en sentido positivo y aquellas en sentido negativo.

Para las afirmaciones planteadas en sentido positivo, los puntajes asignados fueron: no, 0; poco, 1; indeciso, 2; sí, 3 y mucho, 4. Por el contrario, para las afirmaciones formuladas en sentido negativo, la asignación de puntajes se invirtió: no, 4; poco, 3; indeciso, 2; sí, 1 y mucho, 0. Es importante aclarar que nosotros únicamente aplicamos la escala; las afirmaciones ya estaban previamente clasificadas en positivas o negativas en el libro de Ursini y Sánchez (2021).

En la Ilustración 14 se observan las respuestas de un estudiante a dos afirmaciones. La afirmación 1, “Me gusta la clase de matemáticas”, está formulada en sentido positivo. Dado que el estudiante respondió **mucho**, el puntaje asignado a este ítem es **4**. Por otra parte, la afirmación 2, “La clase de matemáticas es aburrida”, está formulada en sentido negativo. En este caso, el estudiante respondió **poco**, por lo que el puntaje asignado a esta afirmación es **3**.

1.	Me gusta la clase de matemáticas	*MUCHO*	*SI*	*INDECISO*	*POCO*	*NO*
2.	La clase de matemáticas es aburrida	*MUCHO*	*SI*	*INDECISO*	*POCO*	*NO*

Ilustración 14 – Ejemplo de la escala de actitudes

La escala que respondieron los estudiantes está compuesta por 18 afirmaciones: las afirmaciones número 2, 3, 10 y 12 están planteadas en sentido negativo; mientras que las afirmaciones 1, 4, 5, 6, 7, 8, 9, 11, 13, 14, 15, 16, 17 y 18 están formuladas en sentido positivo. Posteriormente, se asignó a cada una de las respuestas obtenidas el puntaje correspondiente, según el procedimiento explicado previamente y se organizó esta información en una hoja de Excel (Ilustración 15).

Afirmaciones																		
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Promedio
3	1	3	2	2	3	4	4	0	3	4	0	1	4	4	4	4	4	2,78
3	3	3	2	3	2	4	4	2	0	3	4	2	1	3	1	3	3	2,56
2	1	2	0	2	2	3	3	3	3	3	3	2	3	3	2	1	3	2,28
3	3	1	2	1	3	4	3	1	4	3	4	1	1	3	2	3	1	2,39
3	4	2	0	2	3	3	0	3	2	3	4	2	1	3	2	3	3	2,39
4	3	2	3	3	3	4	2	0	0	1	2	3	2	1	3	3	3	2,33
1	3	1	0	1	1	3	3	2	1	3	1	3	0	0	0	3	1	1,50

Ilustración 15 – Excel con la cuantificación de las respuestas de la escala de actitudes

El principal objetivo era cuantificar la actitud general de cada estudiante hacia las matemáticas. Para ello, se calculó el promedio de los puntajes obtenidos en las 18 afirmaciones para cada estudiante. Este promedio varía en una escala de 0 a 4: cuanto más alto sea este valor, más positiva será la actitud del estudiante hacia las matemáticas, mientras que valores bajos indican una actitud negativa o neutra. A continuación, la Ilustración 16 presenta la clasificación de estos promedios:

PUNTUACIONES OBTENIDAS	TIPO DE ACTITUD
Puntuación = 0	NEGATIVA
0 < puntuación < 1.5	TENDENCIA A NEGATIVA
1.5 ≤ puntuación ≤ 2.5	NEUTRA
2.5 < puntuación < 4	TENDENCIA A POSITIVA
Puntuación = 4	POSITIVA

Ilustración 16 – Clasificación de los promedios de la escala de actitudes

Los archivos de Excel que se descargaron de la aplicación contenían el registro del tiempo de uso del celular expresado en horas, minutos y segundos (Ilustración 13). Debido a esto, se decidió pasar todos los tiempos registrados únicamente a horas.

Posteriormente, se clasificaron los registros según días escolares y días no escolares, y se calculó el promedio del tiempo de uso del celular para cada uno de estos momentos. Así, se definieron dos variables claves: tiempo promedio de uso del celular en días escolares y tiempo promedio de uso del celular en días no escolares.

Finalmente, la base de datos quedó conformada por siete variables (Anexo 4):

1. Calificación del último final de matemáticas.
2. Calificación del cuarto periodo de matemáticas.
3. Tiempo promedio de uso del celular en días escolares (en horas).
4. Tiempo promedio de uso del celular en días no escolares (en horas).
5. Tiempo total de uso del celular en días escolares (en horas).
6. Tiempo total de uso del celular en días no escolares (en horas).
7. Puntuación en la escala de actitudes hacia las matemáticas.
8. Grado escolar (8° o 10°)

Fase 5 – búsqueda del modelo adecuado

Para determinar el modelo estadístico adecuado para analizar los datos, se siguió el procedimiento sugerido por Hair et al. (2019), quienes ofrecen pautas claras para identificar el modelo estadístico más apropiado según las características específicas de las variables y los objetivos del estudio.

Siguiendo estos lineamientos, primero se evaluó el tipo y número de variables involucradas en el estudio. Según Hair et al. (2019, p. 265) "El objetivo del análisis de regresión múltiple es usar variables independientes cuyo valor son conocidos para predecir el único valor dependiente seleccionado por el investigador". Dado que nuestro propósito es examinar la relación que surge entre la variable dependiente calificación final y algunas variables independientes, se optó por utilizar este modelo.

Elegimos específicamente la regresión lineal múltiple porque cumple con los objetivos planteados, ya que nos permite estimar cómo las variables independientes mencionadas influyen simultáneamente sobre la variable dependiente. Además, este modelo admite la inclusión de variables categóricas, lo cual resulta práctico dado que algunas de las variables recolectadas son de esta naturaleza. Adicionalmente, los supuestos que requiere el modelo fueron validados para hacer uso correcto del mismo, como se verá en detalle en la sección de resultados.

Fase 6 - análisis descriptivo de los datos

En esta fase se realiza un análisis descriptivo de los datos, mostrando medidas de tendencia central (media, mediana y moda) y de dispersión (varianza, desviación, rango intercuartil). Este apartado es esencial para comprender la distribución de las variables e identificar posibles anomalías. Además, se presentan histogramas y diagramas de caja y bigotes, acompañados de un análisis de sus resultados.

Para el análisis descriptivo de los datos en el software R, se emplea la nomenclatura que figura en la Tabla 3. Esto se realiza con el propósito de mejorar la legibilidad del código, ya que existen variables con nombres muy extensos, facilitando así la manipulación y la interpretación de la información.

Tabla 3 – Nombres de las variables empleadas en R

Variable	Nombre de la variable en R
Curso	Curso
Genero	Genero
Edad	Edad
Tiempo total de uso del celular en días escolares	Tot_dias_esc
Tiempo promedio de uso del celular en días escolares	Prom_dias_esc

Tiempo total de uso del celular en días no escolares	Tot_dias_noesc
Tiempo promedio de uso del celular en días no escolares	Prom_dias_noesc
Nota final en matemáticas del año 2024	Nota_final
Nota del cuarto corte de matemáticas del año 2024	Nota_4_corte
Puntación de actitudes hacia las matemáticas	Actitudes

La Ilustración 17 muestra un resumen estadístico de las variables contempladas en este estudio. Se incluyen medidas como el valor mínimo (min), el primer cuartil (1st Qu.), la mediana (median), la media (mean), el tercer cuartil (3rd Qu.) y el valor máximo (max) de cada variable.

```
      Edad          Tot_dias_esc     Prom_dias_esc     Tot_dias_noesc
 Min.   :14.00    Min.   : 55.1    Min.   : 3.241    Min.   : 19.80
 1st Qu.:15.00    1st Qu.: 95.8    1st Qu.: 5.746    1st Qu.: 56.05
 Median :16.00    Median :115.3    Median : 6.783    Median : 67.27
 Mean   :15.57    Mean   :114.7    Mean   : 6.753    Mean   : 71.80
 3rd Qu.:16.00    3rd Qu.:130.9    3rd Qu.: 7.702    3rd Qu.: 88.33
 Max.   :17.00    Max.   :189.5    Max.   :11.149    Max.   :121.88
 Prom_dias_noesc    Nota_4_corte      Nota_final       Actitudes
 Min.   : 1.980   Min.   :2.00    Min.   :2.100   Min.   :0.9444
 1st Qu.: 5.606   1st Qu.:3.00    1st Qu.:3.100   1st Qu.:1.8611
 Median : 6.727   Median :3.50    Median :3.350   Median :2.0556
 Mean   : 7.180   Mean   :3.58    Mean   :3.492   Mean   :2.1486
 3rd Qu.: 8.833   3rd Qu.:4.10    3rd Qu.:3.825   3rd Qu.:2.6667
 Max.   :12.188   Max.   :5.00    Max.   :4.900   Max.   :3.3333
```

Ilustración 17– Resumen de estadísticas descriptivas de los datos usando R

La variable Prom_dias_esc tiene como valor mínimo 3,241 y valor máximo 11,149, esto indica que el estudiante que menos utilizó el celular en días escolares pasó un tiempo de 3 horas,12 minutos y 28 segundos diario en el celular; el que más lo utilizó un tiempo de 11 horas, 8 minutos y 56 segundos diario. La media de tiempo que pasan los estudiantes en sus celulares es 6 horas, 44 minutos y 38 segundos; la mediana (2nd Qu.) de los datos es 6 horas, 46 minutos y 59 segundos.

El primer cuartil es 5,746 esto indica que el 25% de los datos esta entre este valor y el valor mínimo, mientras que el tercer cuartil indica que el 75% de los datos están entre este valor y el mínimo.

La Tabla 4 muestra la desviación estándar, la varianza y el rango intercuartil. En las actitudes la desviación estándar de 0.532 y la varianza de 0.283 reflejan una dispersión moderada, ya que sus valores no son ni demasiado bajos (lo que implicaría muy poca variabilidad) ni excesivamente altos (lo que indicaría una amplia dispersión). Por su parte, el rango intercuartil de 0.806 señala que el 50% central de los datos se concentra en el intervalo $(1{,}8611; 2{,}6667)$, que es relativamente estrecho.

Tabla 4 – Desviación, varianza, y rango intercuartil de las variables numéricas del estudio

Variable	Desviación	Varianza	Rango inter cuartil
Tot_dias_esc	29,71	882,77	35,13
Prom_dias_esc	1,74	3,03	1,95
Tot_dias_noesc	24,93	621,54	32,27
Prom_dias_noesc	2,49	6,21	3,22
Nota_final	0,69	0,48	0,725
Nota_4_corte	0,77	0,60	1,1
Actitudes	0,53	0,28	0,805

La Ilustración 18 muestra las actitudes y las notas finales por género. En el gráfico de la izquierda, dedicado a las actitudes por género, se aprecia que las mujeres exhiben una mediana ligeramente menor en comparación con los hombres y un rango intercuartílico no muy compacto, el del género masculino muestra menor variabilidad en los datos. En el gráfico de la derecha, correspondiente a la distribución de las notas finales, se evidencia que la mediana de las calificaciones en el grupo femenino es más alta, mientras que el grupo masculino muestra menor variabilidad y notas más bajas.

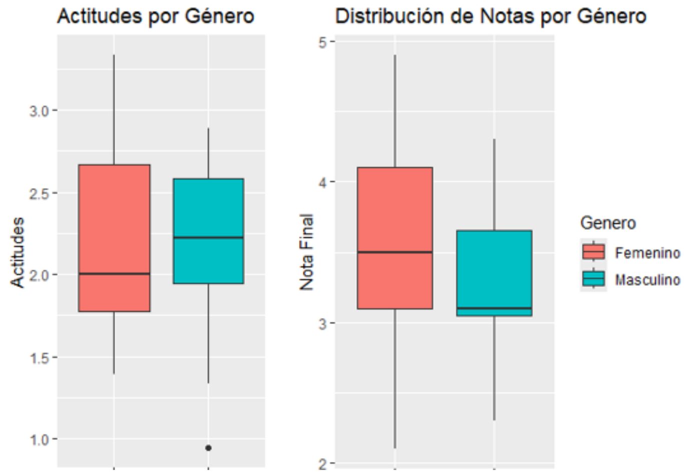

Ilustración 18 – Gráfico de caja de las actitudes y la nota final por genero

Así mismo, en la parte izquierda de la Ilustración 19 se presentan los diagramas de caja y bigotes correspondientes al promedio de horas diarias de uso del celular tanto en días escolares (Prom_dias_esc) como en días no escolares (Prom_dias_noesc). Allí se aprecia que, en general, el promedio de uso tiende a ser mayor en días no escolares y se identifica un valor atípico por encima del extremo superior en los días escolares.

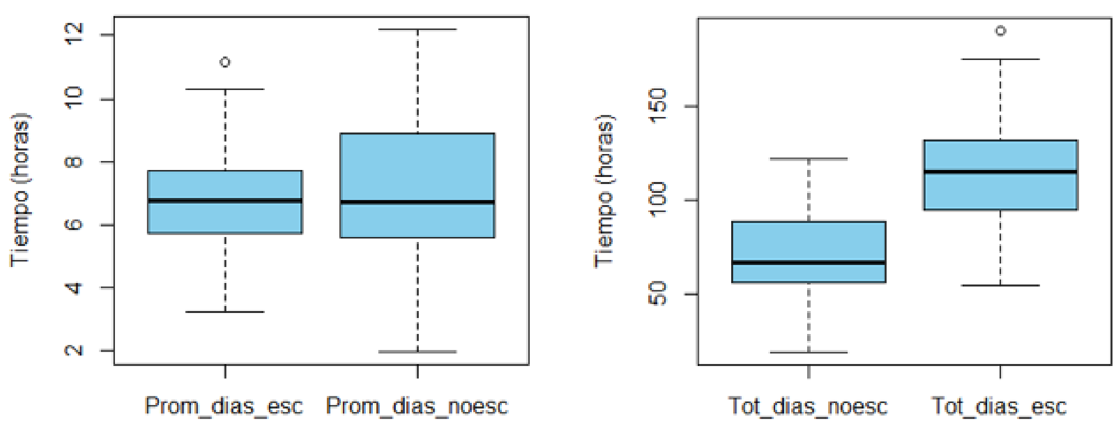

Ilustración 19 – Gráfico de caja de los promedios y el total de horas de uso

Por su parte, en el lado derecho se muestran los diagramas que representan el tiempo total, a lo largo de casi un mes de uso del celular durante días escolares (Tot_dias_esc) y días no escolares

(Tot_dias_noesc). En este caso, se observa que el uso total en días no escolares es inferior al de los días escolares.

En la parte izquierda de la Ilustración 20 se presentan los diagramas de caja y bigotes correspondientes a las puntuaciones de las actitudes hacia las matemáticas (Actitudes). Allí se aprecia, que la mediana se encuentra alrededor de 2,0, esto indica que no hay una tendencia clara a una actitud negativa o positiva, el 50 % de los datos están por encima y el 50% de los datos están por debajo.

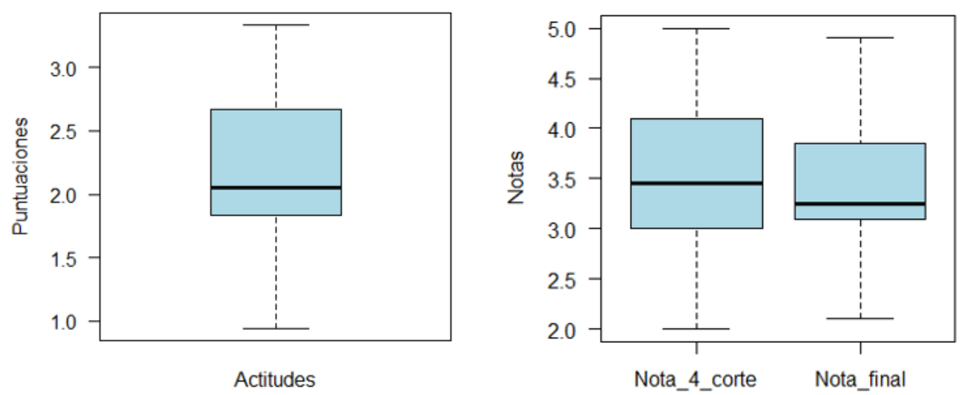

Ilustración 20 – Grafico de caja de las puntuaciones de actitudes y las notas de todos los estudiantes

Por su parte, en el lado derecho se muestran los diagramas que representan las calificaciones, la nota del cuarto periodo en matemáticas del año 2024 (Nota_4_corte) y la nota del final periodo en matemáticas del año 2024 (Nota_final). Ambas distribuciones presentan una mediana bastante cercana. Esto indica que en promedio las calificaciones en el cuarto periodo y las del final son similares. También, se puede apreciar que en las notas finales la caja (rango intercuartil) es más estrecha, lo que indica que la mayoría de los estudiantes tienden a concentrarse más cerca de la mediana.

En términos de Edad, los estudiantes tienen entre 14 y 17 años, con una mediana de 16 años y una media de 15.57 años. En la Ilustración 21 se puede visualizar la gráfica que muestra la distribución de las edades de los estudiantes en el conjunto de datos. Se observa que la mayoría de los estudiantes tienen 16 años, siendo este el grupo más numeroso. Las edades de 15 años y 14 años también tienen una representación considerable, mientras que los estudiantes de 17 años son los menos frecuentes.

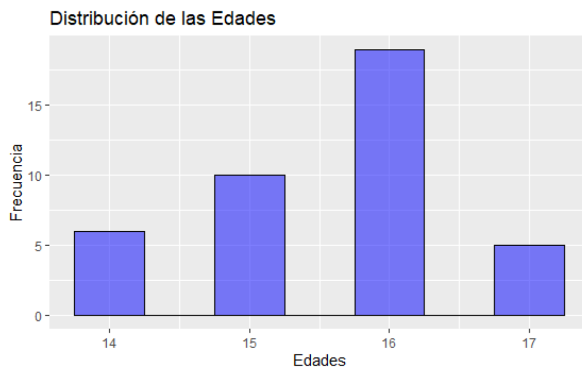

Ilustración 21 – Grafico de barras de la distribución de las edades

En cuanto al uso del celular en días escolares, el Prom_dias_esc tiene un mínimo de 3,241 horas y un máximo de 11,149 horas, con una media de 6,744 horas. En la Ilustración 22 se observa que la mayoría de los estudiantes usan el celular entre 5 y 8 horas durante los días de clase, con una concentración mayor alrededor de las 6 y 7 horas. Aunque hay algunos estudiantes que reportan un uso menor a 4 horas o superior a 10 horas, estos casos son menos frecuentes.

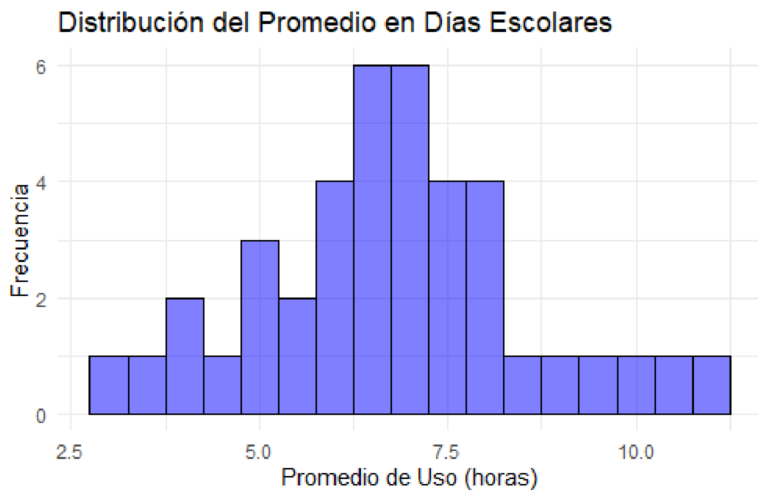

Ilustración 22 – Grafico de barras de la distribución del promedio de horas de uso del celular en días escolares

En la Ilustración 23 se observa que la mayoría de los estudiantes utilizan el celular entre 5 y 7 horas, con un pico de frecuencia alrededor de las 6 horas en días no escolares. Sin embargo, la distribución es más dispersa que en los días escolares, con algunos estudiantes reportando un uso de más de 10 horas y otros con menos de 3 horas.

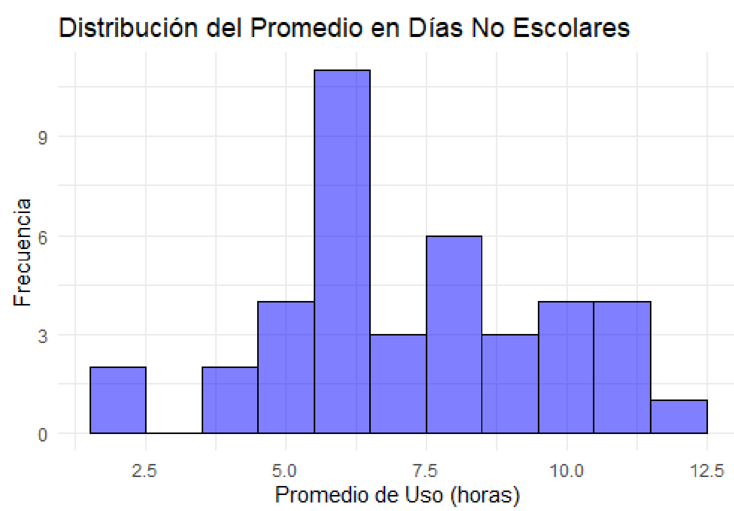

Ilustración 23 – Grafico de barras de la distribución de horas de uso del celular en días no escolares

En días no escolares, el Prom_dias_noesc presenta un promedio de 7,305 horas, con un mínimo de 1,980 y un máximo de 12,188 horas, lo que indica un aumento del uso del celular en comparación con los días escolares. Respecto al tiempo total de uso del celular, Tot_dias_noesc muestra un promedio de 71,80 horas, con una mediana de 67,27 horas y un máximo de 121,88 horas.

En cuanto a la nota final en matemáticas, los valores oscilan entre 2,1 y 4,9, con una mediana de 3,25 y una media de 3,475, lo que indica que la mayoría de los estudiantes tienen un rendimiento medio en esta asignatura. En la Ilustración 24 se observa que la mayoría de los estudiantes obtuvieron calificaciones en torno a 3.0, siendo este el rango con mayor frecuencia. También se pueden identificar grupos de estudiantes con notas más altas, cercanas a 4.0 y 5.0, aunque en menor cantidad.

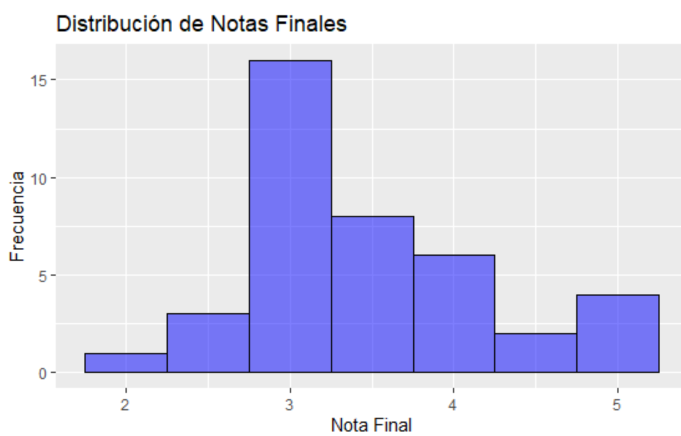

Ilustración 24 – Grafico de barras de la distribución de las notas finales de los estudiantes en matemáticas

Finalmente, la variable Actitudes, que mide la actitud hacia las matemáticas, varía entre 0,9444 y 3,3333, con una media de 2,1486, lo que podría sugerir una tendencia neutra en la percepción de los estudiantes hacia la materia. En la Ilustración 25 se observa que la mayoría de los estudiantes tienen un puntaje de actitud cercano a 2.0, con una alta concentración en este valor.

También hay un grupo considerable con puntajes entre 1.5 y 2.5, lo que indica una actitud mayormente neutral o ligeramente positiva hacia la materia.

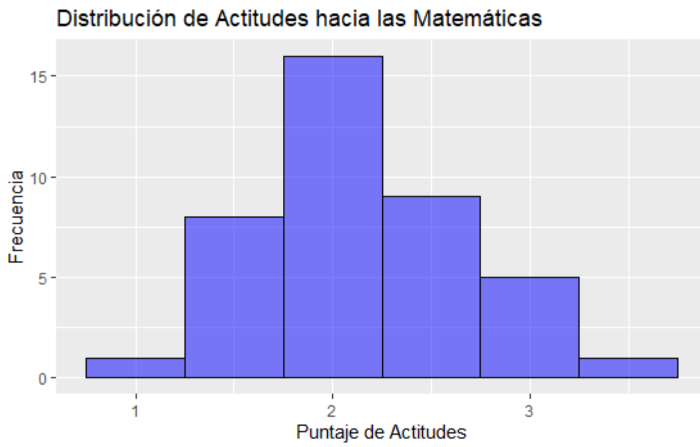

Ilustración 25 – Grafico de barras de la distribución del puntaje sobre las actitudes de los estudiantes hacia las matemáticas

4 CAPÍTULO IV. IMPLEMENTACIÓN DEL MODELO

En este capítulo se describen los resultados obtenidos del análisis utilizando R al conjunto de datos de 40 estudiantes, el código se presenta en el Anexo 5. Respecto al código cabe mencionar que este se realizó teniendo en cuenta las necesidades que se venían generando en el trabajo de grado, es decir, se fue modificando e implementando mientras se limpiaban los datos, se ajustaba el modelo de regresión lineal múltiple, se realizaban las pruebas para comprobar los supuestos y se creaban los gráficos necesarios para la explicación del análisis.

En este sentido, el código en R fue construido poco a poco con base en videos de YouTube, foros de internet como Stack Overflow y RPubs, de tal manera que se fue recopilando información y se revisaron múltiples formas de escribir el código para hacerlo más eficiente utilizando las librerías disponibles de R, optimizando el tiempo tanto en la escritura del código como en su depuración y posterior verificación de funcionalidad.

De esta manera se muestran las pruebas de normalidad (Shapiro-Wilk y Kolmogorov-Smirnov) aplicadas a los valores de las variables, se presentará la matriz de correlación de las variables y el gráfico de dispersión de cada par de variables. Luego de ello se incluye la variable *dummy* para el curso la cual toma el valor de uno, si el curso corresponde a décimo y de cero si el curso es octavo.

Luego de agregada la variable *dummy* y de haber escogido solo los valores numéricos, se realiza un modelo de regresión con todas las variables, para aplicarle a este modelo la prueba VIF ya que se sospecha de multicolinealidad. Con base en estos resultados se eliminan las variables referentes al total de horas que los estudiantes pasan en el teléfono y se vuelve a calcular la matriz de correlación para evidenciar si existe algún cambio entre la correlación de las variables que se mantienen.

Una vez eliminadas las variables se establece un modelo de regresión sin estas variables y se aplica la prueba VIF a este modelo, para corroborar que ya no se tienen problemas de multicolinealidad, posteriormente se utiliza el método de stepwise para determinar el mejor modelo matemático. Luego se revisa el contexto que rodea a este modelo y se llega a la conclusión de que existe una variable (Nota_4_corte) que por sí sola podría explicar la nota final de los estudiantes, una vez realizada la reflexión sobre el contexto del modelo y de las variables se toma la decisión de aplicar el método de *stepwise*, ahora sin la variable Nota_4_corte.

Una vez obtenido el mejor modelo utilizando el método *stepwise*, se procede con la verificación de los supuestos de este último modelo; para ello, se presenta un gráfico de correlación con las variables finales del modelo, para evidenciar la linealidad, posteriormente se revisa la homocedasticidad de los residuales utilizando el test de Breusch-Pagan, luego se aplican las pruebas de normalidad (Shapiro-Wilk y Kolmogórov-Smirnov) a los residuales, se continúa con el análisis de la independencia de los errores con la prueba de Durbin-Watson y por último se revisa que no exista multicolinealidad aplicando la prueba VIF al modelo final.

4.1 Establecer el modelo de regresión

Revisión de la normalidad de las variables

En un principio para establecer si los análisis se podrían realizar sin transformar las variables, se revisó la normalidad de cada variable numérica, para ello se utilizaron dos pruebas de normalidad, una fue la de Shapiro-Wilk y la otra prueba empleada se llama Kolmogorov-Smirnov, de esta manera se contrastaron los valores obtenidos y se revisó si con ambas pruebas se podría afirmar que las variables seguían una distribución normal.

Al realizar la prueba de Shapiro-Wilk, se obtuvieron los siguientes resultados:

```
✅ Tot_dias_esc: Distribuye normal (p = 0.7496)
✅ Prom_dias_esc: Distribuye normal (p = 0.7422)
✅ Tot_dias_noesc: Distribuye normal (p = 0.6381)
✅ Prom_dias_noesc: Distribuye normal (p = 0.6381)
✅ Nota_4_corte: Distribuye normal (p = 0.3015)
✅ Nota_final: Distribuye normal (p = 0.0627)
✅ Actitudes: Distribuye normal (p = 0.2707)
```

Ilustración 26 – Resultados de la prueba de normalidad de Shapiro-Wilk

Para contrastar estos resultados, se implementó la prueba de Kolmogorov-Smirnov, para revisar si con este test, se puede concluir que los valores de las variables distribuyen de manera normal. Al realizar la prueba en R se obtuvieron los siguientes resultados:

```
✅ Tot_dias_esc : Distribuye normal, con un p-value = 0.7784
✅ Prom_dias_esc : Distribuye normal, con un p-value = 0.7725
✅ Tot_dias_noesc : Distribuye normal, con un p-value = 0.7772
✅ Prom_dias_noesc : Distribuye normal, con un p-value = 0.7772
✅ Nota_4_corte : Distribuye normal, con un p-value = 0.7687
✅ Nota_final : Distribuye normal, con un p-value = 0.439
✅ Actitudes : Distribuye normal, con un p-value = 0.7187
```

Ilustración 27 – Resultados de la prueba de normalidad de Kolmogorov-Smirnov

Con base en estos resultados no hay evidencia para rechazar la distribución normal de las variables teniendo en cuenta un nivel de significancia del 0.05.

Correlación lineal de las variables

Una vez verificada la normalidad de las variables se procede a revisar la correlación lineal que existe entre estas, para ello se realizó un mapa de calor con base en la matriz de correlación de los datos, utilizando el coeficiente de Pearson. La matriz muestra cuáles correlaciones son positivas, negativas, y si es cercano a uno el cuadro aparece de un color azul y si es cercano a menos uno el cuadro aparece coloreado de rojo, y si la correlación es cercana a cero el cuadro se muestra de un color gris casi blanco (Ilustración 28).

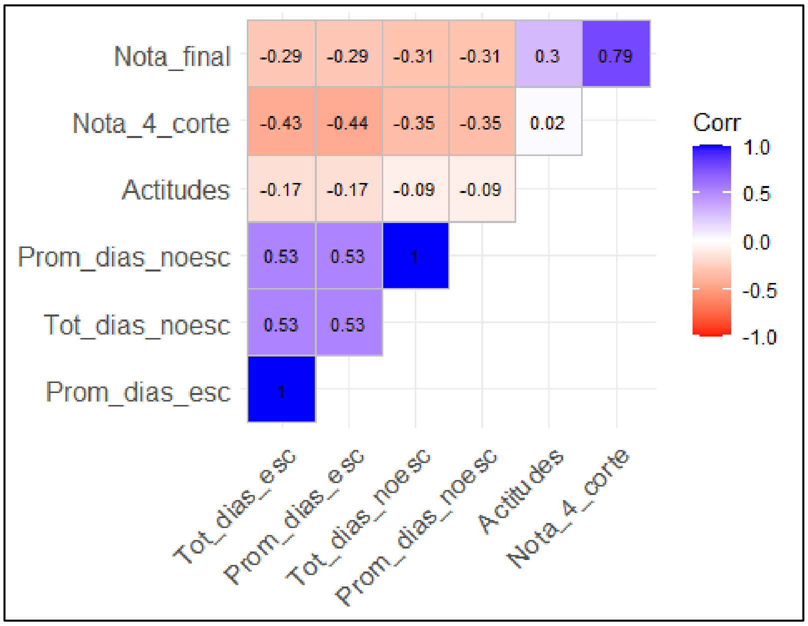

Ilustración 28 – Matriz y mapa de calor de la correlación entre las variables de estudio

Posteriormente, para revisar de una manera visual el comportamiento lineal de la correlación se realizó el gráfico de dispersión de las variables (Ilustración 29).

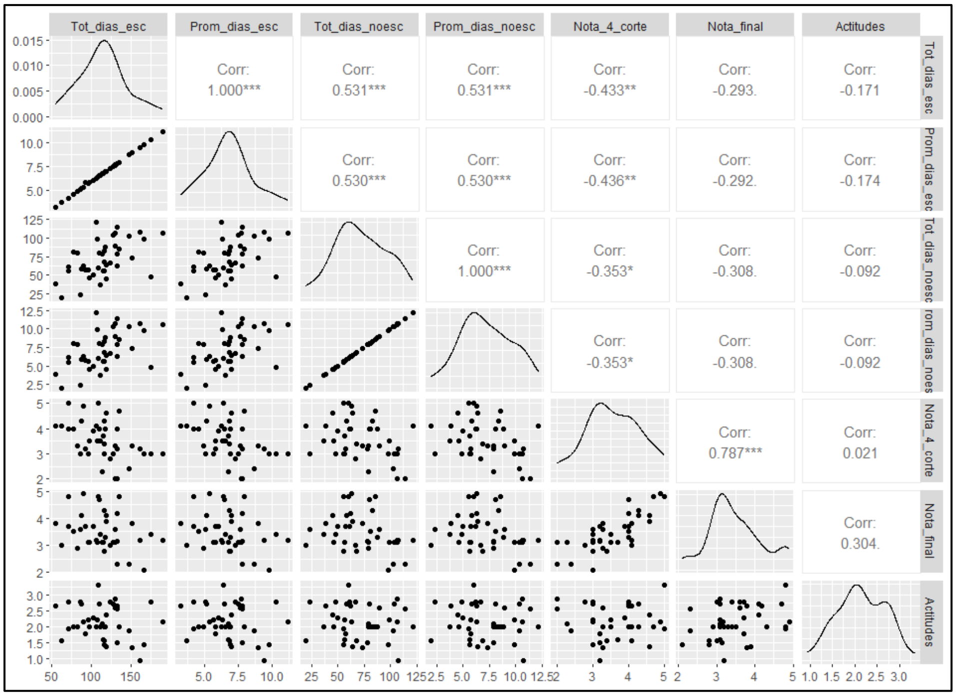

Ilustración 29 – Grafico de dispersión de las variables de estudio

Con este gráfico logramos evidenciar la dispersión de los datos y también nos damos una idea de la relación aparentemente lineal entre las variables independientes con la variable dependiente (Nota_final). Adicionalmente de visualizar una alta correlación entre algunas variables independientes lo cual nos advierte de posibles problemas de multicolinealidad.

Para atender el potencial problema de multicolinealidad se realiza la prueba VIF a un modelo de prueba, con el objetivo de cuantificar la multicolinealidad aparente entre las variables independientes. Para ello primero se incluyen en este modelo todas las variables disponibles en la base de datos, las cuales son: Tot_dias_esc, Prom_dias_esc, Tot_dias_noesc, Prom_dias_noesc, Nota_4_corte, Nota_final, Actitudes, y el Curso la cual se nombrará como Curso_dummy y tomará los valores de 0 para el curso de Octavo y 1 para el curso de Décimo. Dicho modelo de prueba queda de la siguiente manera (Ilustración 30):

```
Call:
lm(formula = Nota_final ~ ., data = df_compilado_clean)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.81864 -0.23398 -0.06139  0.21766  0.72306 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)       -0.55573    0.67222  -0.827  0.41453    
Curso_dummy        0.01368    0.16575   0.083  0.93472    
Tot_dias_esc      -0.11399    0.06937  -1.643  0.11012    
Prom_dias_esc      2.01382    1.18407   1.701  0.09869 .  
Tot_dias_noesc  -347.80009  677.18380  -0.514  0.61106    
Prom_dias_noesc 3477.98269 6771.83572   0.514  0.61106    
Nota_4_corte       0.75299    0.09582   7.859 5.76e-09 ***
Actitudes          0.44239    0.13365   3.310  0.00232 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3905 on 32 degrees of freedom
Multiple R-squared:  0.7435,	Adjusted R-squared:  0.6874 
F-statistic: 13.25 on 7 and 32 DF,  p-value: 7.025e-08
```

Ilustración 30 – Modelo de prueba con todas las variables de estudio

Una vez obtenido el modelo se procede a utilizar la prueba VIF, obteniendo los siguientes valores (Ilustración 31):

```
  Variables.independientes          VIF
1              Curso_dummy 1.580499e+00
2             Tot_dias_esc 1.086141e+03
3            Prom_dias_esc 1.086515e+03
4           Tot_dias_noesc 7.287848e+10
5          Prom_dias_noesc 7.287845e+10
6             Nota_4_corte 1.424369e+00
7                Actitudes 1.293636e+00
```

Ilustración 31 – Resultados de la prueba VIF aplicada al modelo de prueba

Con base en estos resultados se toma la decisión de eliminar del modelo las variables Tot_dias_esc y Tot_dias_noesc, las cuales tienen valores VIF de $1.086141 \cdot 10^3$ y $7.287848 \cdot 10^{10}$ respectivamente. Estos se valores se pueden considerar altos ya que como se mencionó en el apartado de Multicolinealidad y el test VIF, cuando el valor del factor de inflación de la varianza es mayor a 10 los datos tienen un problema de colinealidad, y una manera de solucionarlo es eliminar variables que generen dicho problema.

Una vez eliminadas estas variables, se calculó el coeficiente de correlación de Pearson para revisar si existían cambios en la correlación de las variables luego de eliminar las variables de los totales, obteniendo lo siguiente (Ilustración 32):

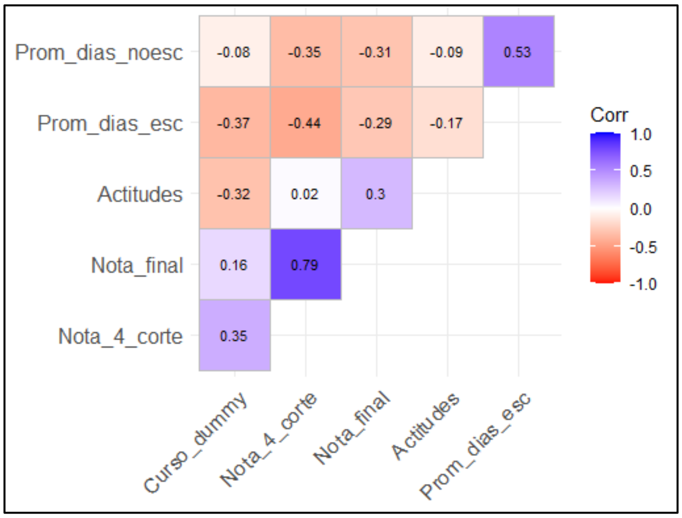

Ilustración 32 – Matriz y mapa de calor de la correlación entre las variables luego de eliminar las variables con el total de horas

Se evidencia con esta matriz que al quitar las variables mencionadas la correlación existente no cambia, lo cual nos permite afirmar que la correlación entre cada par de variables se mantiene a pesar de eliminar variables de nuestro modelo, con base en esto podemos seguir aplicando cambios al modelo sin afectar el valor de la correlación y la linealidad aparente entre las variables independientes y la variable dependiente.

Una vez evidenciada que la correlación no cambia, se procede a eliminar las variables de los totales como se había contemplado, y se obtiene un modelo de prueba 2 al cual se le aplica la prueba VIF, con el objetivo de obtener valores menores a 10 y así lograr superar el problema de la multicolinealidad, obteniendo lo siguiente (Ilustración 33).

```
  Variables.independientes      VIF
1              Curso_dummy 1.549775
2            Prom_dias_esc 1.885017
3          Prom_dias_noesc 1.491913
4              Nota_4_corte 1.360356
5                Actitudes 1.266444
```

Ilustración 33 – Resultados de la prueba VIF aplicada al modelo sin las variables de los totales

Con base en estos resultados podemos afirmar que se ha superado el problema de la multicolinealidad y por lo tanto se considera que las variables explicativas son independientes entre ellas. Con base en este resultado se procede a seleccionar el mejor modelo disponible utilizando el método de *stepwise*, obteniendo que el mejor modelo matemáticamente es el siguiente (Ilustración 34):

```
> summary(modelo_1)

Call:
lm(formula = Nota_final ~ Nota_4_corte + Actitudes, data = df_compilado_clean %>%
    select(-c(Tot_dias_noesc, Tot_dias_esc)))

Residuals:
     Min       1Q   Median       3Q      Max 
-0.72042 -0.26913 -0.04279  0.29751  0.81129 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)   0.17404    0.38411   0.453   0.6531    
Nota_4_corte  0.70060    0.08042   8.712 1.71e-10 ***
Actitudes     0.37701    0.11770   3.203   0.0028 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3911 on 37 degrees of freedom
Multiple R-squared:  0.7026,	Adjusted R-squared:  0.6865 
F-statistic:  43.7 on 2 and 37 DF,  p-value: 1.809e-10
```

Ilustración 34 – Modelo_1 obtenido con el metodo Stepwise

Cuya ecuación está dada por la expresión:

$$\widehat{\text{Nota final}} = 0.70060 \times \widehat{\text{Nota\_4\_corte}} + 0.37701 \times \widehat{\text{Actitudes}} + 0.17404 \qquad (14)$$

Este modelo predice la Nota final de los estudiantes en matemáticas, con base en dos variables, la nota del cuarto corte y las actitudes de los estudiantes hacia las matemáticas. Se puede interpretar que, por cada unidad adicional en la nota del cuarto corte, se espera que la nota final aumente 0.7 unidades, manteniendo constante las actitudes, por su parte por cada punto adicional en las actitudes, la nota final aumenta en 0.37 unidades, siempre y cuando la nota del cuarto corte se mantenga constante.

Con base en este resultado, se logra apreciar que la variable más significativa es la de Nota_4_corte, y aunque esta variable muestra una alta correlación con la variable Nota_final, al revisar el contexto de los datos nos damos cuenta que esto se debe a que Nota_final es el resultado del promedio de las notas de los cortes anteriores; por lo tanto, si se tiene la variable Nota_4_corte está por si sola explica la nota de los estudiantes al final del año académico, lo cual es algo que no nos permitiría analizar otros posibles factores que influyen en la nota final de los estudiantes como el tiempo que usan el celular o el curso al que pertenecen, por lo tanto se decidió quitar esta variable para formular otro modelo de regresión múltiple.

Luego se volvió a implementar el método de *stepwise* para determinar el mejor modelo quitando la variable de Nota_4_corte, como se había mencionado previamente, obteniendo lo siguiente (Ilustración 35).

```
> summary(modelo_2)

Call:
lm(formula = Nota_final ~ Prom_dias_noesc + Actitudes + Curso_dummy,
    data = df_compilado_clean %>% select(-c(Tot_dias_noesc, Tot_dias_esc,
        Nota_4_corte)))

Residuals:
    Min      1Q  Median      3Q     Max
-0.9888 -0.4732 -0.1868  0.4466  1.3543

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      2.73428    0.63577   4.301 0.000124 ***
Prom_dias_noesc -0.07163    0.04146  -1.728 0.092639 .
Actitudes        0.47525    0.20463   2.323 0.025974 *
Curso_dummy      0.37205    0.22929   1.623 0.113401
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6387 on 36 degrees of freedom
Multiple R-squared:  0.2282,    Adjusted R-squared:  0.1639
F-statistic: 3.549 on 3 and 36 DF,  p-value: 0.02385
```

Ilustración 35 – Modelo_2 obtenido con el metodo Stepwise sin la variable Nota_4_corte

Cuya ecuación es:

$$\widehat{\text{Nota final}} = 0.4752 \times \widehat{\text{Actitudes}} - Prom\_dias\_noesc \times 0.0716 + \text{Curso\_dummy} \times 0.372 + 2.734 \qquad (15)$$

Con base en la ecuación del modelo se puede concluir que por cada punto adicional en las actitudes la nota final aumenta en 0.47 unidades, manteniendo las demás variables constantes; en cuanto al uso del celular en los días no escolares, se tiene que por cada hora adicional que los estudiantes pasan en el celular su nota disminuye en 0.07 unidades, por su parte si el estudiante está en decimo (Curso_dummy = 1) en lugar de estar en Octavo (Curso_dummy = 0), la nota final en matematicas aumenta en 0.37 unidades, asumiendo que las otras variables no cambian.

Ahora con este modelo podemos evidenciar que se puede explicar la nota de los estudiantes al finalizar el año escolar teniendo en cuenta el curso al que pertenecen, el total de horas que pasan en el celular los fines de semana y festivos, y contemplando la actitud que tienen hacia las matemáticas, por lo tanto, se obtiene un modelo con múltiples parámetros que permiten revisar que otros aspectos influyen en su nota, sin tener en cuenta la nota del último corte.

4.2 Comprobación los supuestos del modelo de regresión

Una vez obtenido el modelo de la expresión (15) se procede a verificar cada uno de los supuestos, para ello se agrega una columna con los residuales (la diferencia entre la nota final y las predicciones) y se aplican algunos métodos que permiten evidenciar el cumplimiento o no de los siguientes supuestos:

1. El vector de errores ϵ sigue una distribución normal multivariada, cuyo vector de medias o valor esperado es cero, es decir $E(\epsilon) = 0$
2. La varianza de ϵ denotada como σ^2 es constante, es decir es la misma para todos los valores de las variables independientes x_1, x_2, ..., x_p.
3. Los valores de ϵ son independientes entre sí, por lo tanto, el valor de ϵ para un determinado conjunto de valores de las variables independientes no está relacionado con el valor de ϵ de ningún otro conjunto de valores.
4. Las variables independientes se relacionan de manera lineal con la variable dependiente.
5. Las variables explicativas (independientes) son linealmente independientes entre sí.

Normalidad

Primero se revisa la normalidad de los errores, obteniendo que la media tiene como valor: -0.0000000000000047739, evidenciando que es bastante cercana a cero. Luego para visualizar la distribución de los residuales se realizó el siguiente histograma:

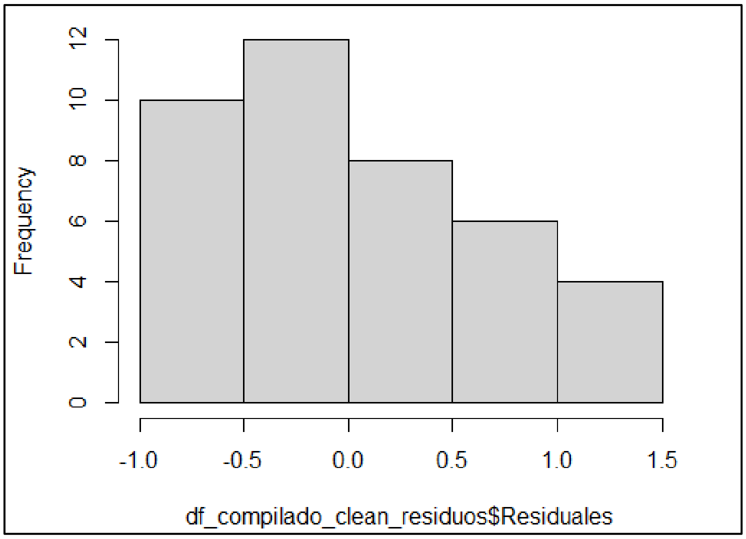

Ilustración 36 – Grafico de barras de la distribución de los residuales

Con base en este gráfico se puede deducir que la distribución puede o no ser normal, por lo tanto, es necesario hacer las pruebas de Shapiro-Wilk y Kolmogórov-Smirnov, para determinar si se cumple el supuesto de normalidad. Primero se realizó la prueba de Shapiro-Wilk obteniendo los siguientes resultados (Ilustración 37):

```
        Shapiro-Wilk normality test

data:  df_compilado_clean_residuos$Residuales
W = 0.93574, p-value = 0.0249
```

Ilustración 37 – Resultados de la prueba de Shapiro-Wilk aplicada a los residuales del modelo_2

Con base en este resultado, y con un nivel de significancia de 0.05 los residuos no seguirían una distribución normal; sin embargo, se decide aplicar la prueba de Kolmogórov-Smirnov para contrastar dicha afirmación, obteniendo lo siguiente (Ilustración 38):

```
        Exact one-sample Kolmogorov-Smirnov test

data:  scale(df_compilado_clean_residuos$Residuales)
D = 0.15398, p-value = 0.2702
alternative hypothesis: two-sided
```

Ilustración 38 – Resultado de la prueba de Kolmogorov-Smirnov aplicada a los residuales del modelo_2

Al revisar esta prueba se puede afirmar que los residuales sí siguen una distribución normal con un nivel de significancia de 0.05, ya que el $p - valor$ es mayor al nivel de significancia, por lo tanto, no se rechaza la hipótesis nula de distribución normal con base en la prueba de Kolmogórov-Smirnov.

Homocedasticidad

Para verificar el segundo supuesto se realiza primero una exploración visual, graficando los residuales contra los valores pronosticados, como se muestra a continuación (Ilustración 39):

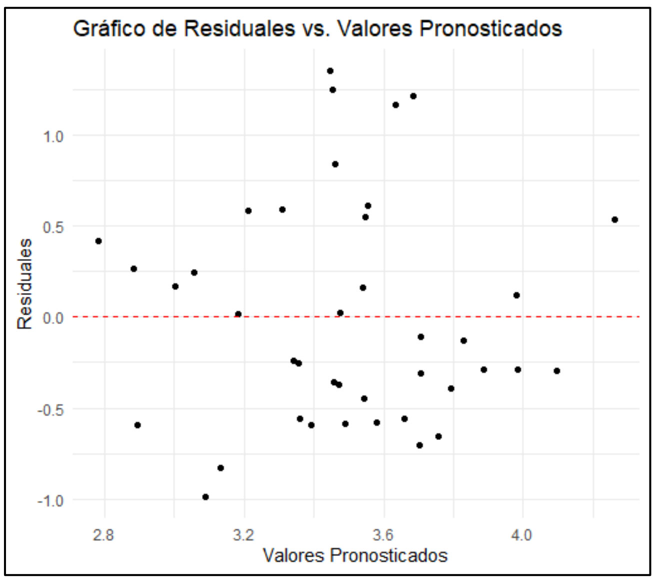

Ilustración 39 – Grafico de Homocedasticidad del modelo_2

Visualmente pareciera que en efecto la varianza de ϵ es la misma para todos los valores de las variables independientes, sin embargo, para verificar ello se realiza la prueba de Breusch-Pagan, obteniendo lo siguiente (Ilustración 40):

```
        studentized Breusch-Pagan test

data:  modelo_2
BP = 0.49699, df = 3, p-value = 0.9196
```

Ilustración 40 – Resultados de la prueba de Breush-Pagan aplicada a los residuales del modelo_2

Por lo tanto, como el $p - valor$ no es menor al nivel de significancia de 0.05 no se rechaza la hipótesis de varianza constante, cumpliendo así el segundo supuesto de la regresión múltiple.

Independencia

Se continúa revisando la independencia de los errores, para ello se utiliza el test de Durbin–Watson, el cual al aplicarlo en R se obtiene lo siguiente (Ilustración 41):

```
        Durbin-Watson test

data:  modelo_2
DW = 1.9631, p-value = 0.3727
alternative hypothesis: true autocorrelation is greater than 0
```

Ilustración 41 – Resultados del test de Durbin-Watson aplicado a los residuales del modelo_2

Como el valor del estadístico es cercano a 2 se puede sospechar que sí se cumple la independencia, lo cual se confirma al analizar el $p - valor$, ya que este es mayor al nivel de significancia de 0.05, por lo tanto, esta prueba permite afirmar que no hay evidencia estadística para asegurar que existe autocorrelación, por lo tanto, se concluye que los residuales son independientes entre sí.

Linealidad con la variable dependiente

Para visualizar esta linealidad se realiza el gráfico de dispersión para las variables que se mantienen en el modelo, Promdias_noesc y Actitudes con la variable Nota_final, obteniendo el siguiente gráfico (Ilustración 42):

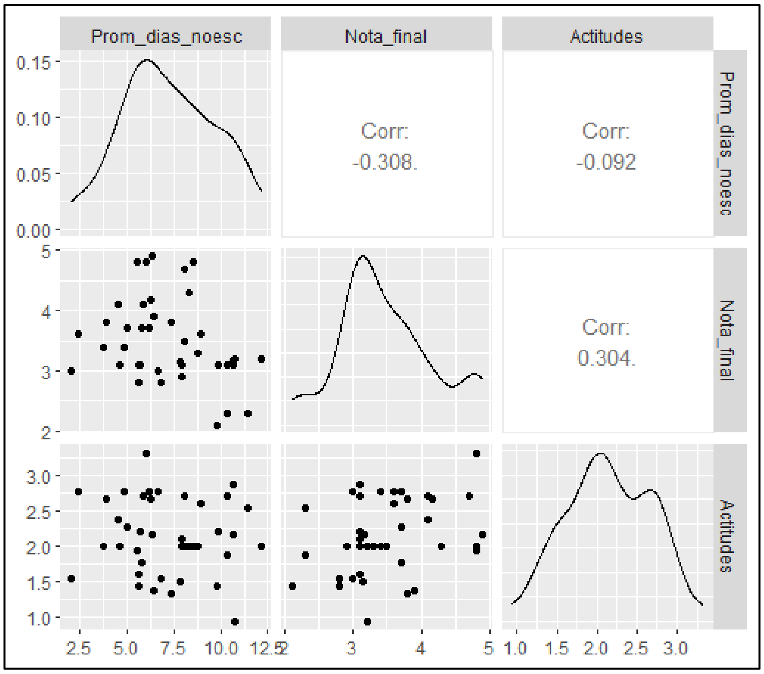

Ilustración 42 – Grafico de dispersión de las variables del modelo_2

Al visualizar este gráfico se puede apreciar que se mantiene la relación lineal observada, que existe entre las variables independientes con la variable dependiente.

Independencia entre las variables explicativas

Para finalizar se revisa la independencia entre las variables Prom_dias_noesc, y Actitudes utilizando el test VIF, obteniendo los siguientes resultados (Ilustración 43).

	Variables.independientes	VIF
1	Prom_dias_noesc	1.021679
2	Actitudes	1.133804
3	Curso_dummy	1.130989

Ilustración 43 – Resultados de la prueba VIF aplicada al modelo_2

Obteniendo que los valores son menores a 10 y por lo tanto se cumple la independencia entre las variables es decir no existe multicolinealidad.

4.3 Análisis del modelo final de regresión

Para finalizar se analizará el modelo (expresión 15) final obtenido con el método de *stepwise*, cuya ecuación de regresión es:

$$\widehat{\text{Nota final}} = 0.4752 \times \widehat{\text{Actitudes}} - Prom\_dias\_noesc \times 0.0716 + \text{Curso\_dummy} \times 0.372 + 2.734$$

Y cuyos resultados en R se presentan en la Ilustración 44:

```
> summary(modelo_2)

Call:
lm(formula = Nota_final ~ Prom_dias_noesc + Actitudes + Curso_dummy,
    data = df_compilado_clean %>% select(-c(Tot_dias_noesc, Tot_dias_esc,
        Nota_4_corte)))

Residuals:
    Min      1Q  Median      3Q     Max 
-0.9888 -0.4732 -0.1868  0.4466  1.3543 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)      2.73428    0.63577   4.301 0.000124 ***
Prom_dias_noesc -0.07163    0.04146  -1.728 0.092639 .  
Actitudes        0.47525    0.20463   2.323 0.025974 *  
Curso_dummy      0.37205    0.22929   1.623 0.113401    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6387 on 36 degrees of freedom
Multiple R-squared:  0.2282,	Adjusted R-squared:  0.1639 
F-statistic: 3.549 on 3 and 36 DF,  p-value: 0.02385
```

Ilustración 44 – Resumen de los principales valores a interpretar del modelo_2

De este modelo se pueden analizar los múltiples valores que nos otorga R, para ello a continuación se presentara con detalle una descripción de los residuales, la interpretación de los coeficientes, el error estándar, el $p-value$ asociado con el valor t para cada parámetro de la ecuación de regresión, el valor del estimador R^2 y el R^2 ajustado, además del valor $f-statistic$ y el $p-value$ asociado al modelo de regresión.

Respecto a los residuales se tiene que la mediana es cercana a cero con un valor de -0.1868, lo cual nos da una idea de que los datos están centrados alrededor del cero. Los valores mínimo y máximo no están alejados de -1 y de 1 respectivamente, lo cual indica que no hay valores atípicos extremos, sin embargo, se evidencia que la distribución no es perfectamente simétrica ya que la mediana, aunque cercana a cero no es cero y a pesar de no tener valores atípicos extremos se logra apreciar una asimetría positiva, como lo sugería el histograma en el supuesto de normalidad.

Respecto a los coeficientes se puede mencionar que, si los estudiantes usan el celular durante sus días no escolares, por cada hora adicional que utilizan el celular la nota final de matemáticas disminuye en 0.07 puntos, y el efecto que esta variable tiene es no es significativo si se tiene en cuenta un nivel de significancia de 0.05, ya que el $p - value$ asociado es de 0.092.

Respecto al coeficiente de la variable $Actitudes$, cuyo valor es 0.47525, se puede concluir que una actitud positiva hacia las matemáticas hace que la nota en esta materia aumente en 0.47 puntos, esto se puede afirmar con un nivel de significancia de 0.05 ya que el $p - value$ asociado es de 0.025, de esta manera se tiene solo un 5% de probabilidad de que dicha afirmación sea falsa.

Por su parte el coeficiente del curso es 0.37205 esto indica que como la variable curso es una variable dummy que toma solo dos valores 0 y 1, cuando la variable toma el valor de 1 los estudiantes representados por este valor, es decir los estudiantes de décimo, tienen una nota en matemáticas de 0.37 puntos más alta que los estudiantes de grado octavo. Sin embargo, esta variable no es significativa con un nivel de significancia de 0.05 pero es significativa con un Nivel de significancia de 0.12, ya que el $p - value$ asociado es de 0.113.

Posteriormente se analizó el valor del error estándar para cada variable, ya que estos valores permiten determinar la precisión que tiene el modelo al estimar los valores desconocidos β de los coeficientes, para ello se calcularon los intervalos de confianza al 95% de confianza para establecer un margen en el cual el verdadero valor del coeficiente poblacional podría estar. En este sentido se utilizó la función de R $confint()$ con la cual se obtuvieron los siguientes intervalos de confianza.

	5 %	95 %
(Intercept)	1.66091839	3.807646583
Prom_dias_noesc	-0.14163636	-0.001627935
Actitudes	0.12977806	0.820719520

Ilustración 45 – Intervalos de confianza para los valores de las variables del modelo

Como estos intervalos no contienen al cero se puede afirmar que el efecto de estas variables es significativo para estimar la nota final en matemáticas.

Por otra parte, se tiene que el error estándar residual del modelo lineal es de 0.63, por lo tanto, como este valor es un estimador de la desviación estándar del error (de los residuales), se puede afirmar que en promedio las predicciones se desvían en 0.63 puntos, es decir, si por ejemplo la nota final en matemáticas de un estudiante fuera de 2, la predicción podría ser de 1.37 o de 2.63.

También es interesante analizar el $p - value$ del modelo de regresión cuyo valor es 0.0238, ya que con este valor se puede efectuar en análisis de la prueba F, para contrastar la hipótesis nula de que $H_0: \beta_1 = \beta_2 = \cdots = \beta_p = 0$ conta la alternativa $H_a: \beta_{\text{i}} \neq 0$, para ello se utilizara la regla de rechazo de la hipótesis nula, la cual dice que se rechaza H_0 si el $p - value \leq \alpha$, en este caso se puede evidenciar que $0.0238 < 0.05$, por lo tanto se rechaza la hipótesis nula y se puede afirmar que hay suficiente evidencia estadística para concluir que uno o más parámetros no es igual a cero y existe una relación global entre la variable dependiente y el conjunto de las variables explicativas.

Para finalizar se analizan los valores R^2 y su valor ajustado R_a^2, los cuales son 0.2282 y 0.1639 respectivamente. Si se multiplica el R^2 por cien se puede afirmar que el modelo explica 22% de la variabilidad en la nota final de matemáticas, sin embargo, como se mencionó en el apartado del modelo de regresión lineal múltiple, agregar variables independientes al modelo hace que el valor del R^2 aumente, por lo tanto, es recomendable analizar el R_a^2.

En este sentido el valor del R_a^2 multiplicado por cien dice que el modelo explica 16% de la variabilidad de la nota final en matemáticas, lo cual puede parecer un porcentaje bajo, sin embargo, como se concluyó en el análisis del $p - value$ del modelo con la prueba f, el modelo regresión es significativo de manera global, lo que permite evidenciar que, aunque estas variables tienen cierta influencia en los resultados de los estudiantes, no son las únicas variables que intervienen en el desempeño académico.

En este orden de ideas se considera que aunque el tiempo de uso del celular, las actitudes que los estudiantes tienen hacia las matemáticas y el curso al que pertenecen influyen en el desempeño en matemáticas, deben existir otros factores, como el ambiente familiar, la salud, las amistades, las horas de sueño, el desarrollo cognitivo, y otras variables que explican los resultados en matemáticas, los cuales no fueron contemplados en este estudio; sin embargo podrían revisarse para complementar y descubrir cuáles influyen y en qué medida en el desempeño académico en matemáticas.

5 CAPITULO V. CONCLUSIONES

Al contrastar los objetivos inicialmente planteados con los resultados alcanzados en este trabajo de grado, se observa que, la revisión de los marcos de referencia permitió afirmar que el celular, si bien se ha integrado como una herramienta esencial en la vida cotidiana, representa un factor ambivalente en el contexto educativo. Su uso excesivo y sin control ha demostrado tener efectos perjudiciales sobre el rendimiento académico, afectando de manera directa la concentración, los hábitos de estudio, la memoria, la calidad del sueño, la salud física y emocional, así como las relaciones interpersonales, todos ellos elementos clave en el proceso de aprendizaje.

No obstante, también se observa en los referentes estudiados que, cuando se utiliza con fines pedagógicos y de manera planificada, el celular puede convertirse en un recurso educativo valioso. En estos casos, favorece la participación activa de los estudiantes, promueve la colaboración, facilita el acceso a información en tiempo real y contribuye significativamente al desarrollo de competencias, lo que, en última instancia, puede traducirse en una mejora del rendimiento académico.

Con relación al segundo objetivo específico, el uso de la programación y de software especializado como R se consolida como una herramienta fundamental en la realización de estudios estadísticos, especialmente cuando se trabaja con bases de datos que requieren organización, transformación y análisis riguroso. En el presente estudio, que contó con 40 observaciones, el empleo de R permitió no solo una depuración eficiente de los datos, sino también una visualización clara mediante gráficos que facilitaron la interpretación de los resultados y la aplicación precisa de pruebas estadísticas. Esta experiencia evidencia que, a medida que se incrementa el volumen de datos, el uso de herramientas como R no solo se vuelve útil, sino imprescindible para garantizar un análisis robusto, ágil y confiable

Adicionalmente, la elección y ajuste del modelo estadístico más adecuado se vio ampliamente beneficiada gracias al uso de R, ya que el método *stepwise*, que implica la evaluación de todas las combinaciones de variables para identificar el modelo más eficiente habría sido extremadamente engorroso y poco práctico de realizar de forma manual. Contar con esta herramienta permitió automatizar el proceso, reduciendo significativamente el tiempo de trabajo y asegurando resultados precisos. De esta manera, el uso de R se consolidó como un recurso fundamental para la correcta aplicación de los métodos estadísticos, optimizando tanto la comprensión de los datos como la construcción y validación del modelo final.

Respecto al tercer objetivo específico, la regresión lineal múltiple constituye una técnica adecuada que es pertinente para analizar la relación entre el rendimiento académico de los estudiantes y múltiples variables que pueden influir en este, ya sean numéricas o categóricas. Su aplicación, respaldada por los lineamientos de Hair et al. (2019) y otros autores, permite evaluar de manera simultánea el efecto de diversas variables independientes sobre una variable dependiente, proporcionando una visión completa de la variable de interés. Si bien esta técnica ofreció resultados apropiados, significativos y estructurados, no debe considerarse como la única alternativa. El análisis de datos puede beneficiarse del uso de otros enfoques estadísticos o modelos más complejos, que podrían generar interpretaciones más profundas o mejorar la capacidad predictiva del estudio.

En cuanto al cuarto objetivo, se logró identificar que en el modelo final la variable actitudes era significativa considerando un nivel de significancia de 0.05, de tal manera que la nota final aumenta en 0.47 unidades, cuando las demás variables se mantienen constantes, respecto al uso promedio de horas en los días no escolares, se logró evidenciar que esta variable no es significativa utilizando el mismo nivel de confianza, igualmente para la variable curso. Sin embargo, el modelo en su conjunto es significativo ya que pasa la prueba f, por lo tanto, se pueden realizar predicciones para el conjunto de datos utilizando este modelo.

Se logró evidenciar qué el curso al que pertenecen los estudiantes, las horas dedicadas al uso del celular en fines de semana y festivos, y la actitud hacia las matemáticas, explican en conjunto una parte de la variabilidad en las calificaciones finales. Además, el modelo cumplió con los supuestos fundamentales de la regresión lineal múltiple: normalidad, homocedasticidad, independencia de los errores, linealidad de las variables independientes con la dependiente y ausencia de multicolinealidad, lo que garantiza la validez de las inferencias realizadas.

El modelo resultó estadísticamente significativo y permitió identificar algunos factores que influyen en el rendimiento académico, el coeficiente de determinación ajustado (R^2 ajustado) fue de 0.1639, por lo tanto, se puede afirmar que el 16% de la variabilidad de la nota final en matemáticas al finalizar el año escolar es explicada por las variables Actitudes, Curso y Promedio de horas que los estudiantes pasan en el celular los días no escolares. Esto indica que existe una proporción considerable de la variabilidad en las calificaciones que no es explicada por las variables incluidas en el modelo, lo que sugiere la influencia de otros factores que no fueron contemplados en este estudio.

Esta situación es comprensible si se considera que el rendimiento académico de un estudiante está determinado por múltiples aspectos, más allá de los analizados en esta investigación. Factores como la alimentación, el ambiente familiar, la calidad de la enseñanza o incluso la relación interpersonal con el docente, entre otros, pueden afectar significativamente el desempeño de un estudiante y explicar parte de la variabilidad no capturada por el modelo.

Como apreciaciones finales se logró evidenciar que en el estudio la variable de Actitudes tiene una influencia mayor que las otras variables ya que si un estudiante aumenta en una unidad su percepción positiva hacia las matemáticas su nota aumentaría en 0.47 puntos, lo cual es algo considerable, porque podría pasar de una nota reprobatoria con valor 2.65 y desempeño bajo a una nota de 3.12 aprobando así la materia con un desempeño básico.

Adicionalmente, como se observó en la expresión 14, el análisis la nota del último corte (Nota_4_corte) está altamente correlacionada con la nota final con un coeficiente de 0.79, y esta variable junto con la de actitudes generaban un modelo que explicaba el 68% de la variabilidad de la nota final en matemáticas, en este sentido surgen como propuesta para nuevos trabajos de grados hacer un estudio que logre responder a ¿Cuál corte influye más en la nota final en matemáticas? ¿Es la misma influencia para todos los cursos? ¿En qué medida influye cada corte?

Otro aspecto interesante ocurrió con el modelo representado en la ecuación (14), este presentó un alto coeficiente de determinación ajustado y un nivel de significancia estadística considerable, pero su utilidad para los fines del presente estudio fue limitada. A pesar de su solidez matemática, dicho modelo se sustentaba principalmente en la variable Nota_4_corte, la cual tiene una relación directa con la nota final, ya que esta última se construye a partir del promedio de los cuatro cortes. Esto significa que, aunque el modelo era altamente predictivo, no ofrecía una comprensión real de los factores adicionales que inciden en el rendimiento académico.

Esta situación evidencia que, en este tipo de investigaciones, no basta con adoptar el modelo que arroje mejores indicadores estadísticos desde el software, sino que es indispensable considerar el contexto y el sentido de los datos. La validez y utilidad de un modelo no radican únicamente en su capacidad predictiva o en sus métricas numéricas, sino también en su coherencia tanto con las variables que se incluyen como con los objetivos del estudio.

Por otro lado, una de las principales limitaciones del presente estudio radica en el tipo de muestreo empleado. Al haberse optado por un muestreo no probabilístico por conveniencia, los resultados obtenidos no pueden generalizarse a toda la población estudiantil. La selección de participantes se limitó a estudiantes de algunos cursos de octavo y décimo grado que cumplían con ciertos criterios, lo cual restringe la representatividad de la muestra. En este sentido, se invita a realizar estudios similares a este con un diseño metodológico que contemple un muestreo probabilístico aleatorio. Este tipo de muestreo permitiría una mayor validez externa, al garantizar que todos los estudiantes tengan la misma probabilidad de ser seleccionados, lo cual fortalecería la precisión y generalización de los hallazgos.

Por otra parte, se hace un llamado a reflexionar sobre como el tiempo que los estudiantes pasan en el celular los días no escolares puede disminuir sus notas, ya que una hipótesis que se puede contemplar es que aquellos estudiantes que usan menos sus celulares obtienen mejores notas porque usan ese tiempo para repasar los temas vistos, o para aprender de manera autónoma, lo cual sería interesante averiguar en un estudio que complemente los resultados obtenidos en este trabajo de grado.

Luego de analizar el modelo de regresión lineal, se considera que la técnica implementada fue adecuada para los fines propuestos, permitiendo no solo explicar relaciones relevantes entre variables, sino también abrir la posibilidad de futuras investigaciones que integren otros factores que puedan incidir en el rendimiento académico de los estudiantes.

Finalmente, queremos expresar las razones del por qué este tipo de trabajos son apropiados para una licenciatura en matemáticas. Aunque este trabajo tiene un enfoque estadístico, no puede desligarse de la pedagógica. Desde nuestra perspectiva como futuros licenciados en matemáticas, nos interesa comprender factores que inciden directamente en el aprendizaje, como el rendimiento académico y las actitudes de los estudiantes hacia la asignatura. Estas variables, más allá de su tratamiento cuantitativo, ofrecen información clave para el ejercicio docente y permiten tomar decisiones fundamentadas en procesos de enseñanza y aprendizaje.

También, consideramos que los docentes de matemáticas deben contar con una formación rigurosa, que no se limite a los contenidos escolares que se enseñan en secundaria, sino que les permita acceder a niveles superiores. Un profesor con una sólida preparación conceptual está mejor equipado para anticipar y abordar las dificultades que pueden surgir en el aprendizaje de sus estudiantes. De igual forma, le permite crear y diseñar mejores propuestas de enseñanza más apropiadas.

Por último, creemos firmemente que el rol del profesor no se agota en la enseñanza de contenidos, sino que implica también una actitud permanente de aprendizaje. El docente que investiga, que se cuestiona y que se interesa por profundizar en líneas temáticas que le llaman la atención, cada día va mejorando su quehacer. El presente trabajo de grado nace del interés académico por explorar y profundizar en el campo de la estadística multivariada, un área que consideramos fundamental y que nos genera interés.

REFERENCIAS

Anderson, D. (2008). *Estadística para administración y economía* (10ª. Ed). Thomson South-Western.

Auzmendi, E. (1992). Las actitudes hacia la matemática-estadística en las enseñanzas media y universitaria. Características y medición, 59-119.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. MacMillan.

Baron, N. S. (2016). Only connect: What the internet might be doing to us.

Bazán, J., Espinosa, G., y Farro, C. (2002). Rendimiento y actitudes hacia la matemática en el sistema escolar peruano.

Besolí, G., Palomas, N., y Chamarro, A. (2018). Uso del móvil en padres, niños y adolescentes: Creencias acerca de sus riesgos y beneficios. *Aloma: Revista de Psicologia, Ciències de l'Educació i de l'Esport*, *36*(1), 29-39.

Breusch, T. S., y Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. Econometrica: Journal of the econometric society, 1287-1294.

Caballero, C. C., Abello LL, R., y Palacio, J. (2007). Relación del burnout y el rendimiento académico con la satisfacción frente a los estudios en estudiantes universitarios. *Avances en psicología latinoamericana*, *25*(2), 98-111.

Cadena, D. M. G., Lugo, N. I. B., Atilano, B. F., Hoyos, G. P. A., y Magaña, A. G. (2016). Problemas de salud derivados del uso del teléfono celular. *Comité editorial*.

Canavos, G. (1988). *Probabilidad y estadística - Aplicaciones y métodos*. Virginia Commonwealth University.

Carmona, M., y Carrión, H. (2015). Potencia de la prueba estadística de normalidad Jarque-Bera frente a las pruebas de Anderson-Darling, Jarque-Bera robusta, Chi cuadrada, Chen-Shapiro y Shapiro-Wilk [Universidad Autónoma del Estado de México].

Castro, S., Guzmán, B., y Casado, D. (2007). Las Tic en los procesos de enseñanza y aprendizaje. *Laurus*, *13*(23), 213-234.

Chatterjee, S., y Hadi, A. (2006). *Regression Analysis by Example*. Wiley Interscience.

Chen, Q., y Yan, Z. (2016). Does multitasking with mobile phones affect learning? A review. *Computers in Human behavior*, *54*, 34-42.

Departamento Administrativo Nacional de Estadística (DANE). (2023). Indicadores básicos de tenencia y uso de Tecnologías de la Información y las Comunicaciones-TIC en hogares y personas de 5 y más años de edad.

Muñoz García, J. A., y Amón Uribe, I. (2013). Técnicas para detección de outliers multivariantes. Revista en telecomunicaciones e informática.

Gómez Chacón, I. M. (2000). Matemática emocional: los afectos en el aprendizaje matemático. *Matemática emocional: los afectos en el aprendizaje matemático*.

González, F. V. (2022). Métodos de Selección de Variables en Modelos de Regresión.

Hair, J. F., Black, W. C., Babin, B. J., y Anderson, R. E. (2010). Multivariate data analysis (7th ed.). Pearson.

Hernández Lalinde, J. D., Espinosa Castro, J. F., Peñaloza Tarazona, M. E., Fernández González, J. E., Chacón Rangel, J. G., Toloza Sierra, C. A., ... y Bermúdez Pirela, V. J. (2018). Sobre el uso adecuado del coeficiente de correlación de Pearson: definición, propiedades y suposiciones.

Hilt, J. A. (2019). Dependencia del celular, hábitos y actitudes hacia la lectura y su relación con el rendimiento académico. *Apuntes Universitarios*, *9*(3), 103-116.

Issa, T., y Isaias, P. (2016). Internet factors influencing generations Y and Z in Australia and Portugal: A practical study. *Information Processing & Management*, *52*(4), 592-617.

Kuss, D. J., y Griffiths, M. D. (2017). Social networking sites and addiction: Ten lessons learned. *International journal of environmental research and public health*, *14*(3), 311.

Lamana-Zapata, R., Ibarra-Mora, J., Henriquez-Beltrán, M., Sepúlveda-Martin, S., Martínez-González, L., y Cigarroa, I. (2021). Aumento de horas de pantalla se asocia con un bajo rendimiento escolar. *Andes pediátrica, 92*(4), 565-575.

Lamas, H. A. (2015). Sobre el rendimiento escolar. *Propósitos y representaciones*, *3*(1), 313-386.

Lahura, E. (2003). *El coeficiente de correlación y correlaciones espúreas* (Vol. 218). Pontificia Universidad Católica del Perú, Departamento de Economía.

Ling Tsai, S., y Walberg, H. J (1983). Mathematics achievement and attitude productivity in junior high school. *The Journal of Educational Research*, *76*(5), 267-272.

Martínez Ortega, R. M., Tuya Pendás, L. C., Martínez Ortega, M., Pérez Abreu, A., y Cánovas, A. M. (2009). El coeficiente de correlación de los rangos de Spearman caracterización. *Revista Habanera de Ciencias Médicas*, *8*(2), 0-0.

Martínez Padrón, O. J. (2008). Actitudes hacia la matemática Sapiens. Revista Universitaria de Investigación, vol. 9, núm. 1, junio, 2008, pp. 237-256 Universidad Pedagógica Experimental Libertador Caracas, Venezuela. *Sapiens. Revista universitaria de investigación*, *9*(1), 237-256.

Martínez, R., Ibarra, J., y Zapata, R. (2021). Aumento de horas de pantalla se asocia con un bajo rendimiento escolar. *Andes Pediátrica*, 92(4), 565-575.

McLeod, D. B. (1988). Affective issues in mathematical problem solving: Some theoretical considerations. Journal for Research in Mathematics Education, 19(2), 134-141.

Méndez, D., y Macía, F. (2007). Análisis factorial confirmatorio de la escala de actitudes hacia la estadística. *Cuadernos de Neuropsicología/Panamerican Journal of Neuropsychology*, *1*(3), 337-345.

Mendivelso, F. y Rodríguez, M. (2021). Prueba no paramétrica de correlación de Spearman. *Revista Médica Sanitas*, *24*(1).

Ministerio de Educación Nacional (MEN). (2009). Evaluación del aprendizaje y promoción de los estudiantes de los niveles de educación básica y media. Decreto 1290. Diario Oficial número 47.322 de fecha 16 de abril de 2009. Colombia.

Monje, C. (2011). Metodología de la investigación cuantitativa y cualitativa. Universidad surcolombiana.

Monsalve Cano, H. D. J. (2016). Aproximaciones hacia una definición de “Bajo Rendimiento Escolar”.

Navarrete,B.(2016). El impacto negativo de las tecnologías en los adolescentes y jóvenes. *Medimay*, *22*(3), 158-172.

Naranjo, D., Buenaño, D., y Mejía, I. T. (2016). Evolución de la tecnología móvil. Camino a 5G. *Revista Contribuciones a las Ciencias Sociales*, 1-13.

Novales, A. (2010). Análisis de regresión. *Universidad Complutense de Madrid: Madrid, Spain*, *116*.

Ochoa, J., Mesa, S., Pedraza, Y., y Orlando, E. (2017). *La lectura inferencial, una clave para potenciar la comprensión lectora. Educación y Ciencia, (20), 249–263*.

Palacios, A., Arias, V., y Arias, B. (2014). Las actitudes hacia las matemáticas: construcción y validación de un instrumento para su medida. *Revista de psicodidáctica*, *19*(1), 67-91.

Pedrero, E., Rodríguez, M., y Ruiz, J. (2012). Adicción o abuso del teléfono móvil. *Adicciones*, 24(2), 139-152.

Rencher, A. C. (2002). Methods of multivariate analysis (2.ª ed.). J. Wiley. (Wiley Series in Probability and Mathematical Statistics).

Prensky, M. (2001). Nativos digitales, inmigrantes digitales. *On the horizon*, *9*(5), 1-7.

Rendon, C. (2023) *Estudio de la técnica multivariante de análisis discriminante y aplicación a datos reales* [Tesis de maestría]. Universidad de Granada.

Rosen, L. D., Lim, A. F., Carrier, L. M., y Cheever, N. A. (2011). An empirical examination of the educational impact of text message-induced task switching in the classroom: Educational implications and strategies to enhance learning. *Psicología Educativa. Revista de los Psicólogos de la Educación*, *17*(2), 163-177.

Santana Vega, L. E., Gómez Muñoz, A. M., y Feliciano García, L. (2019). Uso problemático del móvil, fobia a sentirse excluido y comunicación familiar de los adolescentes.

Siegel, S., y Castellan, N. J., Jr. (1988). Nonparametric statistics for the behavioral sciences (2nd ed.). Mcgraw-Hill Book Company.

Tsitsika, A. K., Tzavela, E. C., Janikian, M., Ólafsson, K., Iordache, A., Schoenmakers, T. M., ... y Richardson, C. (2014). Online social networking in adolescence: Patterns of use in six European countries and links with psychosocial functioning. *Journal of adolescent health*, *55*(1), 141-147.

Ursini, S., Sánchez, G., y Orendain, M. (2004). Validación y confiabilidad de una escala de actitudes hacia las matemáticas y hacia las matemáticas enseñadas con computadora. *Educación matemática*, *16*(3), 59-78.

Ursini, S., y Sánchez, J. (2019). Actitudes hacia las matemáticas. *Qué son. Cómo se miden. Cómo se evalúan. Cómo se modifican. Ciudad de México, México: UNAM, FES Zaragoza*.

Vázquez, D. M., y de la Torre Fernández, E. (2009). Evaluación de las actitudes hacia las matemáticas y el rendimiento académico. In Investigación en Educación Matemática XIII (pp. 285-300). Sociedad Española de Investigación en Educación Matemática, SEIEM.

Wooldridge, y. M. (2010). *Introducción a la econometría. Un enfoque moderno: un enfoque moderno*. Ediciones Paraninfo, SA.

Zan, R., y Di Martino, P. (2007). Attitude toward mathematics: Overcoming the positive/negative dichotomy. *The montana mathematics enthusiast*, *3*(1), 157-168.

ANEXOS

Anexo 1

	FOMATO	
	AUTORIZACION TRATAMIENTO DE DATOS PERSONALES MENORES DE EDAD Resolución 757 de 18 de junio 2018	
FOR005GSI	Fecha de Aprobación: 18-06-2018	Versión: 01	Página 2 de 2

AUTORIZACIÓN TRATAMIENTO DE DATOS PERSONALES DE MENORES DE EDAD

Yo, ______________________________, identificado con C.C. No. ______________ expedida en ______________, representante legal del menor ______________________, identificado con T.I. No. ______________ del curso ________, declaro que he sido informado por LA UNIVERSIDAD PEDAGÓGICA NACIONAL (en adelante la UPN), identificada con NIT. 899.999.124-4, con domicilio en la ciudad de Bogotá y sede principal en la calle 72 No. 11 – 86 de Bogotá, que, de conformidad con los procedimientos establecidos en la Ley 1581 de 2012, Decreto Reglamentario 1377 de 2013 y el Manual de política interna y procedimientos para el tratamiento y protección de datos personales de la Universidad, disponible en la página web www.pedagogica.edu.co, actuará como Responsable del tratamiento de mis datos personales[1], necesarios para el cumplimiento de la misión de la UPN, obtenidos a través de canales y dependencias institucionales y que podrá recolectar, almacenar, usar, actualizar, transmitir, transferir y poner en circulación o suprimirlos, mediante el uso de las medidas necesarias para otorgar seguridad a los registros, evitando su adulteración, pérdida, consulta, uso o acceso no autorizado o fraudulento incluso por terceros.

Que tratándose de datos sensibles[2] y de menores de edad no está obligado a autorizar su tratamiento, salvo las excepciones consagradas en la ley o que medie su consentimiento expreso. Que es de carácter facultativo responder a las preguntas que traten de datos sensibles o menores de edad.

Como representante legal del menor, debo velar por los derechos consagrados en la Constitución y la Ley sobre sus datos, especialmente el derecho a conocer, actualizar, rectificar y suprimir información personal, así como el derecho a revocar el consentimiento otorgado para el tratamiento de datos personales del menor, en los casos en que sea procedente. Las inquietudes o solicitudes relacionadas con el tratamiento de dichos datos pueden ser tramitadas a través del e-mail: quejasyreclamos@pedagogica.edu.co

1 La UPN garantiza la confidencialidad, libertad, seguridad, veracidad, transparencia, acceso y circulación restringida de mis datos y se reserva el derecho de modificar su Política de Tratamiento de datos personales en cualquier momento. Cualquier cambio será informado y publicado oportunamente en la página web.

2 Son **datos sensibles** aquellos que afectan la intimidad del Titular o cuyo uso indebido puede generar su discriminación, tales como aquellos que revelen el origen racial o étnico, la orientación política, las convicciones religiosas o filosóficas, la pertenencia a sindicatos, organizaciones sociales, de derechos humanos o que promueva intereses de cualquier partido político o que garanticen los derechos y garantías de partidos políticos de oposición, así como los datos relativos a la salud, a la vida sexual, y los datos biométricos (Art. 5° Ley 1581 de 2012, art. 3° Decreto 1377 de 2013).

La Universidad garantiza la confidencialidad, libertad, seguridad, veracidad, transparencia, acceso y circulación restringida de los datos y se reserva el derecho de modificar su Política de Tratamiento de datos personales en cualquier momento. Cualquier cambio será informado y publicado oportunamente en la página web.

En este sentido, como parte del proyecto educativo y con el fin de realizar un análisis académico, autorizo la instalación de una aplicación en el dispositivo móvil del estudiante que registrará de manera no invasiva el tiempo de uso de las distintas aplicaciones instaladas en el mismo. Esta herramienta permitirá obtener información de la cantidad de tiempo que usa el estudiante el celular, respetando siempre la privacidad y confidencialidad de los datos recolectados, los cuales serán utilizados exclusivamente con fines académicos y bajo los más altos estándares de seguridad.

Teniendo en cuenta lo anterior, autorizo de manera voluntaria, previa, explicita, informada e inequívoca a la UPN para tratar los datos personales del menor que represento, de acuerdo con el Manual de política interna y procedimientos para el tratamiento y protección de datos personales de la Universidad y para los fines relacionados con su Misión.

Leído lo anterior, manifiesto que la información para el Tratamiento de los datos personales del menor de edad que represento ha sido suministrada de forma voluntaria y es veraz, completa, exacta, actualizada, comprobable y comprensible.

____

FIRMA

Nombre:

Identificación (C.C):

1 La UPN garantiza la confidencialidad, libertad, seguridad, veracidad, transparencia, acceso y circulación restringida de mis datos y se reserva el derecho de modificar su Política de Tratamiento de datos personales en cualquier momento. Cualquier cambio será informado y publicado oportunamente en la página web.

2 Son **datos sensibles** aquellos que afectan la intimidad del Titular o cuyo uso indebido puede generar su discriminación, tales como aquellos que revelen el origen racial o étnico, la orientación política, las convicciones religiosas o filosóficas, la pertenencia a sindicatos, organizaciones sociales, de derechos humanos o que promueva intereses de cualquier partido político o que garanticen los derechos y garantías de partidos políticos de oposición, así como los datos relativos a la salud, a la vida sexual, y los datos biométricos (Art. 5° Ley 1581 de 2012, art. 3° Decreto 1377 de 2013).

Anexo 2

UNIVERSIDAD PEDAGOGICA NACIONAL
Educadora de educadores

FACULTAD DE CIENCIA Y TECNOLOGÍA
DEPARTAMENTO DE MATEMÁTICAS
Licenciatura en Matemáticas

Actitudes hacia las Matemáticas

Nombre: ____________________________Género: __________Edad: _______Curso:_______

Ítem						
1.	Me gusta la clase de matemáticas	*MUCHO*	*SI*	*INDECISO*	*POCO*	*NO*
2.	La clase de matemáticas es aburrida	*MUCHO*	*SI*	*INDECISO*	*POCO*	*NO*
3.	Las matemáticas son difíciles	*MUCHO*	*SI*	*INDECISO*	*POCO*	*NO*
4.	Matemáticas es la materia que más me gusta	*MUCHO*	*SI*	*INDECISO*	*POCO*	*NO*
5.	Las matemáticas son divertidas	*MUCHO*	*SI*	*INDECISO*	*POCO*	*NO*
6.	Me gustan las matemáticas	*MUCHO*	*SI*	*INDECISO*	*POCO*	*NO*
7.	Es importante aprender matemáticas	*MUCHO*	*SI*	*INDECISO*	*POCO*	*NO*
8.	Me gustaría usar las matemáticas cuando vaya a trabajar	*MUCHO*	*SI*	*INDECISO*	*POCO*	*NO*
9.	Me gusta aprender matemáticas con computador	*MUCHO*	*SI*	*INDECISO*	*POCO*	*NO*
10.	Tengo dificultad para entender lo que me piden en las tareas de matemáticas	*MUCHO*	*SI*	*INDECISO*	*POCO*	*NO*
11.	Puedo resolver los problemas planteados en las tareas de matemáticas	*MUCHO*	*SI*	*INDECISO*	*POCO*	*NO*
12.	Me parece aburrido aprender matemáticas usando el computador	*MUCHO*	*SI*	*INDECISO*	*POCO*	*NO*
13.	Me gusta proponer la solución a problemas antes que los demás	*MUCHO*	*SI*	*INDECISO*	*POCO*	*NO*
14.	Me gusta ser el líder de mi equipo	*MUCHO*	*SI*	*INDECISO*	*POCO*	*NO*
15.	Si un problema no sale a la primera, le busco hasta resolverlo	*MUCHO*	*SI*	*INDECISO*	*POCO*	*NO*
16.	Me gusta resolver problemas de matemáticas algo difíciles	*MUCHO*	*SI*	*INDECISO*	*POCO*	*NO*
17.	Me gusta cuando discutimos en grupo cómo resolver un problema de matemáticas	*MUCHO*	*SI*	*INDECISO*	*POCO*	*NO*
18.	En el trabajo en grupo de matemáticas defiendo mis ideas	*MUCHO*	*SI*	*INDECISO*	*POCO*	*NO*

Anexo 3

UNIVERSIDAD PEDAGOGICA NACIONAL
Educadora de educadores

FACULTAD DE CIENCIA Y TECNOLOGÍA
DEPARTAMENTO DE MATEMÁTICAS

DMA-355

Bogotá, martes 01 de octubre de 2024

Señores:
ALVARO HENRY CORREDOR
Rector
SANDRA MÁRQUEZ
Coordinadora
YEIMMY RIAÑO MORENO
Docente de Matemáticas
Colegio Venecia IED
Carrera 55 #49-25
La Ciudad

Asunto: Realización de trabajo de grado – Licenciatura en Matemáticas

Cordial saludo.

El Departamento de Matemáticas agradece la confianza depositada en la propuesta de formación profesional de la Licenciatura en Matemáticas ofertada por la Universidad Pedagógica Nacional, permitiendo el espacio para que algunos de nuestros futuros educadores adelanten el trabajo de campo necesario para la posterior elaboración de su trabajo de grado. Además, agradecemos su apoyo y disponibilidad de tiempo para los procesos de formación de los futuros educadores matemáticos, ya que los mismos han contribuido significativamente para que ellos fortalezcan sus conocimientos pedagógicos, didácticos y matemáticos.

El trabajo de grado que proyectan nuestros estudiantes tiene como objetivo general indagar por posibles relaciones estadísticas entre el rendimiento académico en la asignatura de matemáticas y el uso del dispositivo móvil en estudiantes de secundaria. Para la consecución de este objetivo, nuestros estudiantes estarán recopilando información sobre el uso del celular que hacen algunos de los estudiantes de su institución, específicamente será de interés saber la cantidad de tiempo diario que emplean los estudiantes el celular. De forma simultánea recopilarán información sobre las calificaciones de los estudiantes en lo que respecta al área de matemáticas. La captura y uso de dicha información se hará respetando la normatividad vigente sobre el tratamiento de datos para menores de edad, de

UNIVERSIDAD PEDAGOGICA NACIONAL
Educadora de educadores

forma anónima, con fines expresamente académicos y ciñéndose a los principios de la ética investigativa.

A continuación, relacionamos los datos de los futuros educadores matemáticos que estarán desarrollando el trabajo previamente mencionado:

Documento	Nombres	Correo
1000588580	PEDRAZA HUERTAS CARLOS ANDRES	capedrazah@upn.edu.co
1004005849	VARGAS ORTEGON ANDRES FELIPE	afvargaso@upn.edu.co

Por otra parte, el profesor del Departamento de Matemáticas que estará acompañando el proceso académico del trabajo de grado de los futuros educadores es:

Nombre	Correo
César Guillermo Rendón Mayorga	cgrendonm@upn.edu.co

De antemano agradecemos su contribución con las actividades académicas que se desarrollarán nuestros estudiantes en su institución, así como la disposición que han tenido en el establecimiento de este vínculo académico. Esperamos que la permanencia y trabajo de nuestros estudiantes en su institución aporte al desarrollo de los procesos académicos que en ella se adelantan. Cualquier inquietud puede ser remitida a la Coordinación de la Licenciatura en Matemáticas, que este semestre está a cargo de la profesora Tania Plazas Merchán a través del correo licenciatura_dma@pedagogica.edu.co

Atentamente,

Sarmiento

BENJAMÍN RAFAEL SARMIENTO LUGO
Director - Departamento de Matemáticas
Facultad de Ciencia y Tecnología
(57-1) 594 1894 Ext. 253-254
Calle 72 N°. 11-86 Of. B318

Proyectó y elaboró: Coordinación de Prácticas Educativas - Licenciatura en Matemáticas dma_practica@pedagogica.edu.co

Anexo 4

Curso	Genero	Edad	Tot_dias_esc	Prom_dias_esc	Tot_dias_noesc	Prom_dias_noesc	Nota_4_corte	Nota_final	Actitudes
Octavo	Masculino	16	123,384	7,258	66,469	6,647	2,80	3,00	2,78
Octavo	Masculino	15	118,499	6,971	45,126	4,513	4,60	4,10	2,39
Octavo	Masculino	16	132,659	7,803	114,155	11,416	2,00	2,30	2,56
Octavo	Femenino	14	189,529	11,149	106,297	10,630	3,00	3,17	2,17
Octavo	Femenino	14	110,418	6,495	80,383	8,038	4,00	4,70	2,72
Octavo	Femenino	14	132,798	7,8117	78,546	7,855	3,20	3,15	1,50
Octavo	Femenino	14	132,666	7,804	62,416	6,242	4,00	4,17	2,67
Octavo	Femenino	14	98,020	5,766	45,923	4,592	3,50	3,10	2,00
Octavo	Femenino	15	86,041	5,061	23,480	2,348	3,00	3,60	2,78
Octavo	Masculino	15	147,475	8,675	103,190	10,319	2,40	2,30	1,89
Octavo	Femenino	15	117,708	6,924	87,930	8,793	4,00	3,30	2,00
Octavo	Masculino	14	130,356	7,668	106,680	10,668	2,00	3,10	2,89
Octavo	Femenino	15	174,811	10,283	48,410	4,841	3,00	3,40	2,78
Decimo	Masculino	16	151,606	8,918	73,490	7,349	3,30	3,80	1,33
Decimo	Masculino	15	116,501	6,853	83,180	8,318	4,30	4,30	2,00
Decimo	Masculino	16	160,606	9,415	107,616	10,762	3,20	3,20	0,94
Decimo	Masculino	16	114,223	6,719	78,840	7,884	2,30	3,10	2,11
Decimo	Femenino	16	134,674	7,922	85,320	8,532	4,70	4,80	2,00
Decimo	Masculino	16	129,443	7,614	89,542	8,954	3,30	3,60	2,61
Decimo	Femenino	16	165,808	9,753	98,284	9,828	3,00	2,10	1,44
Decimo	Masculino	16	106,152	6,244	121,875	12,188	4,10	3,20	2,00
Decimo	Femenino	16	90,606	5,330	63,013	6,301	4,90	4,90	2,17
Decimo	Femenino	17	116,084	6,828	68,071	6,807	3,40	2,80	1,56
Decimo	Femenino	16	71,340	4,196	55,393	5,539	5,00	4,80	1,94
Decimo	Femenino	16	93,221	5,826	57,354	5,735	3,20	3,70	1,78
Decimo	Masculino	17	102,313	6,018	50,023	5,002	3,90	3,70	2,28
Decimo	Femenino	16	118,670	6,981	63,822	6,382	4,60	3,90	1,39
Decimo	Femenino	16	83,198	4,894	79,360	7,936	3,30	2,90	2,00
Decimo	Femenino	17	107,933	6,349	98,660	9,866	3,50	3,10	2,22
Decimo	Femenino	15	71,145	4,185	61,580	6,158	4,00	3,70	2,78
Decimo	Masculino	17	96,662	5,686	56,860	5,686	3,00	3,10	2,22
Decimo	Femenino	17	116,909	6,877	55,920	5,592	3,00	2,80	1,44
Decimo	Masculino	15	62,9	3,7	19,800	1,980	4,10	3,00	1,56
Decimo	Masculino	16	127,738	7,514	103,550	10,355	3,00	3,10	2,72
Decimo	Femenino	15	55,097	3,241	38,730	3,873	4,10	3,80	2,67
Decimo	Femenino	16	87,618	5,154	58,440	5,844	4,30	4,10	2,72
Decimo	Femenino	15	108,8	6,4	60,000	6,000	5,00	4,80	3,33
Decimo	Femenino	16	77,469	4,557	80,950	8,095	4,00	3,50	2,00
Decimo	Femenino	16	114,546	6,738	56,100	5,610	3,70	3,10	1,61
Decimo	Femenino	16	111,022	6,531	37,08	3,708	3,50	3,40	2,00

Anexo 5

```r
# Librerias
library(tidyverse)
library(readxl)
library(ggcorrplot)
library(GGally)

# Carga y lectura de los datos
ruta_doc_datos <- file.choose()

df_compilado_datos <- read_excel(ruta_doc_datos, sheet = 2)

View(df_compilado_datos)

colnames(df_compilado_datos)

# Limpieza de los datos
# Seleccionando solo las columnas con valores numericos
df_compilado_clean <- df_compilado_datos %>%
  select(-c(1:3))

colnames(df_compilado_clean)


# Pruebas de normalidad Shapiro-Wilk y Kolmogorov Smirnov para comprobar
           la normalidad de las variables

nivel de significancia <- 0.05  # Nivel de significancia

# Aplicando Shapiro-Wilk a cada columna numérica
for (col in colnames(df_compilado_clean)) {
  test <- shapiro.test(df_compilado_clean[[col]])  # Aplicar prueba
  pvalue <- test$p.value  # Extraer p-valor

  if (pvalue < nivel de significancia) {
    cat(sprintf('⚠️ %s: No distribuye normal (p = %.4f)\n', col, pvalue))
  } else {
    cat(sprintf('✅ %s: Distribuye normal (p = %.4f)\n', col, pvalue))
  }
}
```

```
# Prueba de Kolmogorov smirnov

for (col in colnames(df_compilado_clean)) {

  ks_test <- ks.test(scale(df_compilado_clean[[col]]), # Estandarizar
                     'pnorm') # Comparar con normal estándar

  if (ks_test$p.value < nivel de significancia) {
    cat('⚠️', col, ': No distribuye normal, con un p-value =',
          round(ks_test$p.value, 4), '\n')
  } else {
    cat('✅', col, ': Distribuye normal, con un p-value =',
          round(ks_test$p.value, 4), '\n')
  }
}

# Matriz de correlación entre todas las variables
cor(df_compilado_clean, method = 'pearson')

# Grafico de la matriz de correlación
# install.packages('ggcorrplot')
ggcorrplot::ggcorrplot(
  cor(df_compilado_clean, method = 'pearson'),
  hc.order = TRUE,
  type = 'upper',
  lab = TRUE, lab_size = 3,
  colors = c('red', 'white', 'blue')
)

# Gráfico de dispersión de las variables
install.packages('GGally')
GGally::ggpairs(df_compilado_clean)

colnames(df_compilado_clean)
# Incluyendo la variable Dummy
df_compilado_datos <- df_compilado_datos %>%
  mutate(Curso_dummy = ifelse(Curso == "Decimo", 1, 0)) %>%
  relocate(Curso_dummy, .after = 1)

# Agregando la variable dummy a los datos limpios
df_compilado_clean <- df_compilado_datos %>% select(-Curso,-Genero,-
          Edad)
```

```
# Corriendo el modelo de prueba con todas las variables
summary(
  modelo_prueba_1 <- lm(Nota_final ~., df_compilado_clean)
)

# Prueba VIF
valores_vif_modelo_prueba_1 <- car::vif(modelo_prueba_1)

# Mostrando los valores del test VIF en una tabla
data.frame(`Variables                  independientes`          =
           names(valores_vif_modelo_prueba_1),         VIF          =
           valores_vif_modelo_prueba_1, row.names = NULL)

# Se eliminan los totales y se calcula la matriz de correlación
cor(df_compilado_clean  %>%  select(-c(Tot_dias_esc,  Tot_dias_noesc)),
           method = 'pearson')


# Grafico de la matriz de correlación
# install.packages('ggcorrplot')
ggcorrplot::ggcorrplot(
  cor(df_compilado_clean %>% select(-c(Tot_dias_esc, Tot_dias_noesc)),
           method = 'pearson'),
  hc.order = TRUE,
  type = 'upper',
  lab = TRUE, lab_size = 3,
  colors = c('red', 'white', 'blue')
)

# modelo de prueba 2
summary(
  modelo_prueba_2 <- lm(Nota_final ~., df_compilado_clean %>% select(-
           c(Tot_dias_esc, Tot_dias_noesc)))
)

# Vif aplicado al modelo de prueba 2
valores_vif_modelo_prueba_2 <- car::vif(modelo_prueba_2)

valores_vif_modelo_prueba_2

# Mostrando los valores del test VIF en una tabla
data.frame(`Variables                  independientes`          =
           names(valores_vif_modelo_prueba_2),         VIF          =
           valores_vif_modelo_prueba_2, row.names = NULL)
```

```
# Metodo de Stepwise para las variables escogidas
# Regresión sin variables explicativas
rlm_vacio <- lm(Nota_final ~ 1, df_compilado_clean %>%
                  select(-c(Tot_dias_noesc, Tot_dias_esc)))

# Regresión con todas las variables explicativas disponibles
rlm_completo <- lm(Nota_final ~ .,
                   df_compilado_clean %>%
                     select(-c(Tot_dias_noesc, Tot_dias_esc)))

# Regresión forward
rlm_forward <- step(rlm_vacio,
                    scope   =   list(lower   =   rlm_vacio,   upper   =
          rlm_completo),
                    direction = 'forward')
summary(rlm_forward)

# Regresión backward
rlm_backward <- step(rlm_completo,
                     scope   =   list(lower   =   rlm_vacio,   upper   =
          rlm_completo),
                     direction = 'backward')

summary(rlm_backward)

# Regresión stepwise (modelo_1)
modelo_1 <- step(rlm_vacio,
                     scope   =   list(lower   =   rlm_vacio,   upper   =
          rlm_completo),
                     direction = 'both')

summary(modelo_1)

# Sin la nota_4_corte (modelo_2)
# Regresión sin variables explicativas
rlm_vacio <- lm(Nota_final ~ 1, df_compilado_clean %>%
                  select(-c(Tot_dias_noesc,               Tot_dias_esc,
          Nota_4_corte)))

# Regresión con todas las variables explicativas disponibles
rlm_completo <- lm(Nota_final ~ .,
                   df_compilado_clean %>%
                     select(-c(Tot_dias_noesc,            Tot_dias_esc,
          Nota_4_corte)))
```

```
# Regresión forward
rlm_forward <- step(rlm_vacio,
                    scope    =    list(lower    =    rlm_vacio,    upper    =
          rlm_completo),
                    direction = 'forward')

summary(rlm_forward)
# Regresión backward
rlm_backward <- step(rlm_completo,
                     scope    =    list(lower    =    rlm_vacio,    upper    =
          rlm_completo),
                     direction = 'backward')
summary(rlm_backward)

# Regresión stepwise
modelo_2 <- step(rlm_vacio,
                     scope    =    list(lower    =    rlm_vacio,    upper    =
          rlm_completo),
                     direction = 'both')
summary(modelo_2)

# Verificación de supuestos
# Agregando la columna de los residuales
df_compilado_clean_residuos <- df_compilado_clean %>%
  mutate(Residuales      =      Nota_final      -      predict(modelo_2,
          df_compilado_clean))

# 4. Linealidad. (Matriz de correlación con las variables del modelo)
cor(df_compilado_clean %>%
      select(-c(Tot_dias_noesc,       Tot_dias_esc,       Nota_4_corte,
          Prom_dias_esc)), method = 'pearson')

# Grafico de la correlación
GGally::ggpairs(df_compilado_clean %>%
                  select(-c(Tot_dias_noesc, Tot_dias_esc, Nota_4_corte,
          Prom_dias_esc, Curso_dummy)))

# 2. Homocedasticidad
# Gráfica de los residuales contra los valores pronosticados con línea
          horizontal en y = 0
ggplot(df_compilado_clean_residuos,    aes(x    =    predict(modelo_2,
          df_compilado_clean_residuos),
                              y = Residuales)) +
  geom_point() +
  geom_hline(yintercept = 0, linetype = "dashed", color = "red") +
  labs(x = "Valores Pronosticados", y = "Residuales",
```

```r
       title = "Gráfico de Residuales vs. Valores Pronosticados") +
  theme_minimal()

# Realizando un test de homocedasticidad (Breusch-Pagan)
install.packages('lmtest')
lmtest::bptest(modelo_2)

# 1. Normalidad (de los residuos)
# Valor de la media de los residuales
cat('La    media    de    los    errores    es:',    sprintf('%.20f',
           mean(df_compilado_clean_residuos$Residuales)))

# Histograma de los residuales
hist(df_compilado_clean_residuos$Residuales)

# Pruebas de normalidad
# Prueba de shapiro
shapiro.test(df_compilado_clean_residuos$Residuales)

#prueba de kolmogorov
ks.test(scale(df_compilado_clean_residuos$Residuales), # Estandarizar
        'pnorm') # Comparar con normal estándar

# 4. Independencia de los errores
lmtest::dwtest(modelo_2)

# 5. Multicolinealidad (VIF del modelo final)
valores_vif_modelo_final <- car::vif(modelo_2)

# Mostrando los valores del test VIF en una tabla
data.frame(`Variables independientes` = names(valores_vif_modelo_final),
           VIF = valores_vif_modelo_final, row.names = NULL)

# Intervalos de confiaza:
confint(modelo_2, level = 0.90)
```