

Análisis mediante series de tiempo de los resultados en matemáticas de las pruebas Saber 11 y niveles socioeconómicos NSE usando k-medias

Time series analysis of the Saber 11 state test in mathematics and cluster analysis of the socioeconomic NSE levels using the k-means algorithm

Daniel Alejandro Andrade Rodríguez*

Fundación Universitaria los Libertadores
Facultad de Ingeniería y Ciencias Básicas
Departamento de Estadística
Especialización en Estadística Aplicada

1 de Diciembre de 2020

Resumen

Las pruebas Saber 11 se caracterizan no solo por medir las habilidades de los estudiantes en situaciones particulares, sino también por caracterizar su ambiente socioeconómico mediante cuestionarios de respuesta múltiple anexos en las aplicaciones, de los cuales se generan diferentes niveles socioeconómicos a partir de la teoría de respuesta al ítem. En el presente trabajo se propone una nueva metodología para la determinación de estos niveles mediante análisis de clústeres, generando no solo un nuevo número de niveles sino también una nueva caracterización para cada uno de ellos. Adicionalmente se propone el análisis como ejercicio académico de los resultados en el área de matemáticas desde el año 2000-1 hasta el año 2019-2 mediante el uso de series de tiempo univariadas.

Palabras clave: ICFES, k-means, series de tiempo, NSE, saber11

Abstract

The Saber 11 state test is characterized not only by measuring the abilities of students in particular situations, but also by characterizing their socioeconomic environment through multiple-response questionnaires attached to the applications, from which different socioeconomic levels are generated from the item response theory. In the present work a new methodology is proposed for the determination of those levels using cluster analysis, generating not only a new number of levels but also a new characterization for each one of them. Additionally, the analysis of the results in the area of mathematics from the year 2000-1 to the year 2019-2 is proposed as an academic exercise through the use of univariate time series.

Keywords: Icfes, k-means, time series, NSE

*daandrader@libertadores.edu.co — github.com/daandrader

1 Introducción

El ICFES dentro de su labor como entidad encargada de realizar la evaluación de la educación en Colombia en todos los niveles, tiene a su cargo la evaluación de la educación media, para lo cual aplica las pruebas Saber 11 a los estudiantes del último grado de las instituciones públicas y privadas del país. Estas pruebas están enfocadas a evaluar las competencias de los estudiantes, definidas como las habilidades necesarias para aplicar de manera flexible los conocimientos en diferentes contextos.

Las áreas o componentes básicas que son evaluadas desde el año 2014 corresponde a las áreas de matemáticas, lectura crítica, ciencias naturales, sociales y ciudadanas e inglés con puntajes que van desde 0 a 100, sin embargo, estas componentes han visto cambios[16] a lo largo de los años, variando como tal la componente objetivo, el contenido o la metodología de evaluación lo que dificulta la comparabilidad en el tiempo, algo en lo que se ha venido trabajando mediante el SNEE (Sistema de Evaluación Estandarizada).

Para entender los resultados obtenidos por los estudiantes se debe también conocer cuál es su contexto social, económico y familiar de los evaluados; estas características se conocen como factores asociados al aprendizaje. Trabajos anteriores como el realizado por Timarán[2] en el 2009 en el cual se analiza los patrones de bajo rendimiento de los estudiantes de la Universidad de Nariño relacionando los resultados en los primeros semestres junto con los resultados obtenidos en las pruebas Saber 11 mediante minería de datos confirman como estos factores influyen notoriamente en los resultados obtenidos en ambas etapas del estudiante. De la misma manera Barón[3] en el 2012 analiza las diferencias entre los resultados obtenidos por los estudiantes de la ciudad de Bogotá y la ciudad de Barranquilla llegando a las mismas conclusiones.

Es por ello que el ICFES junto con las pruebas específicas de conocimiento pone a disposición de los estudiantes cuestionarios socioeconómicos cuyas preguntas son de respuesta múltiple con el objetivo de conocer cuáles son estas condiciones para los estudiantes y relacionarlas con los resultados obtenidos. Dentro del análisis hecho por el ICFES se ha generado un constructo teórico, a partir de la teoría de respuesta al ítem (TRI) para describir este contexto, llamado el índice de

nivel socioeconómico (INSE). Este puntaje definido de 0 a 100, define a su vez 4 niveles NSE donde los estudiantes son clasificados acorde con las repuestas de sus cuestionarios[17].

En el presente trabajo se propone una nueva forma de clasificación de estos niveles socioeconómicos, no desde la perspectiva de la TRI sino aplicando técnicas de agrupamiento de conglomerados, en específico agrupamiento por k-medias. Esta aproximación al problema de la clasificación de niveles usa como variables principales los resultados en cada una de las componentes evaluadas para clasificar y caracterizar a los estudiantes en 3 niveles bien definidos, de esta manera se presenta una aproximación práctica que permite detallar de una manera diferente estos niveles socioeconómicos.

Además de lo anterior se propone como ejercicio académico el análisis de los resultados en el área de matemáticas mediante series de tiempo y la metodología de Box-Jenkins desde el año 2000 hasta el año 2019, siendo estos lo que menores cambios han presentado en este periodo de tiempo. La serie generada es analizada en sus componentes y ajustado a un modelo ARIMA a la vez que se generan proyecciones en los puntajes para los siguientes 5 periodos de evaluación.

2 Referentes teóricos

2.1 Pruebas Saber 11

Las pruebas SABER 11 son pruebas aplicadas semestralmente por el ICFES para evaluar la calidad de la educación media en Colombia, dichas pruebas están estandarizadas y su objetivo se centra en servir de criterio para el ingreso de los estudiantes a las instituciones de educación superior. La primera versión del examen fue aplicada en el año 1968, tiempo para el cual su único propósito era el de apoyar a las instituciones de educación superior en sus procesos de admisión. Para el año 1980 finalmente se convirtió en un requisito obligatorio y formal para la admisión a este ciclo educativo. Desde el año 2000 la prueba se enfocó no en evaluar los conocimientos de los estudiantes sino en evaluar sus competencias en situaciones particulares y cotidianas[13]. Los datos presentes en las bases de datos también son públicos desde este año.

Las componentes que son evaluadas corresponden de manera principal a las áreas de matemáticas, lenguaje, ciencias naturales, ciencias sociales e idiomas, sin

2000-1 a 2005-2	2006-1 a 2014-1	2014-2 en adelante
Lenguaje Filosofía	Lenguaje Filosofía	Lectura Crítica
Matemáticas	Matemáticas	Matemáticas (Incluye razonamiento cuantitativo)
Física Química Biología	Física Química Biología	Ciencias Naturales
Historia Geografía	Ciencias Sociales (Historia, Geografía)	Sociales y ciudadanas (Incluye Competencias Ciudadanas)
Idioma (inglés, francés o alemán)	Idioma-Inglés (inglés, francés o alemán)	Inglés

Cuadro 1: Componentes evaluadas y los sus respectivos cambios desde el año 2000.

embargo, a lo largo de los años estas componentes han ido variando acorde con las políticas internas del ICFES y de los lineamientos nacionales.

Estas variaciones van encaminadas a consolidar un Sistema de Evaluación Estandarizada (SNEE) alineando todos los exámenes que son aplicados por el ICFES y además permitir la comparabilidad de resultados, en términos de las competencias genéricas con las otras pruebas del SNEE: saber 3, 5, 9 y los exámenes de educación superior[18]. Las pruebas Saber 11, tiene como población evaluada a los estudiantes de último año presentan el examen a través de instituciones educativas, a los validantes aquellos estudiantes que presentan el examen con la finalidad de validar su bachillerato y, por último, están los individuales que son aquellos examinados que presentan su examen de forma individual sin estar ligados a ninguna institución educativa, por lo general son estudiantes ya graduados.

Las preguntas de estos exámenes son construidas en base a 3 pilares fundamentales: competencias, afirmaciones y evidencias. Las competencias son habilidades necesarias para aplicar de manera flexible los conocimientos en diferentes contextos, lo cual implica no solamente conocer conceptos o datos, sino que implica saber cómo emplear dichos conceptos para resolver situaciones cotidianas[14]. Para cada competencia se establecen diferentes afirmaciones, entendidas como descripciones de aquello que se espera que el estudiante este en la capacidad de hacer cuando ha desarrollado una competencia. Cada afirmación se compone a su vez de evidencias, que

representan acciones o ejecuciones observables, mediante las cuales es posible verificar si el estudiante evaluado ha desarrollado lo relacionado con la afirmación a la que pertenece.

Además de las pruebas enfocadas a las componentes básicas, el estudiante también debe responder un cuestionario socioeconómico, este cuestionario se compone de preguntas cortas de selección múltiple enfocadas a obtener información sobre los procesos de enseñanza y aprendizaje de los estudiantes y ayuda a explicar los resultados de los exámenes, asociándolos con las condiciones y características familiares y el estrato socioeconómico, entre otras

2.2 Índice de nivel socioeconómico (INSE)

Acorde con las respuestas de los estudiantes evaluados en el cuestionario socioeconómico adjunto a las pruebas Saber, es posible conocer más acerca de sus antecedentes escolares, competencias socioemocionales y características socioeconómicas y culturales, también conocidos como factores asociados al aprendizaje. Estos se vuelven relevantes en el estudio de la evaluación de la educación ya que tienen influencia sobre el logro académico. Acorde con algunos autores en trabajos anteriores, a mejores condiciones socioeconómicas se presentan mejores resultados en el desempeño académico.

Dentro de la población evaluada existen grupos de estudiantes que comparten características comunes que van más allá del estrato residencial; este indicador se

vuelve insuficiente a la hora describir y caracterizar otros factores como bienes y servicios asociados al rendimiento de los estudiantes en las pruebas Saber 11, es por ello, por lo que se hace necesario otro indicador que englobe este tipo de características[6]. Es debido a esta necesidad que se construye el índice de nivel socioeconómico (INSE), con el objetivo de caracterizar a la población evaluada de manera integral; este índice es un constructo teórico que pretende representar el estatus socioeconómico, midiendo variables como el capital físico, el capital humano o el capital social al cual pertenece el evaluado.

Desde el 2009 hasta el 2012 el INSE era calculado con base en la metodología multivariante que emplea el Departamento Nacional de Planeación (DNP), a partir del 2012 y hasta la fecha se emplea la teoría de respuesta al ítem (TRI). Una de las ventajas de esta metodología es que permite construir una escala histórica, garantizando la comparabilidad entre periodos. La TRI parte del supuesto que un rasgo latente o característico puede ser medido mediante las respuestas a los ítems o preguntas. De manera que una variable no observada como el índice socioeconómico, puede ser medida con base a los ítems propuestos en los cuestionarios asociados[6].

El puntaje INSE tiene un rango de 0-100, con una media en 50 y una variación de 10 unidades. Gran cantidad de los estudiantes evaluados se encuentran entre los rangos de 40 y 60.

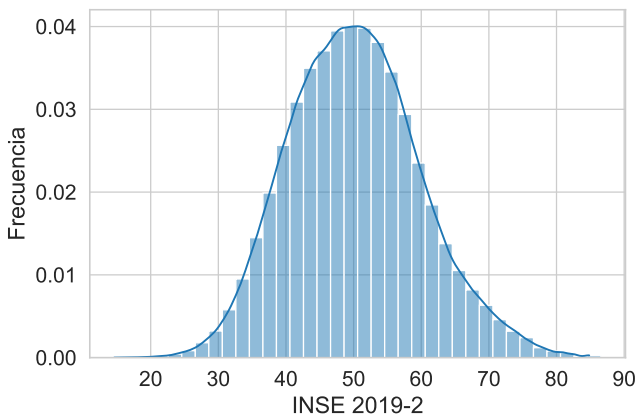


Figura 1: Distribución puntajes INSE para el periodo 2019-2.

El puntaje INSE se emplea como base para el cálculo del nivel socioeconómico categórico (NSE), el cual se

para caracterizar la población que presenta las pruebas Saber. EL NSE se enmarca en una escala de 1-4, donde el primer nivel hace referencia a estudiantes pertenecientes a niveles socioeconómicos bajos y se incrementa hasta el nivel 4 que corresponde a estudiantes con niveles socioeconómicos altos.

Nivel socioeconómico	Puntos corte INSE
NSE 1	0-41.109
NSE 2	Hasta 51.176
NSE 3	Hasta 64.080
NSE 4	Hasta 100

Cuadro 2: Puntajes de corte del INSE por NSE

Para la caracterización de los grupos NSE se emplea una metodología de árboles de decisión que clasifica a los estudiantes evaluados en uno de los 4 niveles posibles de NSE y permite determinar la frecuencia de cada respuesta en las diferentes categorías de los ítems asociados a un nivel NSE. Luego de esto se ordenan las frecuencias por categoría de respuestas en orden descendente y se establece la composición de cada nivel. En el cuadro 2 se muestran las categorías de respuesta de mayor frecuencia asociados a cada NSE, que corresponden a los cuestionarios socioeconómicos usados en las pruebas Saber[6].

En base al cuadro 3, es posible caracterizar los niveles NSE de la siguiente manera[17]:

- Los estudiantes pertenecientes al NSE 1 suelen no tener servicio de internet, ni tener computador, horno microondas ni lavadora. Generalmente la educación de la madre es primaria incompleta.
- Los estudiantes pertenecientes al NSE 2 mantienen la ausencia del servicio de internet, pero la no tenencia de computador es menos marcada. Este nivel se caracteriza por tenencia de lavadora y acceso al servicio de televisión.
- Los estudiantes pertenecientes al NSE 3 se caracterizan por tener acceso a servicios de televisión, tener horno microondas y no poseer automóvil. Las madres de los estudiantes de este nivel generalmente tienen presentan un nivel educativo de secundaria completa.
- Los estudiantes pertenecientes al NSE 4 se caracterizan por tener computador, automóvil, servicio de internet y consola de videojuegos. Los padres

Variable de caracterización	Niveles NSE			
	1	2	3	4
Internet (NO)	X	X		
Computador (NO)	X	X		
Horno microondas o gas (NO)	X			
Lavadora (NO)	X			
Educación madre: primaria incompleta	X			
Lavadora (SÍ)		X		
Computador (SÍ)		X		X
Servicio de televisión (SÍ)		X	X	
Educación madre: secundaria completa			X	
Automóvil particular (NO)			X	
Horno microondas o gas (SÍ)			X	
Automóvil particular (SÍ)			X	X
Educación madre: (profesional completa)				X
Internet (SÍ)				X
Consola de videojuegos				X

Cuadro 3: Descriptores socioeconómicos para los diferentes niveles NSE

se caracterizan por tener un nivel educativo de profesional completo.

2.3 Series de tiempo

Una serie de tiempo puede ser descrita como una secuencia de observaciones (misma variable) tomadas de manera secuencial en el tiempo[11], por ejemplo, el número semanal de accidentes en carretera o las cantidades de lluvia diaria en la ciudad de Bogotá.

Dentro de las características principales de las series de tiempo se encuentran:

- **Tendencia:** en general nos dice si la serie tiende a tomar mayores o menores valores en relación con el tiempo.
- **Estacionalidad:** nos habla acerca de patrones regulares que pueden observarse fácilmente periodos regulares de tiempo (trimestres, semestres, anos, etc).
- **Estacionaridad:** implica que las propiedades de la serie no cambian sin importar el tiempo en que se observe (media y varianza constantes). Es importante ya que muchos de los modelos aplicados en series de tiempo requieren series estacionarias.

Dependiendo de la naturaleza de la serie de tiempo existen diferentes modelos que pueden ser usados para

describir y predecir el comportamiento para tiempos futuros. Existen modelos autorregresivos (AR), modelos de medias móviles (MA) y combinaciones de estos dos (ARMA, ARIMA, SARIMA, etc).

Los modelos autorregresivos AR(p), representan procesos aleatorios, en la que la variable de interés x_t (la serie en el tiempo actual) depende linealmente de sus observaciones pasadas y de un término estocástico w_t (ruido blanco), que por lo general proviene de una distribución normal[11].

$$x_t = w_t + \sum_{i=1}^p \phi_i x_{t-i} \quad (1)$$

Los modelos de media móvil MA(q) se pueden interpretar conceptualmente como procesos de regresión lineal, en este caso, del valor actual de la serie x_t en función de los valores actuales y pasados del ruido blanco w_t y w_{t-1} respectivamente.

$$x_t = w_t + \sum_{i=1}^q \theta_i w_{t-i} \quad (2)$$

Los modelos autorregresivos de media móvil ARMA(p,q) son modelos que integran tanto la parte autorregresiva, como la parte de medias móviles mencionadas anteriormente ARMA(p,q). En caso de que la serie no sea estacionaria y requiera procesos de diferenciación surge el

modelo ARIMA(p,d,q), donde la d hace referencia a los grados de diferenciación requeridos para que la serie en estudio sea estacionaria. Otro posible caso es cuando la serie muestra algún grado de estacionalidad de manera que es necesario tener esta en cuenta esta componente, de acá surgen los modelos SARIMA(p,d,q)×(P,D,Q), donde las componentes en mayúscula hacen referencia a la descripción de la parte estacional.

$$x_t = w_t + \sum_{i=1}^p \phi_i x_{t-1} + \sum_{i=1}^q \theta_i w_{t-1} \quad (3)$$

2.3.1 Metodología Box-Jenkins

Una vez que el fenómeno tras una serie de tiempo es comprendido en su totalidad es posible escribir una fórmula matemática que la describa de manera aproximada (1)(2)(3) permitiéndonos no solo describir la serie en el tiempo presente, sino realizar predicciones en lapsos de tiempo futuros. Dentro de esta aproximación matemática también debe tenerse en cuenta los parámetros asociados, los cuales pueden variar de acuerdo al modelo usado.

Box y Jenkins formularon una metodología para el análisis de series de tiempo, en la cual de manera estructurada y organizada es posible determinar estos modelos y refinarlos en caso de que no sean los adecuados para el modelo que se esté trabajando, esta metodología se basa en los siguientes pasos:

- Desde la interacción con la teoría y la práctica varios modelos que se ajusten a la serie son propuestos.
- De los modelos propuestos se identifica el mejor modelo que se adapte a la serie, este modelo será puesto a prueba en los pasos subsecuentes.
- El modelo preliminar que está siendo analizado es ajustado a la serie, calculando los parámetros asociados.
- Pruebas de consistencia estadística son aplicadas al modelo: independencia de los residuos, aleatoriedad y distribución.
- En caso de que el modelo sea el adecuado en el ajuste y en la consistencia estadística, se usa para generar predicciones de la serie en estudio.

Esta es la metodología de Box-Jenkins para el análisis de series de tiempo: identificación del modelo, estimación, diagnóstico y predicción[20].

2.4 Análisis de agrupamiento por conglomerados

Las técnicas de agrupamiento se basan en la formación de clúster o conglomerados de tal manera que los objetos que conforman estos conglomerados sean similares entre sí, similitud que es derivada de los datos que identifican a cada uno de los individuos. Algunas de estas técnicas de agrupamiento se basan en identificar los objetos acorde a lo que se denominan centroides, que en palabras básicas son puntos centrales que pueden ser definidos en cada uno de los posibles grupos que existan en los datos; otros algoritmos se basan en el agrupamiento jerárquico haciendo grupos menores a partir de grupos mayores[12]. Dentro de las aplicaciones posibles de las técnicas de clustering se encuentran: segmentación de clientes, detección de anomalías y segmentación de imágenes, entre otras aplicaciones.

La distancia d_{ij} entre un par de individuos k y $k+1$ se puede definir de manera clásica a partir de la distancia de Euclides, donde las coordenadas de los puntos corresponden a las variables numéricas de la base de datos, sin embargo, no es la única definición de distancia en este contexto.

$$d_{ij} = \sqrt{\sum_{k=1}^q (x_{ik} - x_{jk})^2} \quad (4)$$

En particular el algoritmo de k-medias busca particionar los n individuos de un grupo de datos multivariantes en k grupos o clústeres, (G_1, G_2, \dots, G_k) donde G_i denota el conjunto de n_i individuos en el i -ésimo grupo y k es minimizado mediante algún criterio. La implementación más común de k-medias, minimiza k acorde con la suma de cuadrados dentro de cada grupo, WGSS[19] por sus siglas en inglés, valor que está dado por:

$$WGSS = \sum_{j=1}^q \sum_{l=1}^k \sum_{i \in G_l} (x_{ij} - \bar{x}_j^{(l)})^2 \quad (5)$$

Donde $\bar{x}_j^{(l)} = \frac{1}{n_l} \sum_{i \in G_l} x_{ij}$ es la media de los individuos en el grupo G_l en la variable j .

3 Metodología

La metodológica usada en este trabajo es de tipo exploratoria y descriptiva de corte longitudinal en la cual se realiza el análisis de los resultados de las pruebas Saber

11, en el área de matemáticas desde el periodo 2000-1 hasta el 2019-2, como serie de tiempo con un total de 40 puntos muestrales correspondientes a cada uno de los periodos de evaluación realizados por el ICFES, con el fin de modelar y predecir el comportamiento para periodos posteriores. En la segunda parte del trabajo se usan los resultados específicos para el periodo de 2019-2, sobre el cual se realizó un análisis de conglomerados usando los puntajes asociados a las componentes evaluadas, de esta manera que se estable una nueva metodología para la de los niveles socioeconómicos asociados a los resultados.

Las bases de datos del ICFES se encuentran como datos públicos dentro del repositorio¹ del ICFES, junto con la documentación y diccionarios correspondientes no solo de las pruebas Saber 11, sino de todas las pruebas aplicadas por el ICFES. En el caso particular de las pruebas Saber, se realizan aplicaciones cada semestre del año, una correspondiente al calendario académico A y otra aplicación correspondiente a las institucionales educativas de calendario B. Históricamente es mayor la cantidad de estudiantes que presentan la prueba en el calendario A que en el calendario B, dado que la mayoría de las instituciones públicas y privadas desarrollan sus actividades acorde a este calendario.

En total para las ultimas aplicaciones del examen hay alrededor de 84 variables descriptoras que engloban información relacionada con el contexto social y familiar del estudiante, así como información relacionada con las características institucionales a las que pertenece cada estudiante, además de contar con la información de los puntajes de cada una de las componentes evaluadas

Las fases desarrolladas dentro del proyecto se describen a continuación:

- **Obtención y limpieza de datos:** Los datos de los resultados de las pruebas Saber 11 y de las demás pruebas realizadas por el ICFES son de dominio público, sin embargo los datos publicados no se pueden considerar listos para ser usados en investigación dado que poseen en un grado bajo datos no disponible y algunas de las variables no han sido codificadas de manera correcta. Esta limpieza implica descartar esta información que no esté disponible y corregir la codificación de las variables.
- **Serie de tiempo y análisis:** Una vez que los da-

tos están listos para ser procesados que toman los valores medios del puntaje de matemáticas y sus correspondientes varianzas. Ordenados de manera cronológica dan lugar a la serie de tiempo que será analizada. Posterior a esto, la metodología Box-Jenkins es usada para determinar el mejor modelo que se ajuste a la serie y se realizan las predicciones correspondientes a 5 periodos.

- **Niveles por agrupación:** En este punto se toman los resultados correspondiente al periodo 2019-2, con cerca de 540000 individuos, con la finalidad de aplicar el método de k-medias para generar nuevos niveles que estén asociados a las características socioeconómicas de los estudiantes. Para esto se realiza la clasificación usando los resultados de cada una de las componentes evaluadas, generándose tres niveles o conglomerados característicos.
- **Análisis de conglomerados y niveles NSE:** Una vez que los niveles han sido generados, estos son asociados con el individuo correspondiente para posteriormente realizar un análisis de frecuencias que permita determinar cuáles son las características asociadas a cada nivel, en términos del cuestionario socioeconómico que realizan los evaluados. Tras esto se realiza una comparación de los niveles generados por la técnica de k-medias y niveles generados a través de la teoría de respuesta al ítem (TRI), en busca de similitudes o diferencias.

4 Resultados

A continuación, se mostrarán los resultados obtenidos al realizar el análisis de la serie de tiempo correspondiente al periodo de 2000-1 al 2019-2, del área de matemáticas haciendo uso de la metodología de Box-Jenkins y el análisis por conglomerados realizados sobre las condiciones socioeconómicas asociadas a los estudiantes que presentaron el examen Saber 11 en el periodo 2019-2.

4.1 Análisis de los resultados en matemáticas desde el año 2000

Tras la obtención de los 40 puntos muestrales correspondientes a los resultados en las pruebas de matemáticas desde el periodo 2000-1 hasta el periodo 2019-2, se grafican los puntajes en relación con la fecha de aplicación

¹DataIcfes

de los exámenes (2 por año) para la obtención de la serie de tiempo.

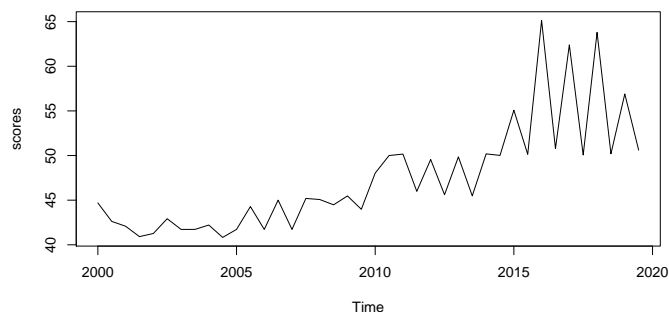


Figura 2: Serie de tiempo de los resultados de matemáticas en las pruebas Saber 11 desde el año 2000

Cabe resaltar como la serie presenta picos o valores máximos en puntos o fechas específicas, estas fechas específicas corresponden generalmente a la primera aplicación de cada uno de los años en estudio o calendario B; en estas aplicaciones participan generalmente colegios privados con niveles socioeconómicos altos, sin embargo, el número de estudiantes que desarrollan la pruebas en este calendario es menor al número de estudiantes que presenta la prueba en el calendario A.

Al descomponer la serie en sus componentes (4.1) de tendencia, estacionalidad y aleatoriedad se observa como la serie definida tiene una tendencia creciente que se puede relacionar directamente con las mejoras observadas en esta área desde el año 2000 y que refleja en cierta manera los esfuerzos realizados por los gobiernos nacionales y locales en la mejora de la educación, algo que también es apreciable en los resultados de las pruebas internacionales PISA aplicadas por la OECD², estando aun por debajo de la media internacional.

En cuanto a la estacionalidad se observa que la serie tiene picos y valles, asociados como se mencionó con anterioridad a las diferentes instituciones que se presentan en los diferentes calendarios académicos. La componente aleatoria de la serie también presenta un comportamiento similar debido a las mismas razones. En relación con el comportamiento de la serie se pueden observar 3 momentos en el que el comportamiento de la serie cam-

bia, al menos en el aumento relacionado con los puntajes descritos: desde el periodo inicial 2000-1 hasta aproximadamente el periodo 2009-2, en el cual hay un valor de medios y de varianza similares. Al igual que para los periodos de 2010-1 hasta 2014-1 y desde el año 2014-2 hasta 2019-2 donde los valores medios de los puntajes se han ido incrementando de manera paulatina.

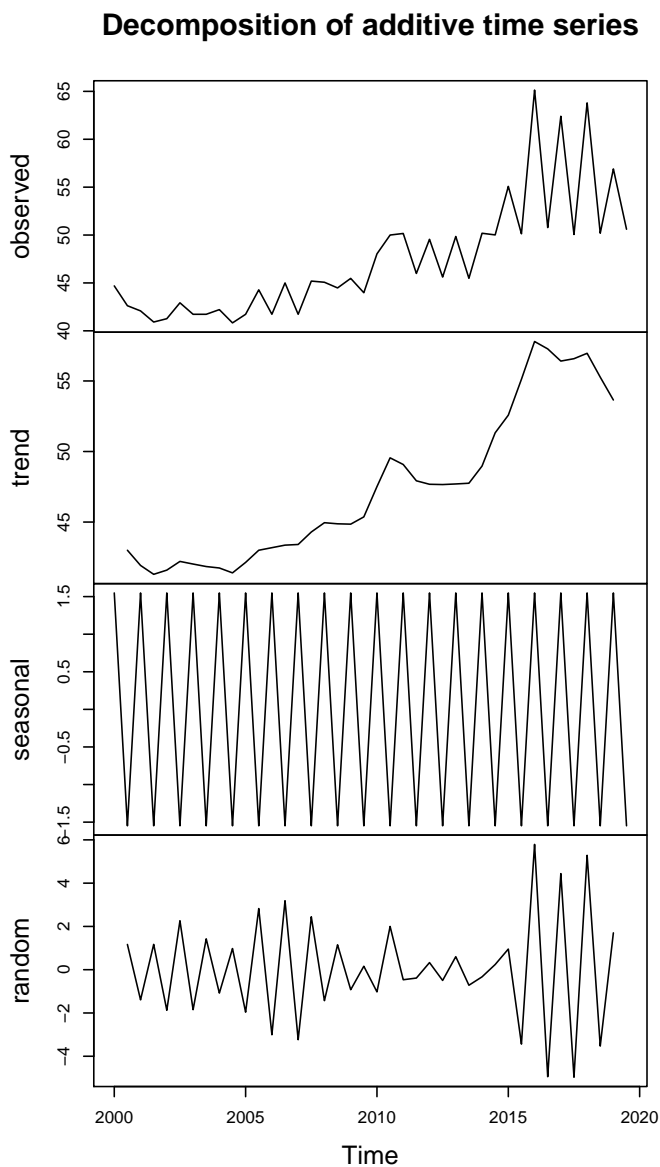


Figura 3: Descomposición de la serie de tiempo mostrada de resultados en matemáticas. Componente de tendencia, estacionalidad y aleatoriedad.

²https://www.oecd.org/pisa/publications/PISA2018_CN_COL_ESP.pdf

Una primera mirada a la media y la varianza de la serie nos indica de manera inmediata que la serie no es estacionaria dado que estos valores no son constantes, resultado que se puede comprobar mediante las pruebas de Dickey-Fuller aumentada y la prueba Phillips-Perron para la estacionariedad de las series con p-valores mayores a 0.4, con una significancia del 5 % y siendo la hipótesis alternativa la estacionariedad de la serie. Desde este punto se puede notar que los valores de autocorrelación tienden a disminuir y converger, y dentro de la gráfica de autocorrelación parcial ya destaca el lag 1, sin embargo, este comportamiento debe ser confirmado una vez que la serie sea estacionaria.

Para una diferencia de $d = 2$ las pruebas de estacionariedad nos indican que la serie en este estado ya es estacionaria con p-valores menores a 0.01 tanto para la prueba de Dickey-Fuller y Phillips-Perron. Una vez que la serie es estacionaria es posible determinar el orden p y q de la serie mediante las funciones de autocorrelación y autocorrelación parcial. Para el caso de la función de autocorrelación ACF, que nos da una idea del orden q de la serie presenta valores alternantes que tienden a 0. Dentro de los valores que poseen mayor significancia está el lag 1 y lag 2. En el caso de la función de autocorrelación parcial, la cual nos brinda una idea del orden p de la serie, se observa también valores decrecientes de forma sinusoidal y con el lag siendo el más significativo. Con esto en mente es posible determinar algunos posibles modelos que se ajusten a la serie en estudio dados los valores de p, q, d y teniendo en cuenta que no hay un componente estacional periódica dentro de la serie. Algunos de los modelos a considerar son ARIMA(0,2,1), ARIMA(1,2,1), ARIMA(1,2,2), etc. El mejor de estos modelos teniendo en cuenta el criterio de información Bayesiano BIC y el criterio de información de Akaike[20] es el ARIMA(1,2,1)

Modelo	BIC	AIC
ARIMA(1,2,0)	216.17	212.89
ARIMA(1,2,2)	205.41	198.86
ARIMA(1,2,1)	202.18	197.27
ARIMA(0,2,1)	256.21	252.94

Cuadro 4: Valores de los criterios de información de AIC y BIC para los posible modelos a ajustar en la serie de estudio.

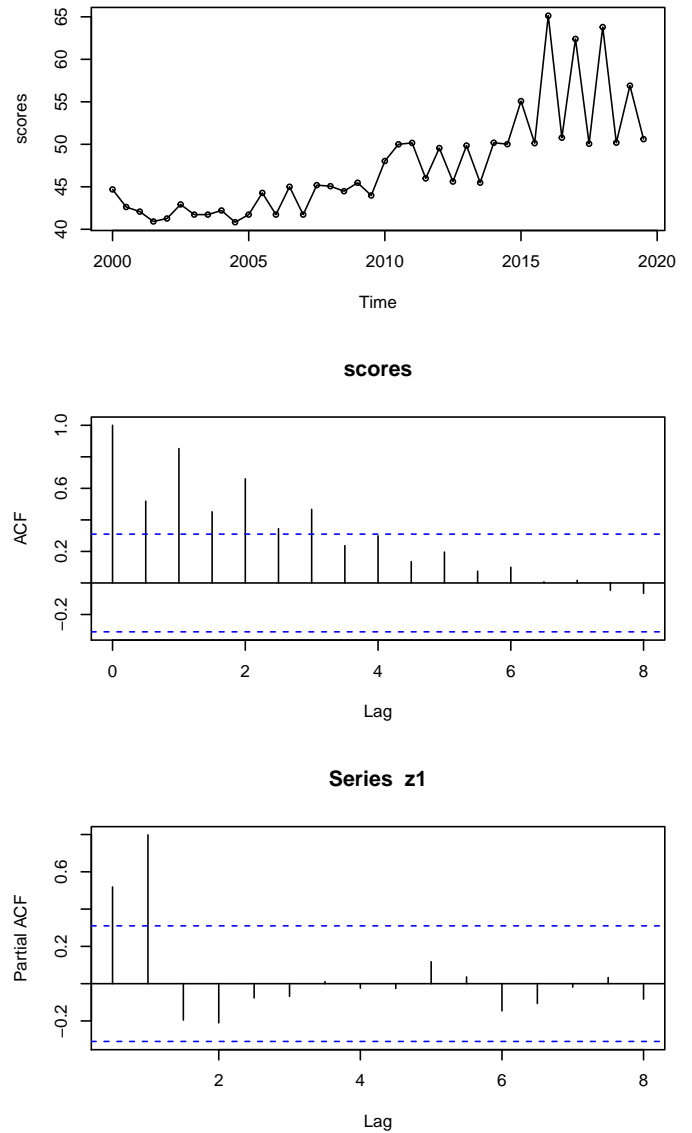


Figura 4: Serie original junto con sus graficas de autocorrelación y autocorrelación parcial. La serie en este estado no es estacionaria.

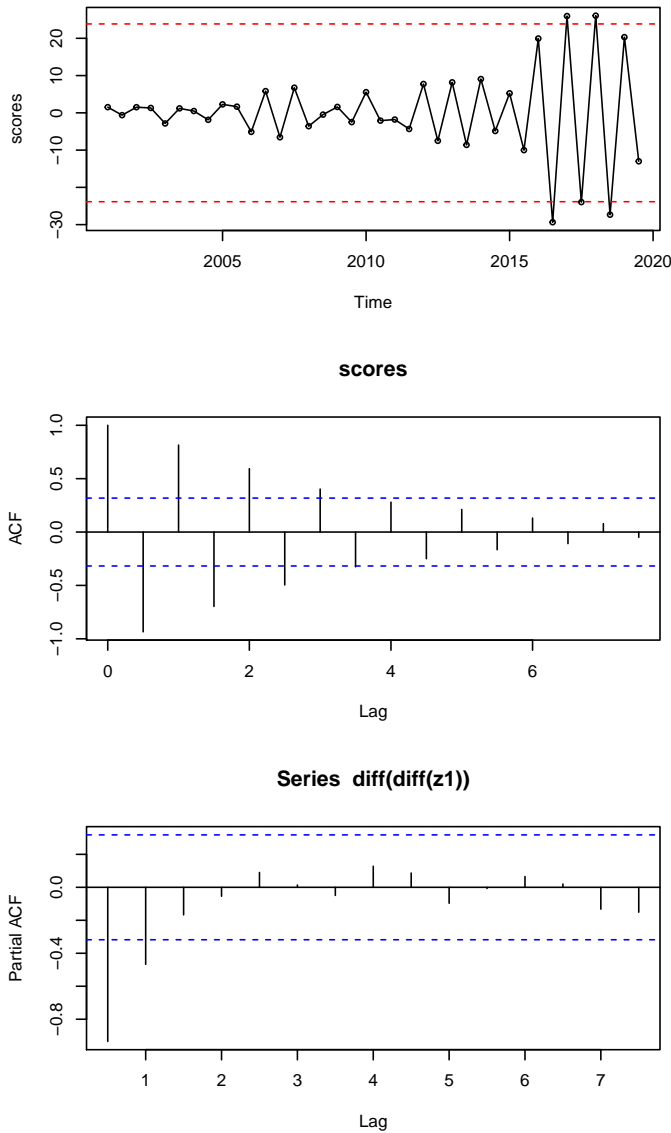


Figura 5: Serie diferenciada $d = 2$ con sus graficas de autocorrelación y autocorrelación parcial. La serie en este estado es estacionaria.

Ahora que el modelo ha sido seleccionado, la siguiente etapa acorde al modelo Box-Jenkins[20], corresponde a evaluar el modelo. Una primera aproximación se puede realizar a través de los residuales del modelo (Fig. 6). En una primera instancia se evalúa el modelo realizando una superposición del modelo generado y de la serie estudiada, esta grafica puede ser observada en al Fig. 6 parte superior izquierda. A pesar de tener algunos rezagos en los picos de la serie, comportamiento

por periodos es positivo y es capaz de predecir los valles de la serie. En relación con los residuales la mayoría de los puntos no presentan variaciones significantes lo que implica una buena descripción de la serie ajustada. En cuanto a los valores de autocorrelación y autocorrelación ninguno de ellos es significativo, aunque la gráfica Q-Q de normalidad implica que estos residuales no provienen de una distribución normal, al menos en los extremos.

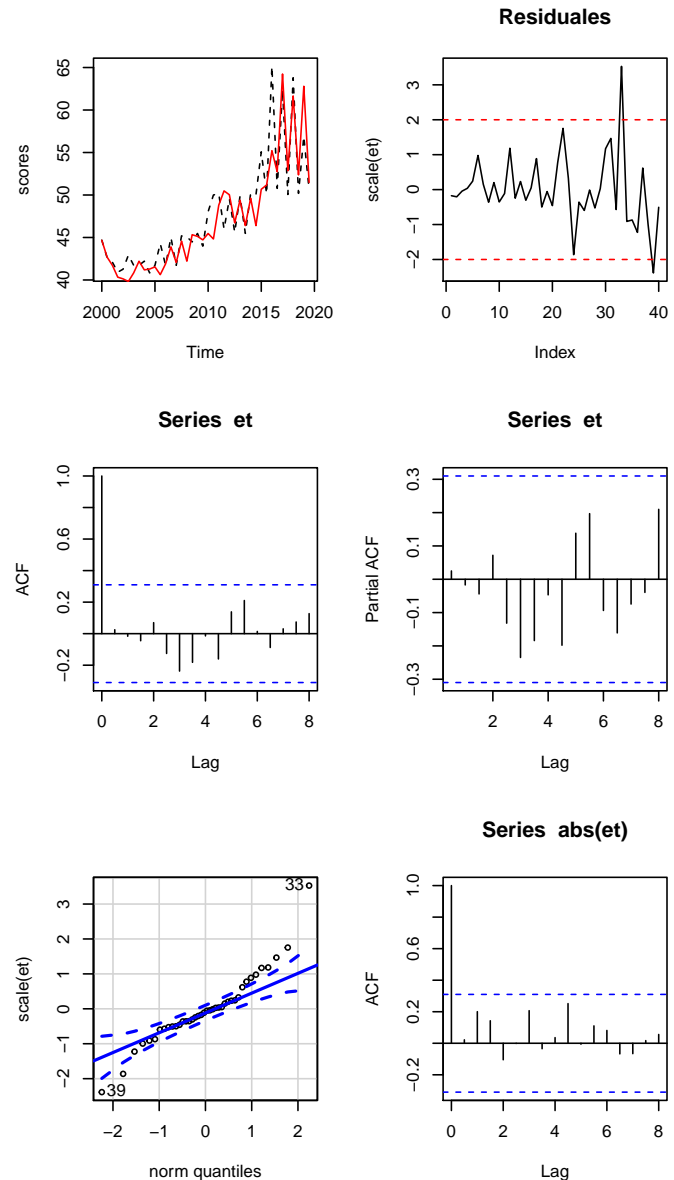


Figura 6: Comprobación de residuales generados con el modelo ARIMA(1,2,1)

Dentro de las pruebas realizadas al modelo se en-

cuenta la prueba de independencia de residuos (Box test), cuya hipótesis nula establece que los residuales están independientemente relacionados y con un p-valor de 0.96 no hay evidencia estadística para rechazar la hipótesis nula. La prueba de normalidad de los residuos confirma lo que gráficamente se presentó con anterioridad: los residuos no provienen de una distribución normal con un p-valor de $8.02e-27$ se rechaza la hipótesis nula. Finalmente, la prueba de aleatoriedad con un p-valor de 0.3 nos indica que no hay suficiente evidencia estadística para rechazar la hipótesis nula que estable que los residuales son aleatorios.

De esta manera se comprueba que el modelo propuesto y ajustado para la serie de puntajes en el área de matemáticas desde el año 2000-1 hasta el 2019-2 es estadísticamente consistente o dicho de otra manera, el modelo propuesto es lo suficientemente bueno para describir y predecir la serie en estudio (Fig. 7).

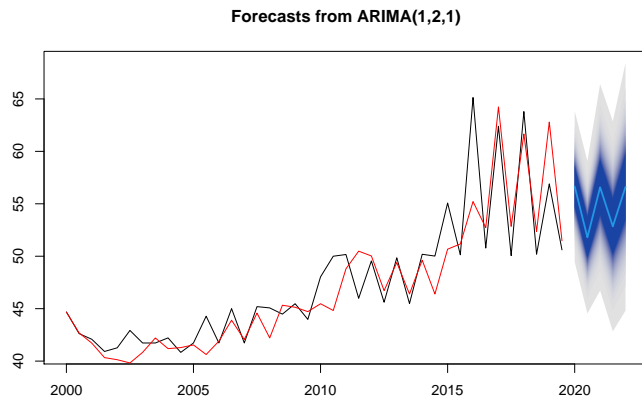


Figura 7: Predicción para un periodo de $h=5$ (2.5 años) para los puntajes de matemáticas en las pruebas Saber 11 en el área de matemáticas.

La predicción del modelo se hace para 5 periodos en el futuro, lo que corresponde a una predicción de 5 aplicaciones en la pruebas Saber 11 junto con los intervalos de confianza del 80 % y 95 %, lo primero que se puede observar es como la relación de máximos y mínimos esa mantenida en relación con los periodos de aplicación. Adicionalmente acorde con el modelo generado se espera una pequeña disminución en los valores medios de los resultados en matemáticas para los siguientes 5 periodos.

Periodo	Predicción	L 80	H 80	L 95	H 95
2020-1	56.65	53.07	60.22	51.18	62.12
2020-2	51.79	48.18	55.41	46.27	57.32
2021-1	56.57	51.67	61.46	49.08	64.06
2021-2	52.84	47.86	57.82	45.22	60.46
2022-2	56.62	50.74	62.49	47.63	65.60

Cuadro 5: Valores predichos para los siguientes 5 periodos de aplicación desde el periodo 2019-2, junto con los intervalos de confianza al 80 % y 90 %.

4.2 Generación de niveles mediante socio-económicos k-medias y comparación con niveles NSE

Dentro de los datos existentes en la base de datos publicadas por el ICFES existe información relacionada no solo con los puntajes de los estudiantes en las diferentes áreas evaluadas, sino también la información socioeconómica relacionada al contexto de los estudiantes, información proveniente de los cuestionarios socio-económicos anexos a las pruebas. Los niveles socio-económicos NSE generados por el ICFES como se mencionó anteriormente proviene de las respuestas de los estudiantes al cuestionario y del análisis hecho mediante la teoría de respuesta al ítem. Otra metodología que vale la pena explorar en la definición de estos niveles es asociar y agrupar directamente las características de los estudiantes y sus puntajes, esto mediante un análisis de clúster, en concreto mediante el algoritmo de k-medias asumiendo el problema como no supervisado. Para realizar esto de manera inicial es posible generar directamente grupos únicamente con los puntajes generados para cada una de las componentes evaluadas, estos grupos darán cuenta de los puntajes que menores distancias posean o dicho de otra manera los que sean más similares. Una vez que los grupos (etiquetas) sean generados se asocian directamente con las características socioeconómicas de los estudiantes con la finalidad de crear una caracterización cada uno de los grupos generados. A continuación, se estudiará esta metodología usando los resultados del periodo 2019-2, cuyos individuos rozan los 500000 una vez que la base de datos está limpia. Esta metodología requiere inicialmente determinar cuál es número óptimo de clases, para ello se aplica el método de k-medias para los puntajes definidos, agrupándolos inicialmente rango de $k = 2$ a $k = 10$ y calculando el WGSS óptimo. Este valor teóricamente siempre será menor a medida que la

cantidad de grupos aumente, sin embargo, hay un valor óptimo de k dado por la regla del codo, el cual nos habla de que el valor óptimo será aquel en el cual se genere un codo en la gráfica de WGSS en función de k .

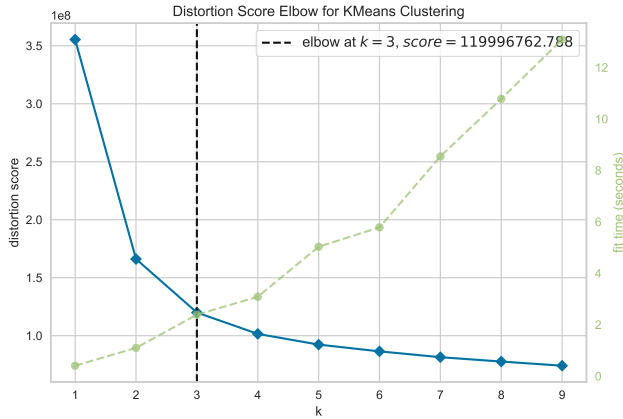


Figura 8: Número óptimo de clúster o conglomerados usando como criterio el WGSS y los tiempo de agrupamiento asociados.

Para el caso concreto de nuestro estudio este valor corresponde a 3, la primera diferencia con la metodología usada por el ICFES con base a la TRI la cual genera 4 grupos. También cabe resaltar que el tiempo de agrupamiento aumenta en relación con el número de grupos definidos, entre mayor sea este mayor será el tiempo que tarda el algoritmo k-medias.

Otro de las métricas útiles a la hora de definir el número óptimo de clases es el coeficiente de silueta, el cual nos da una idea de cuan similar es un objeto en relación con su propio clúster comparados a otros clústeres (cohesión), este valor tiene un rango de -1 a 1. Un valor de 1 indica que el punto está bien emparejado en su clúster y mal emparejado con los clústeres vecinos y de manera contraria para el valor de -1. En el caso de estudio este valor es más cercano a 1 $k = 2$ sin embargo la gráfica de WGSS en función de k , nos indica que el valor óptimo es de 3 y coeficiente de silueta es alto a pesar de no ser el mayor posible.

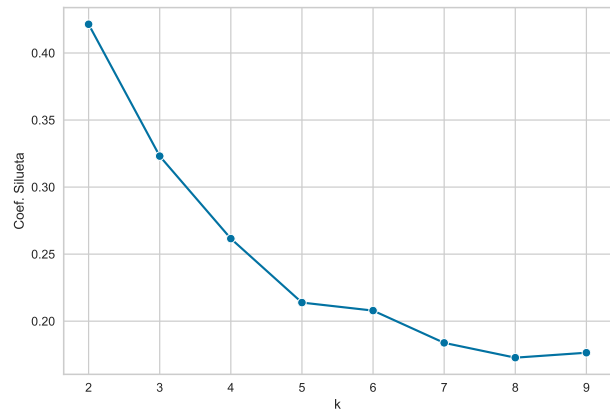


Figura 9: Coeficiente de silueta para conglomerados en un rango de 1 a 10. Entre más cercano el valor sea de 1, más consistente es el agrupamiento.

Una vez definido este valor se realiza la agrupación de la base obteniendo 3 grupos característicos etiquetados como 0,1,2, en términos de puntajes globales estos grupos se caracterizan de la siguiente manera:

Grupo k	Individuos	Puntaje global (μ)
0	172145	194.13
1	206476	253.17
2	121044	316.70

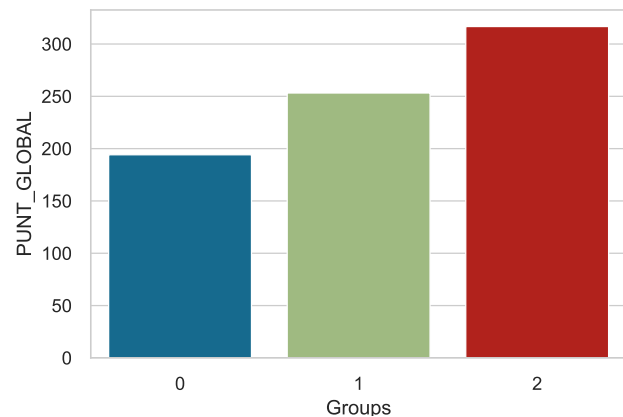


Figura 10: Valores medios de los puntajes globales para cada uno de los niveles generados mediante k-medias.

De manera más específica estos grupos se caracte-

rizan en términos de los puntajes medios de cada uno de los compontes evaluadas. El grupo 0 se caracteriza por valor bajos, el grupo 2 se caracteriza por valores medios y el grupo 2 se caracteriza por valores altos, lo cual sugiere una clasificación orgánica en estos tres niveles: bajo, medio, alto.

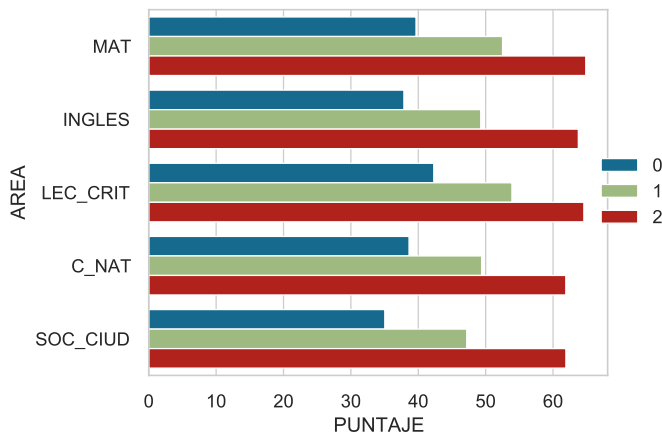


Figura 11: Valores medios de los puntajes por componente para cada uno de los niveles generados mediante k-medias.

Variable de caracterización	Niveles K		
	0	1	2
Internet (NO)	X		
Computador (NO)	X		
Horno microondas o gas (NO)	X	X	
Lavadora (SI)	X	X	X
Computador (SÍ)		X	X
Automóvil particular (NO)			X
Horno microondas o gas (SÍ)			X
Automóvil particular (SÍ)			X
Internet (SÍ)		X	X
Educación madre: primaria incompleta	X		
Educación madre: bachillerato completo	X	X	
Educación madre: (profesional completa)			X
Educación padre: primaria incompleta	X		
Educación padre: bachillerato completo		X	
Educación padre: (profesional completa)		X	X

Cuadro 6: Descriptores socioeconómicos para los diferentes niveles K

El cuadro 5 asocia los grupos con algunas de las características socioeconómicas presentes en cada uno

de los individuos o estudiantes, es notorio como los estudiantes del grupo 1 se caracterizan por la no posesión de computador, servicio de internet y horno microondas, además los padres y madres de los estudiantes de este nivel se caracterizan mayoritariamente por tener una educación primaria incompleta, aunque en algunos casos también hay madres con educación secundaria completa.

En el caso del nivel $k = 1$, que podría asociarse con estudiantes con resultados medios en la prueba por componentes se caracterizan por tener acceso a computador y servicio de internet, también por tener en sus viviendas acceso a horno microondas o de gas y en cuanto a las madres y padres mayoritariamente poseen educación secundaria completa.

Para el caso del nivel $k = 2$ el cual podría asociarse con estudiantes con resultados altos en las pruebas por componentes, se caracterizan tener acceso servicios como internet, computador, horno microondas o de gas, acceso a automóvil particular y mayoritariamente con padres y madres que tienen formación profesional.

Este análisis de agrupamiento o de clústeres surge como una metodología alterna, algo más practica que la actualmente se usa en el ICFES para el análisis del contexto socioeconómico de los estudiantes la cual se basa en inferir a partir de la teoría de respuesta al ítem este contexto teniendo en cuenta las respuestas a los cuestionarios asociados. En contra partida esta metodología realiza la separación de los niveles directamente de los resultados asociados a las componentes evaluadas y los caracteriza en base a las respuestas más frecuentes de los estudiantes en los cuestionarios socioeconómicos.

5 Conclusiones

Los puntajes de las pruebas Saber 11 en el área de matemáticas son las que menos variaciones en el tiempo han tenido en relación con los contenidos y la forma de evaluación lo que permite como una aproximación académica realizar un análisis mediante series de tiempo con la finalidad de modelar la evolución y posibles escenarios a futuro de esta componente en particular de las pruebas Saber 11. El modelo resultante de la serie corresponde a un modelo ARIMA(1,2,1) integrado de grado 2, con componentes de media móvil y autorregresivo; acorde con este modelo se espera el mismo comportamiento de picos y valles acorde con el calendario

de aplicación, y con puntajes algo menores a los registrados en los últimos 5 periodos. Este modelo no solo demostró ser consistente estadísticamente sino también coherente con las predicciones hechas y el comportamiento general de la serie. Como trabajo futuro queda comparar el modelo con otras implementaciones de predicciones mediante series de tiempo como lo pueden ser las aplicaciones mediante redes neuronales recurrentes comparando no lo solo la calidad del ajuste sino la calidad y la precisión en las predicciones realizadas

Al explorar los niveles socioeconómicos NSE que derivan de la TRI el ICFES especifica cuatro niveles asociados a un constructo teórico que ha mostrado ser funcional, pero es posible de ser mejorado. La aproximación realizada en este trabajo aborda de manera practica este problema generando niveles de agrupación mediante el algoritmo de clúster de k-medias y los resultados de

cada una de las componentes evaluadas. Ya no se tiene 4 niveles, sino 3 que su vez son perfectamente caracterizables por medio de las respuestas a los cuestionarios socioeconómicos realizados por los estudiantes evaluados en sus jornadas de aplicación. Esta propuesta permite generar una agrupación más orgánica y que a su vez da cuenta de los fenómenos más característicos estos niveles para que sean abordados como problemáticas sociales por los programas estatales y locales, con la finalidad de mejorar estas condiciones históricas que han demostrado influir en el desarrollo de los estudiantes y sus resultados académicos. Como trabajos futuros queda la aplicación de estos niveles mediante algún algoritmo de clasificación que permita establecer directamente el nivel al que pertenece el estudiante, algoritmos como arboles de decisión o bosques de decisión permiten abordar este problema.

Referencias

- [1] Carvajal, P., Trejos, A., & Soto, J. (2004). APLICACIÓN DEL ANÁLISIS DISCRIMINANTE PARA EXPLORAR LA RELACIÓN ENTRE EL EXAMEN DE ICFES Y EL RENDIMIENTO EN ALGEBRA LINEAL DE LOS ESTUDIANTES DE INGENIERÍA DE LA UTP EN EL PERÍODO 2001-2003. *Scientia et Technica*, 25, 191–196. <https://doi.org/10.1007/s10236-012-0520-1>
- [2] Timarán, R. (2009). Detección de Patrones de Bajo Rendimiento Académico y Deserción Estudiantil con Técnicas de Minería de Datos. *Memorias de La 8a Conferencia Iberoamericana En Sistemas, Cibernética e Informática CISCI*, 8, 45–50. <http://www.iiis.org/CDs2009/CD2009CSC/CISCI2009/PapersPdf/C692YV.pdf>
- [3] Baron, J. (2014). Desempeño relativo de los graduados en el área de educación en el examen de estado del ICFES. *Educación y Desarrollo Regional En Colombia. Serie: Colección de Economía Regional Banco de La República*, 129–157. <https://repositorio.banrep.gov.co/handle/20.500.12134/6708>
- [4] Tobón, D., Posada, H., & Ríos, P. (2009). Determinants of performance of the Schools in Medellin in the high-school graduation-year test (ICFES). 22, 22, 313–333. http://www.scielo.org.co/scielo.php?pid=S0120-35922009000100015&script=sci_abstract&tlng=es
- [5] Barón, J. (2012). Diferencias en las características de los estudiantes y la brecha de rendimiento académico entre Barranquilla y Bogotá: una descomposición semiparamétrica. *Ensayos Sobre POLÍTICA ECONÓMICA*, 30, 164–215. http://www.scielo.org.co/scielo.php?pid=S0120-44832012000200005&script=sci_abstract&tlng=es
- [6] ¿Como se construye el índice de nivel socioeconómico (INSE) en el contexto de las pruebas saber? (2019). *Saber al Detalle. ICFES*, 4, 2–6. <https://www.icfes.gov.co/documents/20143/1519705/Edicion%204%20-%20boletin%20saber%20al%20detalle.pdf>
- [7] Duque, A., & Ortiz, J. (2013). Pruebas ICFES Saber 11 y su relación con el desempeño académico en estudiantes de primer semestre de psicología. *Cuadernos Hispanoamericanos De Psicología*, 26–35. <https://revistas.unbosque.edu.co/index.php/CHP/article/view/1355>

- [8] Restrepo, P., & Alviar, M. (2004). El logro académico y el efecto colegio en las pruebas ICFES en Antioquia. *Lecturas de Economía*, 60, 67–95. <https://revistas.udea.edu.co/index.php/lecturasdeeconomia/article/view/2723/2176>
- [9] Timarán, R., Calderón, A., & Jiménez, J. (2013). Descubrimiento de perfiles de deserción estudiantil con técnicas de minería de datos. *Revista Vínculos*, 10, 373–383. <https://doi.org/10.14483/2322939X.4687>
- [10] Bahamón, M., & Reyes, L. (2014). Caracterización de la capacidad intelectual, factores sociodemográficos y académicos de estudiantes con alto y bajo desempeño en los exámenes Saber Pro - año 2012. *Avances En Psicología Latinoamericana*, 32, 459–476. <https://revistas.urosario.edu.co/index.php/apl/article/view/apl32.03.2014.01>
- [11] Shumway, R., & Stoffer, D. (2017). *Time Series Analysis and its Applications with R Examples* (4th ed.). Springer.
- [12] Everitt, B., & Hothorn, T. (2011). *An Introduction to Applied Multivariate Analysis with R* (2nd ed.). Springer.
- [13] ICFES. (2017, September). Examen de Estado de la educación medial. Resultados del período 2005 – 2010, Datalcfes, Documentación SABER 11 (No. 17). <https://www.icfes.gov.co/web/guest/investigadores-y-estudiantes-posgrado/acceso-a-bases-de-datos>
- [14] ICFES. (2018). GUÍA INTRODUCTORIA AL DISEÑO CENTRADO EN EVIDENCIAS (No. 1). <https://www.icfes.gov.co/documents/20143/1500084/Guia+introdutoria+al+Diseno+Centrado+en+Evidencias.pdf>
- [15] Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann Books.
- [16] ICFES. (2019). Datalcfes, Documentación SABER 11. <https://www.icfes.gov.co/web/guest/investigadores-y-estudiantes-posgrado/acceso-a-bases-de-datos>
- [17] ICFES. (2019b). Descriptores NSE Saber 11, Datalcfes. https://www2.icfes.gov.co/documents/20143/1885630/8.+Descriptores_NSE_Saber11.pdf
- [18] ICFES. (2017a). Informe de resultados nacionales Saber 11 (2014-2 – 2016-2). <https://www.icfes.gov.co/documents/20143/193784/Informe%20nacional%20de%20resultados%20del%20examen%20saber%2011%20-%202014-2%20-%202016-2.pdf>
- [19] Geiron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras and TensorFlow: concepts, tools, and techniques to build intelligent systems* (2nd ed.). O'Reilly.
- [20] Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control* (Wiley Series in Probability and Statistics) (5th ed.). Wiley.