



UNIVERSIDAD NACIONAL DE COLOMBIA

MAESTRIA EN CIENCIAS ESTADISTICA
FACULTAD DE CIENCIAS

TRABAJO DE GRADO

*Análisis del rendimiento académico y de los
resultados en pruebas Saber 11° usando Técnicas
Estadísticas Multivariadas (TEM) como un insumo de
autoevaluación de un colegio de Medellín*

Iván Andrés Giraldo Salguero

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Estadística
Medellín, Colombia
Enero de 2024

ANÁLISIS DEL RENDIMIENTO ACADÉMICO Y DE LOS
RESULTADOS EN PRUEBAS SABER 11° USANDO
TÉCNICAS ESTADÍSTICAS MULTIVARIADAS
(TEM) COMO UN INSUMO DE AUTOEVALUACIÓN DE
UN COLEGIO DE MEDELLÍN.

*Trabajo de Grado como requisito parcial para optar
al título de: Magíster en Ciencias-Estadística*

Estudiante: Iván Andrés Giraldo Salguero
Director: Raúl Alberto Pérez Agamez, Ph.D

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Estadística
Medellín, Colombia
Enero de 2024

Índice general

Prólogo	V
Resumen	VII
Abstract	VIII
Índice de Figuras	X
Índice de Tablas	XIII
1. Introducción	1
1.1. Estado del Arte	2
2. Objetivos	9
2.1. Objetivo General	9
2.2. Objetivos Específicos	9
3. Marco Teórico	11
3.1. Introducción al Análisis de Componentes Principales (ACP)	12
3.2. Proceso para hallar las Componentes Principales	13
3.3. Fundamentación Algebraica del ACP	14
3.3.1. Ajuste de Mínimos Cuadrados	14
3.3.2. ¿Por qué el uso de la palabra “mejor”?	15
4. Análisis Descriptivo	17
4.1. Base de Datos	17
4.2. Análisis de las notas de los diferentes Cursos	17
4.2.1. Análisis de notas del grado 6°	17
4.2.2. Análisis de notas del grado 7°	20
4.2.3. Análisis de notas del grado 8°	23
4.2.4. Análisis de notas del grado 9°	26
4.2.5. Análisis de notas del grado 10°	29
4.2.6. Análisis de notas del grado 11°	32
4.2.7. Análisis de Resultados Prueba Saber 11°	34

5. Métodos Estadísticos Multivariados	39
5.1. Análisis de Componentes Principales (ACP) Grado 6°	40
5.2. Análisis de Componentes Principales (ACP) Grado 7°	42
5.3. Análisis de Componentes Principales (ACP) Grado 8°	44
5.4. Análisis de Componentes Principales (ACP) Grado 9°	47
5.5. Análisis de Componentes Principales (ACP) Grado 10°	49
5.6. Análisis de Componentes Principales (ACP) Grado 11°	51
5.7. Análisis de Componentes Principales (ACP) Prueba Saber 11°	58
6. Conclusiones	65
6.1. Generales	65
6.2. Recomendaciones Futuras	67
A. Apéndice: Soportes Algebraicos	69
A.1. Primera sección	69
A.1.1. Teorema	69
A.1.2. Proposición 1:	69
A.1.3. Proposición 2:	70
A.1.4. Componentes principales de variables estandarizadas	70
A.1.5. Proposición 3:	71
A.1.6. Componentes principales muestrales	71
A.1.7. Proposición 4:	72
Bibliografía	75

Prólogo

Cuando me embarqué en esta apasionante travesía académica en el campo de la estadística, nunca imaginé que me llevaría a descubrir un mundo fascinante de números, patrones y significados ocultos. El camino ha sido arduo, repleto de desafíos y momentos de profundo asombro, pero cada paso ha valido la pena. Hoy, con una mezcla de humildad y orgullo, presento este trabajo de grado, el fruto de años de dedicación y esfuerzo.

La estadística, como el arte de extraer conocimiento de los datos, se erige como una ciencia que influye en las decisiones que dan forma a nuestras vidas cotidianas. Desde analizar fenómenos naturales hasta estudiar complejas interacciones sociales, la estadística se ha vuelto una herramienta esencial para comprender nuestro mundo en constante cambio.

En mi formación como matemático con énfasis en estadística, he sido testigo de la poderosa capacidad de esta disciplina para revelar verdades ocultas y respaldar decisiones informadas. También he aprendido a valorar su capacidad para destacar la incertidumbre inherente a la realidad y cómo la búsqueda de la verdad en los datos puede ser un proceso enriquecedor, a veces incluso más que el resultado en sí.

Quiero expresar mi profundo agradecimiento a mi asesor, el profesor Raúl Alberto Agamez, PhD, y a todos los profesores que me han guiado con su sabiduría, contribuyendo al desarrollo de una visión más profunda y perspicaz de la estadística. Sus consejos invaluable y su retroalimentación crítica han sido fundamentales para mi crecimiento académico y profesional.

Asimismo, agradezco a mis compañeros de clase y colegas, quienes compartieron conmigo su pasión por la estadística y me motivaron a seguir adelante cuando los desafíos parecían abrumadores. Sus debates y discusiones enriquecieron mi experiencia, enseñándome a abordar los problemas desde diversas perspectivas.

No puedo pasar por alto el apoyo incondicional de mi familia, especialmente de mi esposa Nazly Cuadrado y amigos como Sandra, quienes estuvieron a mi lado durante este viaje. Sus palabras alentadoras y su acompañamiento me dieron la fuerza necesaria para enfrentar cada obstáculo con determinación.

Este trabajo de grado es un testimonio de mi crecimiento personal y académico. A través de sus páginas, he intentado explorar nuevas fronteras en la estadística y contribuir con mi granito de arena a esta disciplina que tanto me ha dado. Acepto que mi trabajo es solo un pequeño eslabón en la cadena del conocimiento, pero espero que pueda ser un punto de partida para futuros investigadores que deseen profundizar en estos temas.

Con humildad y entusiasmo, presento esta tesis a la comunidad académica y al mundo en general. Que la estadística continúe guiándonos hacia nuevas revelaciones y perspectivas que nos permitan tomar decisiones informadas y transformadoras.

Iván Andrés Giraldo Salguero

Bello, enero 30 de 2024

Resumen

El éxito de un colegio depende en gran medida de su nivel académico, que se puede evaluar a través del rendimiento de los estudiantes en las asignaturas que cursan a lo largo de sus años académicos, así como a través de los resultados en pruebas externas como las Pruebas Saber 11°. Estas pruebas son presentadas por todos los estudiantes de grado 11° en Colombia bajo la supervisión del Ministerio de Educación Nacional (MEN, 2015). Los resultados de estas pruebas otorgan a los colegios una clasificación a nivel nacional, lo que representa un elemento crucial para que nuevas familias elijan una institución educativa como modelo académico para sus hijos.

Las Pruebas Saber tienen un impacto significativo en la calificación de un colegio en Colombia. Estas pruebas, administradas por el Instituto Colombiano para la Evaluación de la Educación (ICFES), son una herramienta importante para medir el desempeño académico de los estudiantes en áreas clave como matemáticas, ciencias naturales, ciencias sociales y comprensión lectora. Los resultados de las Pruebas Saber se utilizan para evaluar el nivel de logro de los estudiantes en relación con los estándares educativos establecidos por el Ministerio de Educación Nacional. Además, estos resultados se utilizan para clasificar a las instituciones educativas en diferentes categorías de rendimiento, que van desde A+ (más alto rendimiento) hasta D (más bajo rendimiento).

En el presente estudio, se evaluó una base de datos de un colegio en Medellín que incluía 2586 observaciones de información académica de los estudiantes, recopilada durante tres años, desde 2019 hasta 2021. Esta información abarcó los grados sexto a undécimo. Inicialmente, se llevó a cabo un análisis descriptivo de los promedios de calificaciones generados por los estudiantes de sexto a undécimo grado. Posteriormente, se aplicó la técnica estadística multivariada conocida como Análisis de Componentes Principales (ACP).

A partir de este análisis, se presentó una tabla de valores propios para cuantificar la varianza explicada por cada componente principal. Luego, se seleccionaron las seis primeras componentes principales, que explicaban la mayor variabilidad posible de los datos en cada grado. Finalmente, se identificó que la propuesta académica del colegio, junto con los resultados en las asignaturas, contribuye a la mejora de los resultados en las pruebas Saber.

Palabras claves: Análisis multivariado, Análisis multivariable, Técnicas estadísticas multivariadas (TEM), Pruebas Saber 11°, Análisis de componentes principales (ACP).

Abstract

The success of a school depends largely on its academic level, which can be evaluated through the performance of students in the subjects they take throughout their academic years, as well as through the results in external tests such as Knowledge Tests 11th. These tests are presented by all 11th grade students in Colombia under the supervision of the Ministry of National Education (MEN, 2015). The results of these tests give schools a national ranking, which represents a crucial element for new families to choose an educational institution as an academic model for their children.

The Saber Tests have a significant impact on the rating of a school in Colombia. These tests, administered by the Colombian Institute for the Evaluation of Education (ICFES), are an important tool to measure students' academic performance in key areas such as mathematics, natural sciences, social sciences and reading comprehension. The results of the Saber Tests are used to evaluate the level of achievement of students in relation to the educational standards established by the Ministry of National Education. Additionally, these results are used to classify educational institutions into different performance categories, ranging from A+ (highest performing) to D (lowest performing).

In the present study, a database from a school in Medellín was evaluated that included 2,586 observations of student academic information, collected over three years, from 2019 to 2021. This information covered grades six through eleven. Initially, a descriptive analysis was conducted on the grade point averages generated by students in grades six through eleven. Subsequently, the multivariate statistical technique known as Principal Component Analysis (PCA) was applied.

From this analysis, a table of eigenvalues was presented to quantify the variance explained by each principal component. Then, the first six principal components were selected, which explained the greatest possible variability of the data at each grade. Finally, it was identified that the school's academic proposal, together with the results in the subjects, contributes to the improvement of the results in the Saber tests.

Keywords: Multivariate analysis, Multivariate analysis, Multivariate statistical techniques (TEM), Saber 11th Tests, Principal Component Analysis (PCA).

Índice de figuras

3.1. Ajuste de Mínimos Cuadrados del ACP (Fuente: Elaboración Propia) . . .	14
4.1. Promedio de Notas por Asignatura del 6° Grado: Años 2019-2021 (Fuente: Elaboración Propia)	20
4.2. Promedio de Notas por Asignatura del 7° Grado: Años 2019-2021 (Fuente: Elaboración Propia)	23
4.3. Promedio de Notas por Asignatura del 8° Grado: Años 2019-2021 (Fuente: Elaboración Propia)	26
4.4. Promedio de Notas por Asignatura del 9° Grado: Años 2019-2021 (Fuente: Elaboración Propia)	29
4.5. Promedio de Notas por Asignatura del 10° Grado: Años 2019-2021 (Fuente: Elaboración Propia)	31
4.6. Promedio de Notas por Asignatura del 11° Grado: Años 2019-2021 (Fuente: Elaboración Propia)	34
4.7. Promedio de Notas por Asignatura de la Prueba Saber : Años 2019-2021 (Fuente: Elaboración Propia)	37
5.1. Relación entre las 6 mejores variables y las 30 mejores observaciones del Grado 6° (Fuente: Elaboración Propia)	42
5.2. Relación entre las 6 mejores variables y las 30 mejores observaciones del Grado 7° (Fuente: Elaboración Propia)	44
5.3. Relación entre las 6 mejores variables y las 30 mejores observaciones del Grado 8° (Fuente: Elaboración Propia)	46
5.4. Relación entre las 6 mejores variables y las 30 mejores observaciones del grado 9° (Fuente: Elaboración Propia)	49
5.5. Relación entre las 6 mejores variables y las 30 mejores observaciones del grado 10° (Fuente: Elaboración Propia)	51
5.6. Relación entre las 6 mejores variables y las 30 mejores observaciones del grado 11° 2019 (Fuente: Elaboración Propia)	53
5.7. Relación entre las 6 mejores variables y las 30 mejores observaciones del grado 11° 2020 (Fuente: Elaboración Propia)	55

5.8. Relación entre las 6 mejores variables y las 30 mejores observaciones del grado 11° 2021 (Fuente: Elaboración Propia)	57
5.9. Relación entre las 5 mejores variables y las 30 mejores observaciones de la Prueba Saber 11° 2019 (Fuente: Elaboración Propia)	59
5.10. Relación entre las 5 mejores variables y las 30 mejores observaciones de la Prueba Saber 11° 2020 (Fuente: Elaboración Propia)	60
5.11. Relación entre las 5 mejores variables y las 30 mejores observaciones de la Prueba Saber 11° 2021 (Fuente: Elaboración Propia)	62

Índice de tablas

4.1. Calificaciones obtenidas por los estudiantes de grado 6° durante los años escolares 2019, 2020 y 2021 en una institución de Medellín.	18
4.2. Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 6° durante el año 2019 de una institución de Medellín.	18
4.3. Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 6° durante el año 2020 de una institución de Medellín.	19
4.4. Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 6° durante el año 2021 de una institución de Medellín.	19
4.5. Calificaciones obtenidas por los estudiantes de grado 7° durante los años escolares 2019, 2020 y 2021 en una institución de Medellín.	21
4.6. Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 7° durante el año 2019 de una institución de Medellín.	21
4.7. Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 7° durante el año 2020 de una institución de Medellín.	21
4.8. Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 7° durante el año 2021 de una institución de Medellín.	22
4.9. Calificaciones obtenidas por los estudiantes de grado 8° durante los años escolares 2019, 2020 y 2021 en una institución de Medellín.	24
4.10. Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 8° durante el año 2019 de una institución de Medellín.	24
4.11. Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 8° durante el año 2020 de una institución de Medellín.	25
4.12. Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 8° durante el año 2021 de una institución de Medellín.	25
4.13. Calificaciones obtenidas por los estudiantes de grado 9° durante los años escolares 2019, 2020 y 2021 en una institución de Medellín.	27
4.14. Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 9° durante el año 2019 de una institución de Medellín.	27
4.15. Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 9° durante el año 2020 de una institución de Medellín.	27

4.16. Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 9° durante el año 2021 de una institución de Medellín.	28
4.17. Calificaciones obtenidas por los estudiantes de grado 10° durante los años escolares 2019, 2020 y 2021 en una institución de Medellín.	30
4.18. Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 10° durante el año 2019 de una institución de Medellín.	30
4.19. Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 10° durante el año 2020 de una institución de Medellín.	30
4.20. Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 10° durante el año 2021 de una institución de Medellín.	31
4.21. Calificaciones obtenidas por los estudiantes de grado 11° durante los años escolares 2019, 2020 y 2021 en una institución de Medellín.	32
4.22. Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 11° durante el año 2019 de una institución de Medellín.	32
4.23. Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 11° durante el año 2020 de una institución de Medellín.	33
4.24. Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 11° durante el año 2021 de una institución de Medellín.	33
4.25. Calificaciones obtenidas por los estudiantes de grado 11° en las Pruebas Saber durante los años escolares 2019, 2020 y 2021 en una institución de Medellín.	35
4.26. Medidas de dispersión de los resultados obtenidos por los estudiantes de grado 11° en las pruebas saber durante el año 2019 de una institución de Medellín.	35
4.27. Medidas de dispersión de los resultados obtenidos por los estudiantes de grado 11° en las pruebas saber durante el año 2020 de una institución de Medellín.	36
4.28. Medidas de dispersión de los resultados obtenidos por los estudiantes de grado 11° en las pruebas saber durante el año 2021 de una institución de Medellín.	36
5.1. Calificaciones obtenidas por los estudiantes de grado 6° durante los años escolares 2019, 2020 y 2021 en una institución de Medellín.	40
5.2. Tabla de valores propios asociada a cada una de las CP del Grado 6°	41
5.3. Tabla de vectores propios de la matriz de correlación grado 6°	41
5.4. Calificaciones obtenidas por los estudiantes de grado 7° durante los años escolares 2019, 2020 y 2021 en una institución de Medellín.	43
5.5. Tabla de Valores Propios Grado 7°	43
5.6. Tabla de vectores propios de la matriz de correlación grado 7°	43
5.7. Calificaciones obtenidas por los estudiantes de grado 8° durante los años escolares 2019, 2020 y 2021 en una institución de Medellín.	45

5.8. Tabla de Valores Propios Grado 8°	45
5.9. Tabla de vectores propios de la matriz de correlación grado 8°	46
5.10. Calificaciones obtenidas por los estudiantes de grado 9° durante los años escolares 2019, 2020 y 2021 en una institución de Medellín.	47
5.11. Tabla de Valores Propios Grado 9°	48
5.12. Tabla de vectores propios de la matriz de correlación grado 9°	48
5.13. Calificaciones obtenidas por los estudiantes de grado 10° durante los años escolares 2019, 2020 y 2021 en una institución de Medellín.	50
5.14. Tabla de Valores Propios Grado 10°	50
5.15. Tabla de vectores propios de la matriz de correlación grado 10°	50
5.16. Tabla de Valores Propios Grado 11° 2019	52
5.17. Tabla de vectores propios de la matriz de correlación grado 11° 2019	53
5.18. Tabla de Valores Propios Grado 11° 2020	54
5.19. Tabla de vectores propios de la matriz de correlación grado 11° 2020	55
5.20. Tabla de Valores Propios Grado 11° 2021	56
5.21. Tabla de vectores propios de la matriz de correlación grado 11° 2021	56
5.22. Tabla de Valores Propios Prueba Saber 11° 2019	58
5.23. Tabla de vectores propios de la matriz de correlación Prueba Saber 11° 2019	58
5.24. Tabla de Valores Propios Prueba Saber 11° 2020	60
5.25. Tabla de vectores propios de la matriz de correlación grado 11° 2020	60
5.26. Tabla de Valores Propios Prueba Saber 11° 2021	61
5.27. Tabla de vectores propios de la matriz de correlación grado 11° 2021	61

Capítulo 1

Introducción

El marco legal del sistema educativo colombiano se basa en tres pilares fundamentales. En primer lugar, la Constitución Política de 1991, que establece la responsabilidad compartida de la sociedad, la familia y el estado, de garantizar una educación de calidad y equidad Congreso de la República (1991). En segundo lugar, la Ley 115 de 1994, cuyo artículo 80 aborda la creación de un sistema nacional de evaluación de la educación, que evalúa tanto los logros de los estudiantes como el desempeño de los docentes MEN (1994). Por último, la Ley 715 de 2001, en su artículo 5 y Numeral 5.6, se enfoca en la definición, diseño y establecimiento de instrumentos y mecanismos para asegurar la calidad de la educación (MEN, 2001).

A raíz de esta base legal, las instituciones educativas en todo el país están constantemente inmersas en procesos de innovación institucional, que van desde mejoras en la infraestructura hasta la adaptación a la normativa vigente y la búsqueda de mejores resultados en pruebas externas como Las Pruebas Saber 11°. Este último es un indicador de evaluación significativo para las escuelas, ya que el Ministerio de Educación Nacional (MEN), a través del Instituto Colombiano para la Evaluación de la Educación (ICFES), clasifica a las instituciones educativas en cinco categorías: A+, A, B, C y D. La categoría A+ se reserva para las escuelas de mejor desempeño, mientras que la categoría D se asigna a aquellas con un desempeño más bajo (ICFES, 2016). Por lo tanto, los colegios se esfuerzan continuamente por fortalecer y mejorar sus planes académicos para cumplir con los estándares de evaluación del ICFES.

El objetivo de este trabajo es utilizar la herramienta estadística conocida como Análisis de Componentes Principales (ACP) para evaluar e identificar el rendimiento del nivel académico de una institución educativa a partir de los resultados de las Pruebas Saber 11°. Con esta herramienta, se busca identificar áreas de mejora, ya que el Análisis de Componentes Principales (ACP) puede ayudar a detectar las áreas de conocimiento en las que los estudiantes muestran un bajo rendimiento. Al comprender mejor estas áreas, los educadores pueden ajustar sus planes de enseñanza para abordar las deficiencias específicas. Además, se pueden identificar factores influyentes, ya que el ACP puede revelar las variables que más contribuyen a la variabilidad en los resultados académicos, como el tipo de enseñanza utilizada o el acceso a recursos educativos. Esto permitirá el seguimiento del progreso. Al utilizar el ACP en diferentes momentos a lo largo del tiempo, los educadores pueden realizar un seguimiento del progreso académico de los estudiantes y evaluar la efectividad de las intervenciones educativas implementadas. Finalmente, la

personalización del aprendizaje es clave. Al comprender las diferencias individuales en los patrones de rendimiento de los estudiantes, los educadores pueden adaptar sus métodos de enseñanza para satisfacer las necesidades específicas de cada estudiante, lo que podría mejorar su compromiso y desempeño académico.

Es evidente que el Análisis de Componentes Principales (ACP) puede ser una herramienta útil para comprender mejor los factores que influyen en el rendimiento académico de los estudiantes y diseñar estrategias pedagógicas más efectivas para mejorar los resultados en las pruebas Saber. Sin embargo, es importante tener en cuenta que el éxito académico es multifacético y depende de diversos factores. Por lo tanto, el ACP debe utilizarse en conjunto con otras herramientas y enfoques pedagógicos.

El análisis presentado se enfoca en los promedios de notas generados por los estudiantes de sexto a undécimo grado, analizando su desempeño a lo largo del año en las diferentes materias cursadas. Se identifican las deficiencias y se utiliza la herramienta propuesta para abordarlas de manera efectiva. Se seleccionarán las seis primeras componentes principales, que explicarán la mayor variabilidad posible de los datos en cada grado.

1.1. Estado del Arte

En esta propuesta, se analizarán los resultados del rendimiento académico de una institución educativa en Medellín, enfocándose en la aplicación de métodos estadísticos a los resultados académicos y a las pruebas Saber 11°. A continuación, se presenta una revisión de algunos estudios que abordan temas similares al objetivo de este trabajo.

En la Universidad Nacional de Misiones en Argentina, se llevó a cabo un análisis del rendimiento académico de los estudiantes de la Facultad de Ingeniería, abarcando las cohortes desde 1999 hasta 2003, con un total de 589 estudiantes como población objetivo. El rendimiento académico se definió como el promedio de materias aprobadas anualmente. A través de la técnica estadística multivariada de Regresión Logística, se determinó la incidencia de diversos factores, tanto personales, socioeconómicos y académicos, en el rendimiento académico de los estudiantes. Los resultados obtenidos llevaron a la conclusión de que las variables significativas para el rendimiento académico incluyen el promedio de calificaciones del nivel medio, el tipo de institución donde cursaron los estudios de nivel medio y el número de asignaturas aprobadas en el primer año de carrera. Destacaron que este último factor fue el más relevante, subrayando la importancia de la etapa inicial de la carrera en los resultados académicos posteriores del estudiante (Ibarra and Michalus, 2010).

En la Universidad Cooperativa de Colombia en compañía con la Universidad de la Costa y la Corporación Universitaria Rafael Núñez propusieron evaluar modelos predictivos basados en la metodología CRISP-DM para el desempeño de estudiantes en Colombia en la Prueba Saber 11°. El objetivo es determinar, a través de métricas específicas, el modelo más efectivo para predecir el rendimiento de los estudiantes de secundaria en la región Caribe Colombiana en la prueba Saber 11°. Además, se propone una nueva metodología de evaluación de los resultados de la prueba por regiones, con el fin de tener en cuenta las particularidades socioeconómicas de cada una de ellas (Acosta Solano et al., 2022).

Los profesores del departamento de pedagogía experimental de la Universidad de Barcelona llevaron a cabo un análisis múltiple para evaluar las técnicas clásicas incluidas

en los métodos multivariados, como el análisis de regresión, análisis de perfiles y análisis causal. En su enfoque, realizaron una síntesis de los modelos adoptados en la investigación sobre el rendimiento académico para destacar los diversos factores considerados y el nivel de énfasis puesto en ellos. También abordaron las implicaciones metodológicas derivadas de cada uno de estos modelos. En segundo lugar, propusieron, a partir de la explicación de un modelo parcial, una secuencia de análisis múltiple con el propósito de precisar el alcance y las limitaciones de algunas de las técnicas clásicas incluidas en los métodos multivariados. El análisis de regresión, el análisis de perfiles y el análisis causal fueron los aspectos centrales de su atención. Al plantear este trabajo, su objetivo fue llamar la atención sobre la necesidad de tomar las debidas precauciones metodológicas en la investigación sobre el rendimiento académico. Buscaron evitar la mezcla entre el nivel de opinión y el nivel de constatación que la investigación educativa puede ofrecer en este momento (Garanto et al., 2002).

Profesionales de diversas universidades, incluyendo la Universidad Nacional de las Lomas en Argentina, la Corporación Universitaria Minuto de Dios y la Universidad Simón Bolívar, llevaron a cabo una publicación en la Revista Espacios donde abordaron la identificación de variables influyentes en el cumplimiento de estándares curriculares emitidos por el Ministerio de Educación Nacional (MEN) de Colombia, específicamente en instituciones educativas en el distrito de Barranquilla. Para este propósito, utilizaron los resultados de varias instituciones educativas en el distrito de Barranquilla en la prueba estatal Saber 11 del ICFES, realizada en el año 2012. En el análisis estadístico, aplicaron la técnica del Análisis de Componentes Principales (ACP), una herramienta del análisis multivariado. Este enfoque les permitió identificar las variables clave que afectan el cumplimiento de los estándares curriculares establecidos por el MEN, proporcionando así una perspectiva valiosa para la mejora continua en el ámbito educativo (Ruiz Escorcia et al., 2016).

En la Fundación Universitaria Los Libertadores, se llevó a cabo una investigación con el objetivo de definir las principales variables y su incidencia en el puntaje de las pruebas SABER PRO dentro del programa de Ingeniería en Producción Acuícola de la Universidad de Nariño durante los últimos cuatro años (2016-2019). Para este proyecto, se utilizaron 384 datos correspondientes a estudiantes de noveno semestre que presentaron las pruebas SABER PRO en el programa de pregrado. Se implementó un análisis descriptivo de los datos y técnicas multivariadas por componentes principales, así como un análisis de regresión múltiple y cuantílica para establecer las principales variables y su incidencia en el puntaje global. Los resultados del primer análisis determinaron que todas las competencias genéricas inciden en las dos primeras componentes, pero se identificó la necesidad de fortalecer variables como Lectura Crítica, Comunicación Escrita y Competencia Ciudadana, las cuales tienen una contribución mayor al 50 % sobre las tres últimas componentes, impidiendo que la variabilidad retenida en las primeras componentes sea mayor al 60 %. Adicionalmente, a pesar del incumplimiento del supuesto de homocedasticidad, el análisis de regresión múltiple generó un modelo con el 99 % de capacidad predictiva. El análisis de regresión cuantílica generó cinco modelos correspondientes a los cuantiles de 0.1, 0.25, 0.5 (mediana), 0.75 y 0.9, sin diferencias significativas entre ellos. Todos los modelos, sin excepción, demandan la participación de todas las competencias genéricas con aproximadamente el mismo coeficiente para cada una de ellas. Como conclusión, se determinó que todas las variables son indispensables para predecir el puntaje global, cada una con diferente grado de contribución, calidad de asociación y peso sobre las combinaciones

lineales de las dos primeras componentes. Además, el modelo generado por las diferentes técnicas de regresión fue el mismo, siendo cualquiera capaz de predecir el puntaje global de las pruebas SABER PRO para el programa de Ingeniería en Producción Acuícola (Cerón Benavides et al., 2020).

En la Universidad de Santander llevaron a cabo un análisis enfocado en los resultados de las competencias genéricas de las pruebas de Estado Saber 11° por parte de los estudiantes recién ingresados para el periodo A-2016. El objetivo principal de este análisis de los pre saberes de los estudiantes admitidos fue establecer un criterio de clasificación para identificar a aquellos con altos niveles de riesgo, según los puntajes obtenidos en cada área evaluada. Además de considerar los puntajes en las diversas áreas de la prueba, se incorporaron otras variables como el tipo de colegio, género, carrera a la que ingresan y la región de procedencia. Para caracterizar a los estudiantes, se aplicó un análisis de correspondencias múltiples. Posteriormente, se propuso un análisis factorial que identificó tres factores: numérico, verbal e inglés. Finalmente, mediante regresión cuantílica, se intentó explicar la incidencia de las covariables sobre los factores a distintos niveles. Los resultados indicaron que los estudiantes provenientes de colegios privados y de la región de Santander demostraron un mejor rendimiento en las competencias de inglés y matemáticas respectivamente (Pérez Pulido et al., 2016).

En la Universidad de La Salle sede Bogotá se llevó a cabo un trabajo en torno a la información recopilada por el Instituto Colombiano para la Evaluación de la Educación (ICFES) de las Pruebas Saber 11° para los años 2017 y 2018. Esta información incluyó variables relacionadas con los datos de contacto de los estudiantes, información socioeconómica, datos del colegio, detalles de la citación del examen y los resultados obtenidos. El objetivo del estudio fue determinar si alguna de estas variables tenía incidencia en los resultados de las Pruebas Saber 11° y en la brecha existente entre colegios públicos. Se analizaron las variables utilizando el software R Studio y, a través de un modelo de Random Forest, se evaluó la importancia de algunas de estas variables sobre el puntaje promedio obtenido por los estudiantes en estas pruebas, Con este trabajo, se identificaron las variables más influyentes en el puntaje, resaltando que el acceso a internet, a un computador y el número de libros en el hogar fueron las tres más importantes. Además, se confirmó que las diferencias en estas variables contribuyen a la brecha existente entre las instituciones educativas públicas. Como conclusión, se propusieron tres iniciativas basadas en los resultados obtenidos, con el propósito de mejorar la calidad de la educación y los resultados de los estudiantes en las Pruebas Saber 11° (Rodríguez Rodríguez and Hernandez Cano, 2021).

En la Universidad Nacional de Colombia, sede Medellín, se emplearon diversas técnicas estadísticas multivariadas, como Máquinas Vector Soporte (SVM), Análisis Discriminante (AD), K-vecinos más próximos (KNN) y Regresión Logística (RL), con el objetivo de clasificar a los estudiantes de pregrado en dos poblaciones: aquellos con posibilidad de desertar y aquellos sin esa posibilidad. El estudio se basó en la información proporcionada por los estudiantes que ingresaron a la universidad desde el primer semestre de 2009 hasta el primer semestre de 2016. Se utilizaron datos como el rendimiento académico de cada estudiante en cada periodo matriculado y la identificación de aquellos que perdieron la calidad de estudiante debido a bajo rendimiento, así como aquellos que continuaron con sus estudios. Esta información se dividió para el entrenamiento de los modelos y la validación. Los resultados obtenidos permitieron identificar la técnica que proporciona el modelo con el menor porcentaje de error y mayor sensibilidad, lo que podría ser útil para

realizar predicciones de deserción en nuevos individuos basándose en la información de las variables seleccionadas (Madrid Echeverry, 2017).

La revista *Multivariate Behavioral Research* publicó la aplicación del modelo de regresión multinivel contemporáneo a la predicción del desempeño académico de estudiantes universitarios de primer año. En un primer nivel de análisis, los datos provienen de $N > 16.000$ estudiantes que eran estudiantes de primer año de universidad en 1994-1995 y que también participaban en deportes universitarios de alto nivel. En un segundo nivel de análisis, los datos de los estudiantes se relacionaron con las diferentes características de las universidades C D 267 de la División I de la NCAA (National College Athletics Association). Los análisis presentados aquí se centran inicialmente en la predicción del GPA de los estudiantes de primer año a partir de una variedad de variables académicas de la escuela secundaria. Los modelos utilizados son modelos de regresión multinivel estándar, pero examinamos la predicción no lineal dentro de estos modelos multinivel y se consideran variables de resultado adicionales (McArdle et al., 2013).

En la Universidad Tecnológica Nacional de Argentina, se destaca que la estadística es una disciplina universalmente aplicada para describir, clasificar, inferir, decidir y cuantificar la probabilidad de error en la toma de decisiones científicas. Aunque su aplicación en estudios cuantitativos y cualitativos ya está consolidada, su utilidad en áreas sociales, especialmente en investigación educativa, ha experimentado un impulso reciente debido a la difusión del software de código abierto y al crecimiento exponencial de la ciencia de datos. Con el objetivo de estudiar la contribución diferencial de las técnicas estadísticas multivariadas y del lenguaje R, se seleccionaron trabajos de investigación educativa del nivel superior en un período reciente. Se analizaron las técnicas multivariadas empleadas en estas investigaciones y se compararon con trabajos que abordaban contextos muy similares pero no aplicaban estas técnicas. Este análisis permitió identificar la evolución de la aplicación de estas técnicas y del lenguaje R en las investigaciones de corte educativo (Chan and Galli, 2020).

La Universidad Técnica Estatal de Novosibirsk, Federación de Rusia, señala que las universidades modernas, al desarrollar un entorno digital, se basan en un enfoque centrado en el estudiante, relevante para las prácticas pedagógicas y didácticas. El artículo que han trabajado presenta el uso de modelos multivariados para identificar relaciones y patrones comunes entre el rendimiento académico y las características psicométricas. La muestra del estudio incluye datos de pruebas psicométricas de estudiantes de la Universidad Técnica Estatal de Novosibirsk (NSTU), que determinan las características cognitivas, motivacionales y personales, así como datos sobre su rendimiento académico. Los resultados obtenidos se utilizan además con fines de diagnóstico y pronóstico para recomendar contenidos educativos en función de la personalidad del estudiante (Pustovalova and Avdeenko, 2022).

La Federal University Wukari (FUW) en Wukari, Taraba, Nigeria, llevó a cabo un estudio de investigación diseñado para comparar el rendimiento académico de 300 estudiantes de nivel de la Universidad Federal Wukari durante la sesión académica del segundo semestre de 2015/2016. Los grupos considerados fueron las Facultades de Agricultura y Ciencias de la Vida, Ciencias Puras y Aplicadas, y Humanidades, Gestión y Ciencias Sociales. Se seleccionó una muestra aleatoria de cinco (5) departamentos de cada facultad. Los datos recopilados incluyeron el CGPA de 300 estudiantes de nivel en cinco (5) departamentos de cada facultad, accesados desde los expedientes académicos de la Universidad. Para el análisis, se utilizó el Análisis Multivariado de Varianza (MANOVA) y

estadística descriptiva. Los hallazgos revelaron que no existe una diferencia significativa en el rendimiento académico de los estudiantes de las tres (3) Facultades (Okeke et al., 2018).

Profesionales de diversas universidades, incluyendo la Universidad Nacional Agraria La Molina, The University of San Martín de Porres y la Universidad de Lima, llevaron a cabo una investigación que presenta diferentes técnicas multivariadas aplicadas a las calificaciones obtenidas por los alumnos en las diferentes asignaturas de un plan de estudios, así como algunas variables socioeconómicas, para detectar posibles relaciones y comportamientos irregulares de alumnos, secciones y profesores en una institución de educación superior. El objetivo es alcanzar una gestión educativa eficiente, oportuna y confiable. Las técnicas multivariadas utilizadas incluyen análisis de correspondencias simple, análisis clúster, análisis de covarianza, análisis discriminante, regresión logística binaria y regresión logística ordinal y nominal. Los resultados de la investigación indican que la técnica del análisis de covarianza permite comparar las diferentes secciones de un curso, eliminando el efecto del promedio ponderado acumulativo o de cualquier otra covariable. Asimismo, las otras técnicas multivariadas contribuyen a alcanzar el objetivo propuesto (Chue Gallardo et al., 2007).

El estudio realizado en la Facultad de Estadística de la Universidad Santo Tomás destaca la importancia de considerar los efectos marginales de la educación en función de los diferentes niveles de capacidades de los estudiantes en el ámbito académico. La omisión de este comportamiento, junto con la interacción de la educación con las capacidades de los estudiantes, podría resultar en interpretaciones erróneas si se utilizan métodos lineales para la estimación de los parámetros de interés. Se abordó el uso de modelos de regresión semiparamétrica, cuyo objetivo principal es analizar la relación entre la variable respuesta y las variables explicativas. Estos modelos permiten establecer esta relación mediante una función de distribución conocida, facilitando así la estimación de los parámetros y la interpretación de los resultados. Un punto clave resalta la importancia de abordar la estructura desconocida de los errores en el modelo, ya que esta puede afectar la estimación de los parámetros. Los errores pueden contener información significativa que no se ajusta adecuadamente al modelo. Por lo tanto, el estudio se centró en entender y abordar los diferentes tipos de errores que podrían surgir al ajustar un modelo semiparamétrico. El objetivo final era lograr la mejor estimación insesgada del modelo y minimizar la pérdida de información en los errores (Corredor Rivera and Pineda Ríos, 2017).

En la Universidad Tecnológica de Pereira (UTP), se llevaron a cabo análisis multivariados, como la regresión logística y el análisis discriminante, para identificar las variables significativas que afectan el rendimiento académico en la asignatura de Álgebra Lineal. Este estudio incluyó a 131 estudiantes de las carreras de Ingeniería Ambiental y Biomédica en la Universidad Manuela Beltrán. En la construcción de los modelos, se tuvieron en cuenta diez variables predictoras. Se ejecutaron 11 modelos utilizando la técnica de regresión logística en el programa SPSS, de los cuales solo seis mostraron incidencia significativa en el rendimiento académico de la asignatura. Se destacó que el rendimiento académico en matemáticas durante el bachillerato fue la variable que contribuyó con el mayor porcentaje de clasificación. Además, se empleó la técnica de análisis discriminante para evaluar si el porcentaje de clasificación de los tres modelos más significativos respecto a la variable dependiente mejoraba con este método (Marquín Triana and Romero, 2017).

Un trabajo realizado en la Universidad de Granada se enfocó en utilizar la metodología

de Componentes Principales para calcular el índice de empoderamiento de mujeres colombianas que participaron en el programa de gobierno “Mujeres Ahorradoras en Acción” en el año 2012. Este programa tenía como objetivo contribuir a la superación de la pobreza extrema de mujeres en situación de vulnerabilidad, brindándoles acceso al sistema microfinanciero y fomentando la generación de ingresos a través del fortalecimiento microempresarial. Para llevar a cabo este estudio, se recopiló información mediante encuestas a las mujeres inscritas en el programa, analizando su situación económica y dedicando un capítulo completo a preguntas que permitieran comprender el proceso de empoderamiento. El análisis de esta información se basó en un estudio previo realizado en México por el Instituto Nacional de las Mujeres, donde se utilizó el análisis de componentes principales para generar índices de empoderamiento. Estos índices se aplicaron a la base de datos colombiana, estimando así los niveles de empoderamiento de las mujeres participantes en el programa. El objetivo final del trabajo fue proporcionar al programa estos cálculos para su utilización en futuros estudios (Peña Méndez and Gutiérrez Sánchez, 2014).

El artículo propuesto por docentes de la Corporación Universitaria del Caribe (CE-CAR) aborda la importancia de evaluar los procesos educativos mediante el seguimiento de la apropiación de competencias básicas en la comunidad estudiantil para lograr el éxito. El objetivo central del artículo fue realizar un análisis multivariado de los resultados obtenidos por los estudiantes en las diferentes competencias evaluadas en la prueba Saber-Pro 2016 en el Departamento de Sucre, Colombia. Desde el punto de vista metodológico, se llevó a cabo un análisis de clúster y un análisis de correspondencias simples para identificar las asociaciones más significativas entre los niveles y las competencias genéricas, utilizando la categorización en quintiles establecida por el ICFES. Las técnicas multivariadas empleadas para analizar los resultados de las competencias genéricas de las pruebas Saber-Pro fueron consideradas complementarias. Se concluyó que el grupo de referencia de Medicina mostró un nivel alto, mientras que los grupos de referencia de Ingeniería, Ciencias Naturales y Exactas tuvieron un nivel medio. En contraste, se observaron niveles bajos y muy bajos para los demás grupos de referencia. Este análisis proporciona una perspectiva detallada de los niveles de competencia en diferentes áreas y ofrece información valiosa para la mejora de los procesos educativos en el Departamento de Sucre (Rodríguez Manrique et al., 2022).

Capítulo 2

Objetivos

2.1. Objetivo General

- A partir del uso de Técnicas de Estadística Multivariada (TEM), construir algunos instrumentos estadísticos que nos permitan comparar los resultados de las pruebas saber 11° con los resultados académicos.

2.2. Objetivos Específicos

- Analizar por medio de TEM si los procesos académicos internos de un colegio potencian o no los resultados de Las Pruebas Saber 11°.
- Realizar un comparativo de los años 2019 a 2021 de los resultados académicos y de Las Pruebas Saber 11° usando TEM.
- Generar un insumo a partir de TEM que permita evaluar la institución año a año, según sus proyecciones académicas institucionales.

Capítulo 3

Marco Teórico

Las técnicas estadísticas multivariadas (TEM) han experimentado un constante aumento en sus aplicaciones a lo largo de los años. Ahora, resulta difícil abordar la amplia gama de aplicaciones del mundo real de estos métodos en discusiones breves, dado que existen diversas técnicas estadísticas que permiten analizar conjuntos de datos multivariados (Jhonson and Wichern, 2007). Por lo tanto, el objeto y la materia prima del trabajo estadístico se encuentran en los datos que proporcionan información sobre un objeto en un momento determinado (Monroy, 2007).

Las técnicas estadísticas multivariadas se centran en datos asociados a conjuntos de medidas sobre un grupo de individuos u objetos. Estos individuos, junto con sus variables, pueden organizarse en una estructura de forma matricial, por ejemplo, $\mathbf{X}_{n \times p}$, donde las filas corresponden a las observaciones individuales y las columnas a las diferentes variables medidas en esas observaciones. Las TEM se distinguen en función del análisis por filas (individuos) y/o columnas (variables) (Monroy, 2007).

En la amplia variedad de tipos de datos que se pueden analizar mediante TEM, se incluyen los resultados académicos de una institución basados en evaluaciones realizadas en distintas asignaturas y áreas cursadas. La evaluación es un concepto fundamental para la calidad de los procesos educativos, ya que proporciona información valiosa que puede respaldar decisiones estructuradas y facilitar la comprensión de los procesos de enseñanza-aprendizaje a través de un análisis adecuado de los datos mediante alguna TEM.

El uso y la interpretación apropiados de los resultados del análisis de datos mediante TEM pueden orientar de manera efectiva el trabajo de las instituciones académicas, docentes, estudiantes y padres de familia (MEN, 2008). En consecuencia, el desempeño académico es esencial al evaluar los procesos educativos en las instituciones, ya que proporciona información que respalda la toma de decisiones estructuradas y permite comprender los procesos de enseñanza-aprendizaje que, de lo contrario, podrían no estar claros sin un análisis adecuado de los datos.

Algunas de las herramientas o TEM que podrían utilizarse para este tipo de análisis de datos multivariados incluyen el análisis de componentes principales (ACP). El ACP permite reducir la dimensionalidad de los espacios en los que residen tanto los individuos como las variables, lo que simplifica la información al eliminar datos redundantes y facilita la interpretación de los resultados. Otra TEM a considerar es el Análisis de Correspondencias, tanto simple como múltiple, que se utiliza con objetivos similares al ACP,

pero se enfoca en conjuntos de datos con variables cualitativas. Mediante un proceso de transformación, estos datos se adecúan para realizar análisis similares al ACP utilizando las distintas categorías de las variables cualitativas como las variables de insumos del ACP.

3.1. Introducción al Análisis de Componentes Principales (ACP)

El análisis de componentes principales está relacionado con la explicación de la estructura de covarianzas de un conjunto de p variables a través de pequeñas combinaciones lineales de estas variables. Estas combinaciones lineales representan la selección de un nuevo sistema de ejes de coordenadas obtenido mediante la rotación del sistema original. Los nuevos ejes representan las direcciones con máxima variabilidad y proporcionan una forma más simple de describir las estructuras de covarianza (Monroy, 2007).

Con las componentes principales se busca principalmente reducir la dimensionalidad e intentar interpretar dichas combinaciones. Aunque p componentes son necesarias para reproducir la variabilidad total del sistema, a veces mucha de esta variabilidad podría ser explicada por un número pequeño k de componentes principales. Es decir, existe casi tanta información en las k componentes principales como la que se obtiene de las p variables originales (Jhonson and Wichern, 2007).

De esta manera, el conjunto de n mediciones en p variables puede ser reducido a un conjunto de n mediciones y k componentes principales, donde $k < p$. Este tipo de análisis puede revelar comportamientos en los datos que previamente no se sospechaban y permitir algunas interpretaciones que no eran evidentes con los datos originales.

Algebraicamente, las componentes principales son combinaciones lineales particulares de p variables aleatorias X_1, X_2, \dots, X_p . Geométricamente, representan un nuevo sistema de coordenadas, el cual se obtiene al rotar el sistema original generado por X_1, X_2, \dots, X_p .

Las componentes principales dependen solamente de la matriz de covarianzas Σ (o de la matriz de correlaciones ρ) asociada al vector de variables $\underline{\mathbf{X}} = (X_1, \dots, X_p)'$. Su desarrollo no requiere del supuesto de normalidad multivariada. Es claro que *la matriz de covarianza* es una medida de la relación entre las variables en un conjunto de datos.

Cuando la distribución del vector de variables $\underline{\mathbf{X}}$ es normal multivariada, las componentes principales adquieren interpretaciones útiles relacionadas con densidades elipsoidales constantes. Además, el nuevo vector que contiene las p componentes principales tendrá una distribución normal multivariada. Este supuesto de normalidad multivariada en el vector $\underline{\mathbf{X}}$ facilita la inferencia sobre las componentes principales muestrales.

Para un vector de variables $\underline{\mathbf{X}} = (X_1, \dots, X_p)'$, considere las siguientes combinaciones lineales:

$$\begin{aligned}
 Y_1 &= \underline{\mathbf{a}}'_1 \underline{\mathbf{X}} = a_{11}X_1 + \dots + a_{1p}X_p \\
 Y_2 &= \underline{\mathbf{a}}'_2 \underline{\mathbf{X}} = a_{21}X_1 + \dots + a_{2p}X_p \\
 &\vdots \\
 Y_p &= \underline{\mathbf{a}}'_p \underline{\mathbf{X}} = a_{p1}X_1 + \dots + a_{pp}X_p.
 \end{aligned}
 \tag{3.1}$$

Es fácil ver que

$$\begin{aligned}
 \text{Var}[Y_i] &= \underline{\mathbf{a}}_i^T \Sigma \underline{\mathbf{a}}_i, \text{ donde } a_i \text{ representa a los coeficientes asociados a la } i\text{-ésima CPs} \\
 &\vdots \\
 \text{Cov}[Y_i, Y_k] &= \underline{\mathbf{a}}_i^T \Sigma \underline{\mathbf{a}}_k, \text{ para } i, k = 1, \dots, p \text{ y } i \neq k.
 \end{aligned}
 \tag{3.2}$$

Las componentes principales son aquellas combinaciones lineales Y_1, Y_2, \dots, Y_p no-correlacionadas, tales que sus varianzas sean tan grandes como sea posible (Jhonson and Wichern, 2007).

3.2. Proceso para hallar las Componentes Principales

El proceso para identificar las componentes principales se lleva a cabo de la siguiente manera:

- La primera componente principal corresponde a la combinación lineal $\underline{\mathbf{a}}'_1 \underline{\mathbf{X}}$ que maximiza $\text{Var}[\underline{\mathbf{a}}'_1 \underline{\mathbf{X}}]$, sujeto a que $\underline{\mathbf{a}}'_1 \underline{\mathbf{a}}_1 = 1$.
- La segunda componente principal corresponde a la combinación lineal que maximiza $\text{Var}[\underline{\mathbf{a}}'_2 \underline{\mathbf{X}}]$, sujeto a que $\underline{\mathbf{a}}'_2 \underline{\mathbf{a}}_2 = 1$ y $\text{Cov}[\underline{\mathbf{a}}'_1 \underline{\mathbf{X}}, \underline{\mathbf{a}}'_2 \underline{\mathbf{X}}] = 0$.
- La i -ésima componente principal corresponde a la combinación lineal $\underline{\mathbf{a}}'_i \underline{\mathbf{X}}$ que maximiza $\underline{\mathbf{a}}'_i \underline{\mathbf{a}}_i = 1$ y $\text{Cov}[\underline{\mathbf{a}}'_i \underline{\mathbf{X}}, \underline{\mathbf{a}}'_k \underline{\mathbf{X}}] = 0$ para $k = 1, 2, \dots, i - 1$.

3.3. Fundamentación Algebraica del ACP

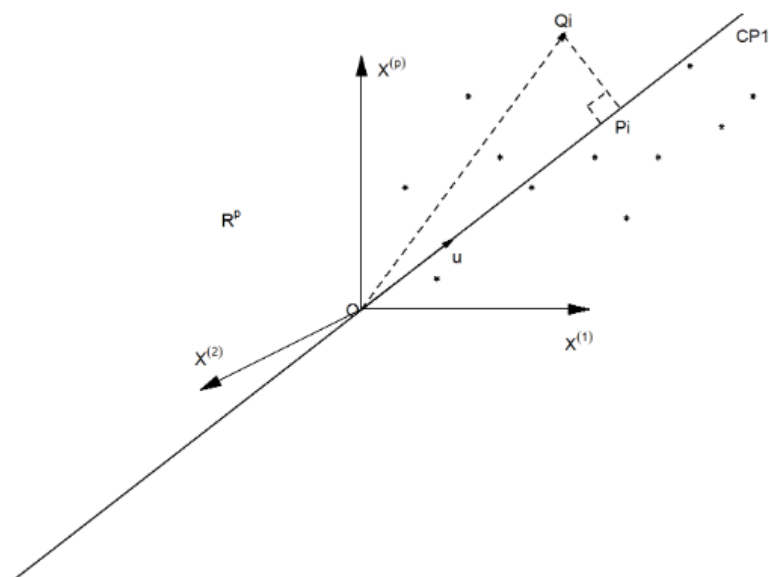


Figura 3.1: Ajuste de Mínimos Cuadrados del ACP (Fuente: Elaboración Propia)

Algebraicamente, las CPs son combinaciones lineales (CLs) particulares de las p -variables aleatorias originales X_1, X_2, \dots, X_p .

Geoméricamente, las componentes principales representan la selección de un nuevo sistema de coordenadas (o ejes) que se obtienen al rotar el sistema de coordenadas original generado por las variables originales X_1, X_2, \dots, X_p , hasta obtener las direcciones ortogonales (o componentes principales) con máxima variabilidad y cuyas direcciones proporcionan una dispersión simple y más parsimoniosa de la estructura de varianzas y covarianzas de los datos (Jhonson and Wichern, 2007).

Las CPs dependen solamente de la matriz de varianzas-covarianzas Σ (o de la matriz de correlación ρ) del vector $\mathbf{X} = (X_1, X_2, \dots, X_p)'$.

Para encontrar un sub-espacio de dimensión q ($q < p$), tal que los n -puntos (ie. las filas de \mathbf{X}) queden aproximadamente representados de la mejor forma en este sub-espacio, se empieza por encontrar un sub-espacio de dimensión uno (ie. $q = 1$), es decir, encontrar una línea recta que contenga al origen, la cual se ajuste lo mejor posible a los n -puntos o datos.

La figura 3.1, muestra el ajuste de los datos, a la línea recta $CP1$. La proyección de un vector cualesquiera OQ_i (ie. del i -ésimo individuo) sobre la recta $CP1$, es el vector OP_i .

Sea \mathbf{u} - el vector unitario del sub-espacio $CP1$, entonces la proyección OP_i es el producto escalar entre OQ_i y \mathbf{u} . De esta forma, el producto de la matriz \mathbf{X} y el vector \mathbf{u} , ie. $\mathbf{X}\mathbf{u}$, es la proyección de las filas de \mathbf{X} sobre el sub-espacio de dimensión uno $CP1$.

3.3.1. Ajuste de Mínimos Cuadrados

Uno de los criterios para encontrar el “mejor” ajuste o mejor sub-espacio es el de mínimos cuadrados.

De la relación pitagórica entre los lados del triángulo OP_iQ_i , de la figura 3.1 se tiene que, al sumar sobre cada uno de los n -triángulos determinados por las proyecciones de los n -puntos sobre el sub-espacio $CP1$, se tiene que:

$$(OQ_i)^2 = (OP_i)^2 + (Q_iP_i)^2, i = 1, 2, \dots, n$$

de donde,

$$(Q_iP_i)^2 = (OQ_i)^2 - (OP_i)^2, \quad i = 1, 2, \dots, n$$

$$Q = \sum_{i=1}^n (Q_iP_i)^2 = \sum_{i=1}^n (OQ_i)^2 - \sum_{i=1}^n (OP_i)^2.$$

Las proyecciones OP_i -reflejan la información recogida en el sub-espacio $CP1$ de cada punto. Se quiere maximizar esta cantidad de información, es decir maximizar las OP_i (lo que equivale a minimizar a Q).

Ahora, como los puntos Q_i -están a una cantidad fija del origen O_i , entonces se tiene que minimizar a Q , es equivalente a maximizar las distancias OP_i . De lo anterior, se tiene que la cantidad a maximizar, en función de la matriz de datos \mathbf{X} , es:

$$\text{Maximizar : } \sum_{i=1}^n (OP_i)^2 = (\mathbf{X}\mathbf{u})'(\mathbf{X}\mathbf{u}) = \mathbf{u}'\mathbf{X}'\mathbf{X}\mathbf{u}$$

con la restricción: $\mathbf{u}'\mathbf{u} = 1$.

La solución al anterior problema de maximización, es el mejor sub-espacio unidimensional.

Ahora, el mejor sub-espacio bi-dimensional, que ajusta los n -puntos es el generado por \mathbf{u}_1 , \mathbf{u}_2 , donde \mathbf{u}_2 -es el segundo vector en la base para este sub-espacio bi-dimensional, el cual es ortogonal a \mathbf{u}_1 y que maximiza a:

$$\mathbf{u}_2'\mathbf{X}'\mathbf{X}\mathbf{u}_2.$$

De forma iterativa, el sub-espacio q -dimensional ($q \leq p$), es el “mejor” en el sentido mínimo-cuadrático, y se determina de manera semejante. Se demuestra que este sub-espacio es generado por los vectores propios:

$$\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_q$$

de la matriz $\mathbf{X}'\mathbf{X}$, los cuales corresponden a los q -valores propios más grandes.

3.3.2. ¿Por qué el uso de la palabra “mejor”?

Se trata de maximizar a : $\mathbf{u}'\mathbf{X}'\mathbf{X}\mathbf{u}$

con la restricción de $\mathbf{u}'\mathbf{u} = 1$.

Lo anterior es un problema de optimización con restricción, el cual se resuelve a través de técnica de multiplicadores de Lagrange. La expresión a maximizar o función langrangiana es:

$$L(\mathbf{u}, \lambda) = \mathbf{u}'\mathbf{X}'\mathbf{X}\mathbf{u} - \lambda(\mathbf{u}'\mathbf{u} - 1)$$

Al derivar con respecto a $\underline{\mathbf{u}}$ e igualar a cero se tiene que:

$$2\mathbf{X}'\mathbf{X}\underline{\mathbf{u}} - 2\lambda\underline{\mathbf{u}} = \underline{\mathbf{0}}$$

o equivalentemente,

$$\mathbf{X}'\mathbf{X}\underline{\mathbf{u}} = \lambda\underline{\mathbf{u}},$$

de donde se observa que $\underline{\mathbf{u}}$ es un vector propio de $\mathbf{X}'\mathbf{X}$. Ahora, se verá cuánto es el máximo alcanzado en $\underline{\mathbf{u}}$ Como:

$$\underline{\mathbf{u}}'\mathbf{X}'\mathbf{X}\underline{\mathbf{u}} = \underline{\mathbf{u}}'(\mathbf{X}'\mathbf{X}\underline{\mathbf{u}}) = \underline{\mathbf{u}}'\lambda\underline{\mathbf{u}} = \lambda\underline{\mathbf{u}}'\underline{\mathbf{u}} = \lambda$$

entonces, se tiene que el valor máximo se consigue cuando λ -es el valor propio más grande de $\mathbf{X}'\mathbf{X}$.

Luego, se denota por $\underline{\mathbf{u}}_1$ -al vector propio asociado con el valor propio más grande λ_1 de la matriz $\mathbf{X}'\mathbf{X}$.

De lo anterior se concluye que $\underline{\mathbf{u}}_1$ genera el sub-espacio $CP1$, llamado la Primera Componente Principal.

Similarmente, si se busca el sub-espacio de dos dimensiones con características similares de ajuste a $CP1$ y que contenga a $CP1$, entonces se debe buscar un segundo vector unitario $\underline{\mathbf{u}}_2$ que maximice a:

$$\underline{\mathbf{u}}_2'\mathbf{X}'\mathbf{X}\underline{\mathbf{u}}_2$$

con la restricción de que sea ortogonal a $\underline{\mathbf{u}}_1$. En este caso se tiene la siguiente ecuación Lagrangiana:

$$L(\underline{\mathbf{u}}_1, \underline{\mathbf{u}}_2, \lambda) = \underline{\mathbf{u}}_2'\mathbf{X}'\mathbf{X}\underline{\mathbf{u}}_2 - \lambda(\underline{\mathbf{u}}_2'\underline{\mathbf{u}}_2 - 1) - \psi\underline{\mathbf{u}}_2'\underline{\mathbf{u}}_1$$

con λ y ψ -los respectivos multiplicadores de Lagrange.

Derivando con respecto a $\underline{\mathbf{u}}_2$, igualando a cero y teniendo en cuenta las restricciones se llega a que, $\underline{\mathbf{u}}_2$ es el vector propio asociado con el segundo valor propio más grande λ_2 de la matriz $\mathbf{X}'\mathbf{X}$.

Capítulo 4

Análisis Descriptivo

4.1. Base de Datos

En este estudio, se evaluó una base de datos de un colegio en Medellín que contenía información académica de tres años, abarcando desde 2019 hasta 2021. Se recopiló información de los grados sexto a undécimo, con un total de 873 estudiantes en el año 2019, 841 estudiantes en el año 2020 y 872 estudiantes en el año 2021, lo que dio un total de 2586 observaciones analizadas durante esos tres años y en esos grados. Se inició con un análisis descriptivo de las calificaciones generadas por los estudiantes de sexto a undécimo grado. Posteriormente, se llevó a cabo un análisis de variabilidad de los datos para cada año escolar con las asignaturas evaluadas y, finalmente, se examinó el comportamiento de los resultados académicos año tras año mediante un gráfico.

4.2. Análisis de las notas de los diferentes Cursos

4.2.1. Análisis de notas del grado 6°

En la tabla 4.1 se observan las calificaciones generadas durante los años 2019 a 2021 del grado 6°, con un total de 369 observaciones. Este conjunto de datos proporciona una visión detallada del rendimiento académico de los estudiantes a lo largo de tres años consecutivos, permitiendo así un análisis comparativo de su evolución y progreso en distintas asignaturas. El análisis de estas calificaciones es crucial para identificar áreas de mejora en el proceso educativo y desarrollar estrategias pedagógicas que fomenten un aprendizaje más efectivo. Además, estos datos pueden servir como referencia para futuras investigaciones sobre la evolución del rendimiento académico en contextos similares.

Tabla 4.1: Calificaciones obtenidas por los estudiantes de grado 6° durante los años escolares 2019, 2020 y 2021 en una institución de Medellín.

	BIO: Biología	CSO: Sociales	DIB: Dibujo	ETI: Ética	EFI: Educación Física	REL: Religión	LCA: Lengua Castellana	ING: Inglés	MAT: Matemáticas	TEC: Tecnología	FIL: Filosofía	EMP: Emprendimiento
AÑO	BIO	CSO	DIB	ETI	EFI	REL	LCA	ING	MAT	TEC	FIL	EMP
2019	4,3	4,1	4,2	4,5	4,1	4,7	4,1	4,3	4,2	4,5	4,4	4,4
2019	3,8	3,5	3,5	4,0	3,9	3,5	3,5	3,6	3,6	3,7	3,8	4,0
2019	4,2	3,7	4,0	4,5	4,2	4,5	3,8	4,1	3,7	4,4	4,1	4,6
2019	3,8	3,6	3,5	3,5	4,0	3,8	3,5	3,5	3,5	4,0	3,7	3,6
2019	3,9	4,0	3,7	4,0	4,2	4,5	3,5	3,5	3,8	4,1	3,9	4,3
.
.
.

En el año 2019, los estudiantes de sexto grado mostraron un desempeño más uniforme y consistentemente alto en materias como emprendimiento (4.29), tecnología (4.22) y educación física (4.14), las cuales también presentaron una menor variabilidad en las notas, indicando consistencia en el rendimiento. Por otro lado, materias como lengua castellana (3.62), sociales (3.83) y ética (4.01) presentaron mayores dispersiones, reflejando una mayor variabilidad en el desempeño de los estudiantes. En particular, lengua castellana y sociales tuvieron los coeficientes de variación más altos (0.097 y 0.093, respectivamente), sugiriendo que, en relación con sus medias, estas materias exhiben una mayor variabilidad en las notas, lo que podría requerir una atención especial para mejorar la uniformidad en el rendimiento estudiantil ver tabla 4.2.

Tabla 4.2: Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 6° durante el año 2019 de una institución de Medellín.

Variable	BIO	CSO	DIB	ETI	EFI	REL	LCA	ING	MAT	TEC	FIL	EMP
Media	3.94	3.83	3.96	4.01	4.14	3.97	3.62	3.93	3.74	4.22	3.91	4.29
Varianza	0.083	0.126	0.084	0.113	0.015	0.092	0.123	0.107	0.093	0.051	0.039	0.068
Desv.Estándar	0.287	0.354	0.290	0.336	0.124	0.304	0.351	0.328	0.305	0.225	0.196	0.261
Coef.Variación	0.073	0.093	0.073	0.084	0.030	0.077	0.097	0.083	0.082	0.053	0.050	0.061

En la tabla 4.3, al analizar el año 2020, los estudiantes de sexto grado mostraron un desempeño más parejo y alto en materias como dibujo (4.32), tecnología (4.30) y emprendimiento (4.28), con notas más consistentes. Por otro lado, en materias como lengua castellana (3.86), religión (3.96) y sociales (4.03), las notas fueron más variadas, con diferencias mayores entre los estudiantes. En particular, inglés y lengua castellana tuvieron los coeficientes de variación más altos (0.081 y 0.076, respectivamente), lo que indica una mayor variabilidad en estas materias. Esto sugiere que sería útil enfocar esfuerzos en estas áreas para mejorar la uniformidad en el rendimiento estudiantil. Comparando con 2019, se observa una mejora general en las medias de las notas, aunque la dispersión sigue siendo un área a abordar en ciertas materias.

Tabla 4.3: Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 6° durante el año 2020 de una institución de Medellín.

Variable	BIO	CSO	DIB	ETI	EFI	REL	LCA	ING	MAT	TEC	FIL	EMP
Media	4.05	4.03	4.32	4.12	4.24	3.96	3.86	4.15	4.07	4.30	4.07	4.28
Varianza	0.050	0.084	0.034	0.063	0.059	0.047	0.086	0.114	0.081	0.052	0.030	0.059
Desv.Estándar	0.225	0.290	0.185	0.250	0.243	0.216	0.293	0.338	0.285	0.227	0.173	0.242
Coef.Variación	0.056	0.072	0.043	0.061	0.057	0.055	0.076	0.081	0.070	0.053	0.042	0.057

En la tabla 4.4 en el año 2021, los estudiantes de sexto grado tuvieron un desempeño más uniforme y alto en materias como emprendimiento (4.30), educación física (4.27) y religión (4.19), con notas más consistentes. Sin embargo, materias como lengua castellana (3.87), biología (3.85) y sociales (3.97) mostraron una mayor variabilidad en las notas, con diferencias más marcadas entre los estudiantes. En particular, lengua castellana tuvo el coeficiente de variación más alto (0.106), indicando la mayor variabilidad en esta materia. Esto sugiere que sería útil enfocar esfuerzos en mejorar la uniformidad en el rendimiento en estas áreas. Comparado con 2020, se mantiene una tendencia de medias altas en muchas materias, aunque la dispersión en algunas sigue siendo un desafío a abordar.

Tabla 4.4: Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 6° durante el año 2021 de una institución de Medellín.

Variable	BIO	CSO	DIB	ETI	EFI	REL	LCA	ING	MAT	TEC	FIL	EMP
Media	3.85	3.97	4.11	4.23	4.27	4.19	3.87	4.10	4.01	4.19	4.04	4.30
Varianza	0.075	0.086	0.050	0.054	0.090	0.104	0.169	0.109	0.124	0.078	0.106	0.033
Desv.Estándar	0.274	0.293	0.223	0.233	0.299	0.323	0.411	0.330	0.352	0.280	0.325	0.182
Coef.Variación	0.071	0.074	0.054	0.055	0.070	0.077	0.106	0.080	0.088	0.067	0.081	0.042

En la figura 4.1 durante los años 2019, 2020 y 2021, los estudiantes de sexto grado mostraron algunas tendencias consistentes en su desempeño académico. En general, las materias como emprendimiento, tecnología y educación física tendieron a tener medias más altas y menor variabilidad en las notas, lo que indica un rendimiento uniforme y consistentemente bueno en estas áreas. En particular, emprendimiento destacó con las medias más altas en los tres años ver tablas 4.2, 4.3 y 4.4.

Sin embargo, materias como lengua castellana y sociales mostraron consistentemente una mayor variabilidad en las notas, con coeficientes de variación relativamente altos, sugiriendo diferencias más marcadas en el desempeño de los estudiantes. Lengua castellana, especialmente, fue notable por su alta variabilidad en los tres años, lo que indica que esta área podría beneficiarse de esfuerzos adicionales para mejorar la consistencia en el rendimiento de los estudiantes, como se puede apreciar en la misma figura 4.1.

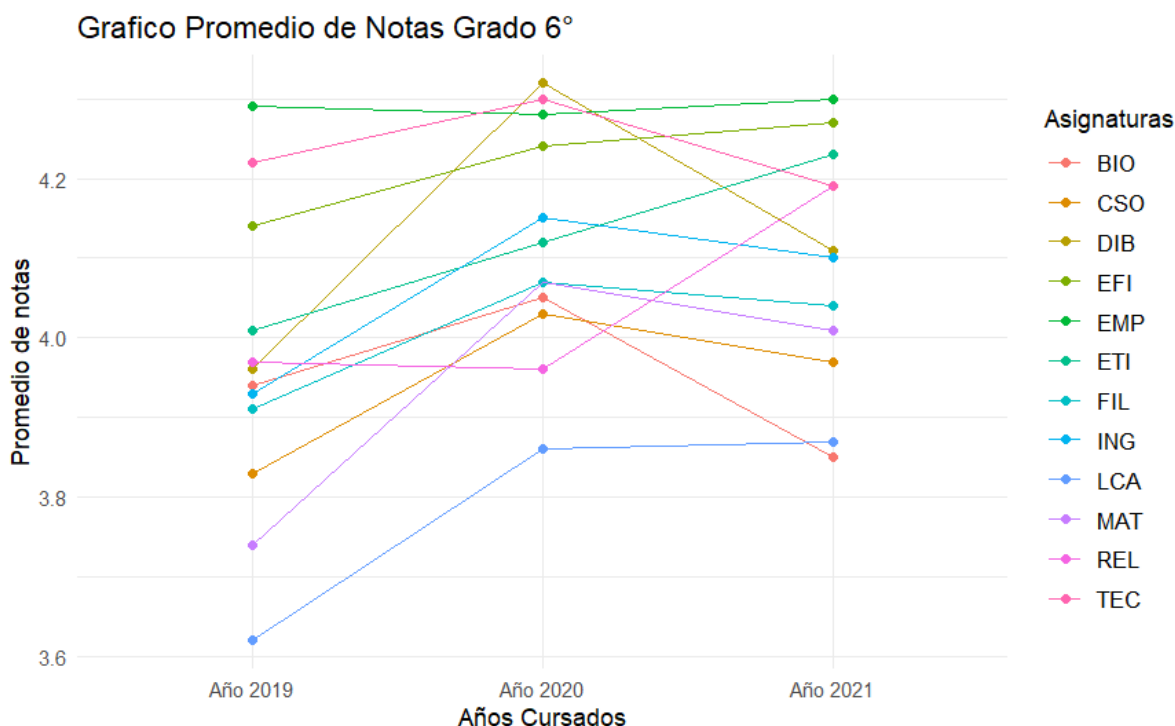


Figura 4.1: Promedio de Notas por Asignatura del 6° Grado: Años 2019-2021 (Fuente: Elaboración Propia)

En términos de tendencias generales, se observa una mejora gradual en las medias de notas de la mayoría de las materias entre 2019 y 2020, con un leve retroceso en algunas áreas en 2021. A pesar de estas fluctuaciones, el patrón de variabilidad y desempeño en materias específicas se mantuvo relativamente constante, destacando tanto áreas de éxito como áreas que necesitan atención continua para mejorar la uniformidad en el rendimiento estudiantil.

4.2.2. Análisis de notas del grado 7°

En la tabla 4.5 se observan las calificaciones generadas durante los años 2019 a 2021 del grado 7°, con un total de 378 observaciones. Este conjunto de datos proporciona una visión detallada del rendimiento académico de los estudiantes a lo largo de tres años consecutivos, permitiendo así un análisis comparativo de su evolución y progreso en distintas asignaturas. El análisis de estas calificaciones es crucial para identificar áreas de mejora en el proceso educativo y desarrollar estrategias pedagógicas que fomenten un aprendizaje más efectivo. Además, estos datos pueden servir como referencia para futuras investigaciones sobre la evolución del rendimiento académico en contextos similares.

En la tabla 4.6, al analizar el año 2019, los estudiantes de séptimo grado mostraron un desempeño uniforme y alto en materias como emprendimiento (4.23), educación física (4.10) y tecnología (4.11), con notas consistentes y baja variabilidad, lo que indica un rendimiento destacado. Emprendimiento tuvo la media más alta, sugiriendo un rendimiento sobresaliente en esta materia.

Tabla 4.5: Calificaciones obtenidas por los estudiantes de grado 7° durante los años escolares 2019, 2020 y 2021 en una institución de Medellín.

BIO: Biología		ETI: Ética		LCA: Lengua Castellana		TEC: Tecnología						
CSO: Sociales		EFI: Educación Física		ING: Inglés		FIL: Filosofía						
DIB: Dibujo		REL: Religión		MAT: Matemáticas		EMP: Emprendimiento						
AÑO	BIO	CSO	DIB	ETI	EFI	REL	LCA	ING	MAT	TEC	FIL	EMP
2019	3,8	3,8	3,6	4,1	4,1	3,7	3,8	3,9	3,7	4,1	4,0	4,3
2019	3,6	3,5	3,8	3,6	4,1	3,5	3,5	3,0	3,5	4,0	3,8	4,2
2019	3,5	3,6	3,7	3,7	4,2	4,0	3,9	3,8	3,5	4,1	4,0	4,3
2019	3,8	3,9	4,1	3,9	4,1	3,8	4,3	4,1	3,9	4,6	4,0	4,3
2019	4,1	4,2	4,1	3,9	4,2	4,1	4,2	4,0	3,7	4,3	4,0	4,5
.
.
.

Por otro lado, materias como lengua castellana (3.92), inglés (3.82) y religión (3.95) presentaron una mayor variabilidad en las notas, con coeficientes de variación más altos, sugiriendo diferencias más marcadas en el desempeño estudiantil. Inglés, en particular, mostró la mayor dispersión en las calificaciones.

En resumen, mientras emprendimiento, educación física y tecnología destacaron por su consistencia y alto rendimiento, lengua castellana, inglés y religión necesitan atención para mejorar la uniformidad en el rendimiento estudiantil.

Tabla 4.6: Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 7° durante el año 2019 de una institución de Medellín.

Variable	BIO	CSO	DIB	ETI	EFI	REL	LCA	ING	MAT	TEC	FIL	EMP
Media	3.78	3.76	3.98	3.93	4.10	3.95	3.92	3.82	3.70	4.11	3.82	4.23
Varianza	0.075	0.075	0.048	0.068	0.025	0.118	0.089	0.128	0.059	0.087	0.057	0.077
Desv.Estándar	0.274	0.274	0.219	0.261	0.160	0.344	0.298	0.358	0.242	0.295	0.239	0.278
Coef.Variación	0.072	0.073	0.055	0.067	0.039	0.087	0.076	0.094	0.066	0.072	0.063	0.066

En la tabla 4.7, al analizar el año 2020, los estudiantes de séptimo grado mostraron un desempeño destacado en materias como ética (4.41), religión (4.28) y tecnología (4.19), con altos promedios y baja variabilidad, indicando un rendimiento académico fuerte y uniforme. Por otro lado, Lengua castellana (3.88) presentó la mayor variabilidad en las calificaciones, con un coeficiente de variación de 0.103, sugiriendo diferencias significativas en el rendimiento de los estudiantes y la necesidad de atención adicional para mejorar la consistencia. En general, mientras ética, religión y tecnología destacaron por su consistencia y altos promedios, materias como lengua castellana, inglés (4.14) y matemáticas (4.09) mostraron mayor dispersión en las notas, lo que indica áreas clave para enfocarse en la mejora académica.

Tabla 4.7: Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 7° durante el año 2020 de una institución de Medellín.

Variable	BIO	CSO	DIB	ETI	EFI	REL	LCA	ING	MAT	TEC	FIL	EMP
Media	4.00	4.01	4.13	4.41	4.16	4.28	3.88	4.14	4.09	4.19	3.98	4.18
Varianza	0.085	0.110	0.076	0.099	0.109	0.116	0.160	0.122	0.127	0.070	0.073	0.085
Desv.Estándar	0.292	0.331	0.276	0.315	0.330	0.341	0.400	0.349	0.356	0.265	0.270	0.292
Coef.Variación	0.073	0.083	0.067	0.071	0.079	0.080	0.103	0.084	0.087	0.063	0.068	0.070

En la tabla 4.8, al analizar el año 2020, se observa que las medias muestran los promedios obtenidos en cada asignatura. Destacan las áreas de Tecnología y Emprendimiento, que registran los puntajes más altos, con valores de (4.29) y (4.31), respectivamente. Esto sugiere un rendimiento generalmente sólido en estas áreas.

Sin embargo, al examinar la varianza y la desviación estándar, se observa que Emprendimiento exhibe una dispersión considerable en los puntajes, indicando una variabilidad notable entre los estudiantes en esta materia. Por otro lado, el coeficiente de variación revela que Filosofía presenta una consistencia notable en los puntajes, con una variabilidad relativa baja en comparación con otras materias.

Estos hallazgos son cruciales para comprender las fortalezas y debilidades de los estudiantes en diferentes áreas de estudio. Esta información puede guiar intervenciones educativas más específicas y efectivas para mejorar el aprendizaje y el rendimiento estudiantil.

Tabla 4.8: Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 7° durante el año 2021 de una institución de Medellín.

Variable	BIO	CSO	DIB	ETI	EFI	REL	LCA	ING	MAT	TEC	FIL	EMP
Media	3.97	3.89	4.20	4.05	4.27	3.94	3.89	4.15	3.95	4.29	4.02	4.31
Varianza	0.098	0.090	0.068	0.054	0.078	0.085	0.105	0.117	0.102	0.051	0.157	0.063
Desv.Estándar	0.313	0.300	0.261	0.233	0.279	0.292	0.324	0.342	0.319	0.227	0.396	0.250
Coef.Variación	0.079	0.077	0.062	0.058	0.065	0.074	0.083	0.082	0.081	0.053	0.099	0.058

En la figura 4.2, se observa una tendencia general al incremento en las medias de las asignaturas evaluadas en séptimo grado durante los años 2019, 2020 y 2021. Por ejemplo, en 2019, la media para Tecnología fue de (3.70), aumentando a (4.29) en 2021. De manera similar, Emprendimiento mostró un aumento significativo en su media, pasando de (3.82) en 2019 a (4.31) en 2021. Sin embargo, algunas asignaturas presentaron fluctuaciones en sus medias a lo largo de los años, como Inglés y Filosofía, lo que sugiere variaciones en el rendimiento estudiantil en estos temas específicos en diferentes momentos.

Respecto a la dispersión de los datos, se observan diferencias notables en la varianza y la desviación estándar entre las diferentes materias y años. Por ejemplo, en 2020, Religión mostró una varianza alta, lo que indica una mayor variabilidad en los puntajes de los estudiantes en comparación con otros años. En cuanto a la variabilidad relativa, el coeficiente de variación generalmente permanece estable en la mayoría de las materias a lo largo de los tres años, aunque se observan algunas excepciones. Por ejemplo, Filosofía exhibió una variabilidad relativamente baja en 2021 en comparación con otros años ver tablas 4.6, 4.7 y 4.8.

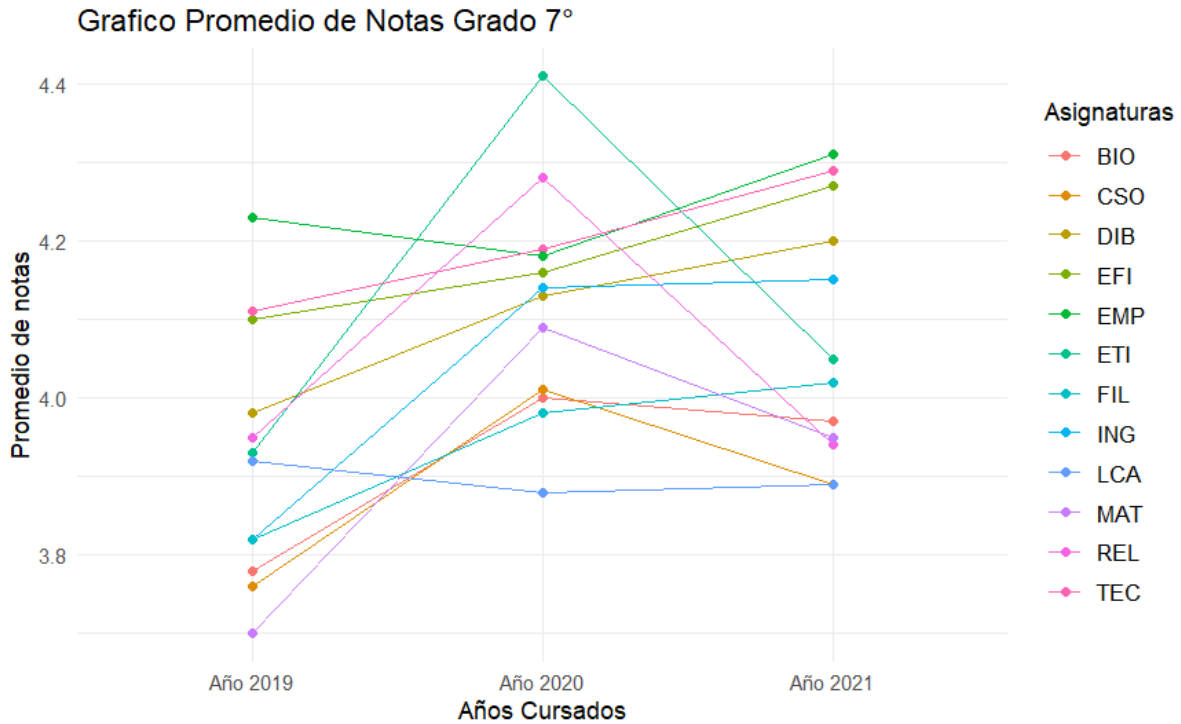


Figura 4.2: Promedio de Notas por Asignatura del 7° Grado: Años 2019-2021 (Fuente: Elaboración Propia)

Estos hallazgos ofrecen una comprensión detallada del rendimiento estudiantil en diferentes áreas a lo largo del tiempo, lo que puede guiar en la identificación de áreas de mejora y en el diseño de programas educativos más efectivos.

4.2.3. Análisis de notas del grado 8°

En la tabla 4.9 se observan las calificaciones generadas durante los años 2019 a 2021 del grado 8°, con un total de 391 observaciones. Este conjunto de datos proporciona una visión detallada del rendimiento académico de los estudiantes a lo largo de tres años consecutivos, permitiendo así un análisis comparativo de su evolución y progreso en distintas asignaturas. El análisis de estas calificaciones es crucial para identificar áreas de mejora en el proceso educativo y desarrollar estrategias pedagógicas que fomenten un aprendizaje más efectivo. Además, estos datos pueden servir como referencia para futuras investigaciones sobre la evolución del rendimiento académico en contextos similares.

Tabla 4.9: Calificaciones obtenidas por los estudiantes de grado 8° durante los años escolares 2019, 2020 y 2021 en una institución de Medellín.

FIS: Física QUI: Química BIO: Biología CSO: Sociales				DIB: Dibujo ETI: Ética EFI: Educación Física REL: Religión				LCA: Lengua Castellana ING: Ingles MAT: Matemáticas TEC: Tecnología			FIL: Filosofía EMP: Emprendimiento			
AÑO	FIS	QUIM	BIO	CSO	DIB	ETI	EFI	REL	LCA	ING	MAT	TEC	FIL	EMP
2019	4,0	4,0	4,0	4,0	4,1	3,8	4,1	4,0	4,1	4,5	4,2	4,1	4,2	4,3
2019	4,1	3,6	3,9	3,9	4,5	3,8	4,1	4,4	4,0	3,9	3,7	3,9	4,4	4,3
2019	4,0	3,9	4,1	4,0	4,1	3,8	4,2	4,4	4,2	4,6	4,4	4,3	4,2	4,4
2019	4,0	3,9	3,8	3,7	4,2	3,7	4,2	4,5	4,0	4,4	4,4	4,1	4,0	3,7
2019	4,2	3,9	4,3	3,8	3,8	4,1	4,4	4,5	4,2	4,6	4,3	4,0	4,3	4,1
.
.
.

En la tabla 4.10, al analizar el año 2019 los estudiantes de octavo grado mostraron un rendimiento destacado y consistente en materias como educación física (4.18), religión (4.09) y filosofía (4.05), con altos promedios y baja variabilidad. En particular, educación física tuvo el coeficiente de variación más bajo (0.039), indicando una notable uniformidad en el rendimiento de los estudiantes. En contraste, materias como inglés (3.97), sociales (3.83) y lengua castellana (3.95) presentaron una mayor variabilidad en las calificaciones, con inglés mostrando la mayor desviación estándar (0.384), lo que refleja diferencias significativas en el desempeño de los estudiantes. En resumen, mientras educación física, religión y filosofía destacaron por su uniformidad y altos promedios, inglés, sociales y lengua castellana requieren atención para mejorar la consistencia en el desempeño estudiantil.

Tabla 4.10: Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 8° durante el año 2019 de una institución de Medellín.

Variable	FIS	QUI	BIO	CSO	DIB	ETI	EFI	REL	LCA	ING	MAT	TEC	FIL	EMP
Media	3.83	3.75	3.81	3.83	4.04	3.77	4.18	4.09	3.95	3.97	3.98	3.91	4.05	3.99
Varianza	0.085	0.069	0.076	0.097	0.069	0.055	0.027	0.130	0.076	0.148	0.114	0.116	0.048	0.098
Desv.Estándar	0.292	0.262	0.276	0.312	0.263	0.234	0.164	0.360	0.275	0.384	0.337	0.340	0.219	0.314
Coef.Variación	0.076	0.070	0.072	0.081	0.065	0.062	0.039	0.088	0.070	0.097	0.085	0.087	0.054	0.079

En la tabla 4.11, al analizar el año 2020, se observa que los estudiantes de octavo grado mostraron un rendimiento destacado en materias como religión (4.32), ética (4.26) y emprendimiento (4.23), con altos promedios y baja variabilidad relativa, destacándose por su consistencia en las calificaciones. Por otro lado, materias como lengua castellana (3.82) y matemáticas (4.01) presentaron una mayor variabilidad en las notas, con coeficientes de variación de 0.077 en ambas, indicando diferencias más marcadas en el rendimiento estudiantil. Física (3.89) y química (3.91) mostraron una variabilidad moderada, con coeficientes de 0.080 y 0.078, respectivamente. En resumen, mientras religión, ética y emprendimiento sobresalieron por su uniformidad y altos promedios, lengua castellana y matemáticas presentaron mayor dispersión en las notas, sugiriendo áreas clave para enfocar esfuerzos en mejorar la uniformidad del rendimiento estudiantil.

Tabla 4.11: Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 8° durante el año 2020 de una institución de Medellín.

Variable	FIS	QUI	BIO	CSO	DIB	ETI	EFI	REL	LCA	ING	MAT	TEC	FIL	EMP
Media	3.89	3.91	4.08	3.74	4.16	4.26	4.07	4.32	3.82	4.15	4.01	4.12	4.07	4.23
Varianza	0.097	0.092	0.067	0.047	0.051	0.055	0.066	0.062	0.087	0.094	0.096	0.061	0.054	0.058
Desv.Estándar	0.311	0.304	0.259	0.217	0.227	0.236	0.257	0.249	0.295	0.307	0.309	0.247	0.233	0.240
Coef.Variación	0.080	0.078	0.064	0.058	0.054	0.055	0.063	0.058	0.077	0.074	0.077	0.060	0.057	0.057

En la tabla 4.12, al analizar el año 2021 los estudiantes de octavo grado mostraron un rendimiento notable en materias como emprendimiento (4.40), tecnología (4.28) y educación física (4.11), con altos promedios y una baja variabilidad relativa, destacándose por su consistencia. En contraste, materias como química (3.92), biología (3.98) y ética (3.93) presentaron una mayor variabilidad en las calificaciones, con coeficientes de variación de 0.108, 0.098 y 0.102 respectivamente, indicando diferencias más significativas en el rendimiento de los estudiantes. Además, física (3.90) mostró una variabilidad considerable con un coeficiente de variación de 0.095. En resumen, mientras emprendimiento, tecnología y educación física destacaron por su uniformidad y altos promedios, materias como química, biología y ética presentaron una mayor dispersión en las notas, sugiriendo áreas clave para enfocar esfuerzos en mejorar la uniformidad del rendimiento estudiantil.

Tabla 4.12: Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 8° durante el año 2021 de una institución de Medellín.

Variable	FIS	QUI	BIO	CSO	DIB	ETI	EFI	REL	LCA	ING	MAT	TEC	FIL	EMP
Media	3.90	3.92	3.98	3.88	4.06	3.93	4.11	3.94	3.83	4.09	3.82	4.28	4.08	4.40
Varianza	0.136	0.178	0.152	0.086	0.074	0.161	0.098	0.122	0.149	0.122	0.136	0.087	0.062	0.049
Desv.Estándar	0.369	0.421	0.390	0.293	0.273	0.402	0.313	0.350	0.387	0.349	0.369	0.295	0.249	0.222
Coef.Variación	0.095	0.108	0.098	0.076	0.067	0.102	0.076	0.089	0.101	0.085	0.097	0.069	0.061	0.050

En la figura 4.3, se observa entre 2019 y 2021, los estudiantes de octavo grado mostraron un rendimiento académico en general sólido con algunas variaciones notables en diferentes materias. En 2019, materias como educación física (4.18), religión (4.09) y filosofía (4.05) destacaron por sus altos promedios y baja variabilidad, mientras que inglés (3.97) y lengua castellana (3.95) presentaron una mayor dispersión en las calificaciones. En 2020, materias como religión (4.32), ética (4.26) y emprendimiento (4.23) continuaron mostrando altos promedios y consistencia, pero lengua castellana (3.82) y matemáticas (4.01) tuvieron una mayor variabilidad. En 2021, las materias emprendimiento (4.40), tecnología (4.28) y educación física (4.11) sobresalieron, aunque química (3.92), biología (3.98) y ética (3.93) mostraron más variabilidad en las calificaciones, ver tablas 4.10, 4.11 y 4.12.

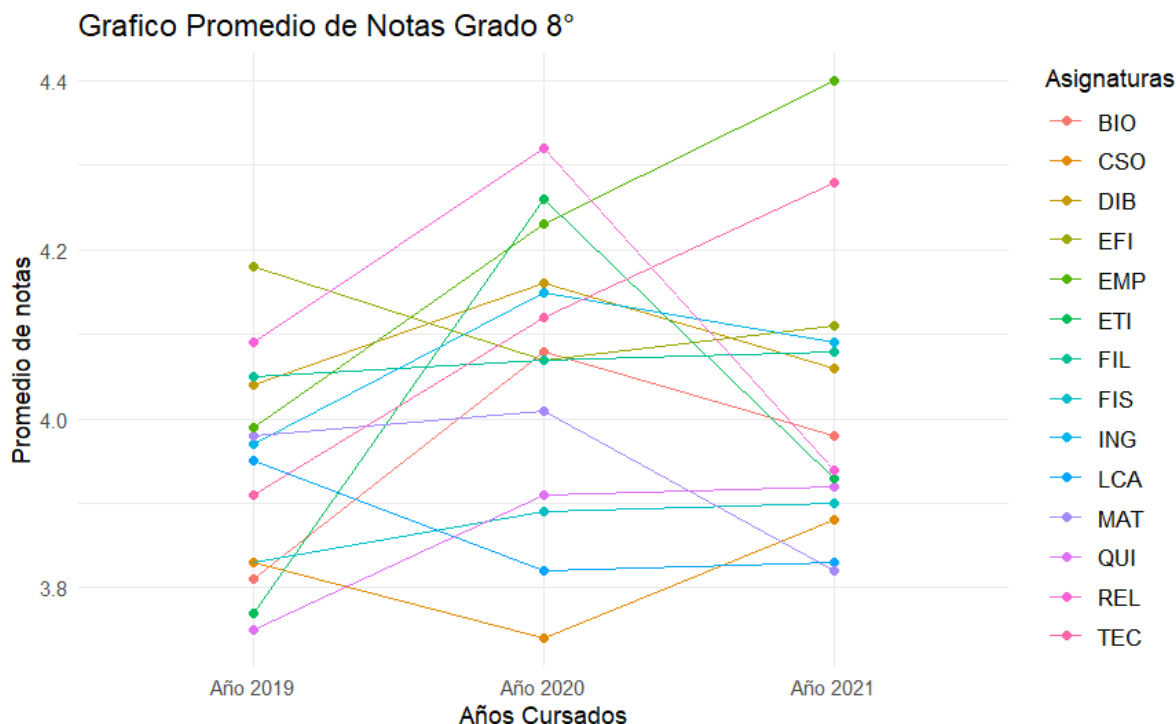


Figura 4.3: Promedio de Notas por Asignatura del 8° Grado: Años 2019-2021 (Fuente: Elaboración Propia)

En resumen, a lo largo de los tres años, materias como educación física, religión y emprendimiento consistentemente mostraron altos promedios y baja variabilidad, indicando un desempeño fuerte y uniforme. En contraste, materias como inglés, lengua castellana, química y biología variaron más en consistencia, sugiriendo áreas para mejorar la uniformidad en el rendimiento estudiantil.

4.2.4. Análisis de notas del grado 9°

En la tabla 4.13 se observan las calificaciones generadas durante los años 2019 a 2021 del grado 9°, con un total de 378 observaciones. Este conjunto de datos proporciona una visión detallada del rendimiento académico de los estudiantes a lo largo de tres años consecutivos, permitiendo así un análisis comparativo de su evolución y progreso en distintas asignaturas. El análisis de estas calificaciones es crucial para identificar áreas de mejora en el proceso educativo y desarrollar estrategias pedagógicas que fomenten un aprendizaje más efectivo. Además, estos datos pueden servir como referencia para futuras investigaciones sobre la evolución del rendimiento académico en contextos similares.

Tabla 4.13: Calificaciones obtenidas por los estudiantes de grado 9° durante los años escolares 2019, 2020 y 2021 en una institución de Medellín.

FIS: Física				DIB: Dibujo				LCA: Lengua Castellana			FIL: Filosofía			
QUI: Química				ETI: Ética				ING: Inglés			EMP: Emprendimiento			
BIO: Biología				EFI: Educación Física				MAT: Matemáticas						
CSO: Sociales				REL: Religión				TEC: Tecnología						
AÑO	FIS	QUIM	BIO	CSO	DIB	ETI	EFI	REL	LCA	ING	MAT	TEC	FIL	EMP
2019	4,0	4,5	3,9	3,6	4,1	4,1	4,0	3,7	3,9	4,4	3,5	3,8	3,5	3,7
2019	3,6	4,0	3,9	3,5	3,7	3,7	4,1	4,1	3,6	3,6	3,5	3,7	3,7	3,7
2019	4,1	4,6	4,4	3,8	4,1	4,3	4,2	4,4	4,2	3,7	3,9	4,1	4,2	4,1
2019	3,7	4,1	3,7	3,7	3,9	3,7	3,9	3,9	3,7	3,7	3,5	3,6	3,5	3,5
2019	3,9	4,5	4,4	3,7	4,3	3,9	4,2	4,1	4,2	4,4	3,8	4,0	4,2	4,2
.
.
.

En la tabla 4.14, al analizar el año 2019, los estudiantes de noveno grado mostraron un desempeño destacado en materias como química (4.16), biología (4.08) y educación física (4.17), con altos promedios y baja variabilidad, especialmente educación física, que tuvo el coeficiente de variación más bajo 0.036. Por otro lado, materias como religión (4.06) e inglés (4.01) presentaron una mayor dispersión en las calificaciones, con coeficientes de variación de 0.094 y 0.090, respectivamente, lo que indica diferencias más significativas en el rendimiento estudiantil. En general, las materias con menor variabilidad relativa, como educación física y dibujo, mostraron un desempeño académico fuerte y uniforme, mientras que materias como religión y matemáticas requieren atención adicional para mejorar la consistencia en las calificaciones.

Tabla 4.14: Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 9° durante el año 2019 de una institución de Medellín.

Variable	FIS	QUI	BIO	CSO	DIB	ETI	EFI	REL	LCA	ING	MAT	TEC	FIL	EMP
Media	3.86	4.16	4.08	3.70	4.09	3.92	4.17	4.06	3.97	4.01	3.65	3.91	3.87	3.96
Varianza	0.066	0.080	0.096	0.041	0.044	0.084	0.023	0.145	0.060	0.132	0.068	0.059	0.061	0.059
Desv.Estándar	0.257	0.283	0.310	0.202	0.211	0.290	0.151	0.381	0.245	0.363	0.261	0.242	0.248	0.242
Coef.Variación	0.067	0.068	0.076	0.055	0.052	0.074	0.036	0.094	0.062	0.090	0.071	0.062	0.064	0.061

En la tabla 4.15, al analizar el año 2020, los estudiantes de noveno grado mostraron un rendimiento académico muy fuerte, especialmente en materias como ética (4.56), ciencias sociales (4.42) y religión (4.37), con altos promedios y baja variabilidad, destacándose ética con el coeficiente de variación más bajo (0.047). Materias como inglés (4.21) y matemáticas (4.04) presentaron mayor variabilidad con coeficientes de variación de 0.090 y 0.087, respectivamente, indicando diferencias más significativas en el rendimiento estudiantil. En general, las materias con menor variabilidad relativa, como ética y religión, mostraron un desempeño académico fuerte y uniforme, mientras que materias como inglés y matemáticas requieren atención adicional para mejorar la consistencia en las calificaciones.

Tabla 4.15: Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 9° durante el año 2020 de una institución de Medellín.

Variable	FIS	QUI	BIO	CSO	DIB	ETI	EFI	REL	LCA	ING	MAT	TEC	FIL	EMP
Media	4.00	4.12	4.16	4.42	4.08	4.56	4.25	4.37	3.87	4.21	4.04	4.32	4.00	4.37
Varianza	0.084	0.105	0.101	0.064	0.110	0.047	0.078	0.051	0.121	0.079	0.123	0.072	0.065	0.060
Desv.Estándar	0.290	0.324	0.319	0.254	0.332	0.216	0.279	0.226	0.347	0.281	0.351	0.269	0.255	0.245
Coef.Variación	0.073	0.079	0.077	0.057	0.082	0.047	0.066	0.052	0.090	0.067	0.087	0.062	0.064	0.056

En la tabla 4.16, al analizar el año 2021, los estudiantes de noveno grado mostraron un rendimiento notable en materias como biología (4.15), química (4.11) y emprendimiento (4.14), con promedios altos y una variabilidad moderada. La materia con la mayor consistencia fue ética, que tuvo un coeficiente de variación de 0.051, indicando un rendimiento bastante uniforme entre los estudiantes. Por otro lado, materias como física (3.87) y ciencias sociales (3.74) presentaron una mayor dispersión en las calificaciones, con coeficientes de variación de 0.090 y 0.066 respectivamente, lo que sugiere diferencias más significativas en el rendimiento estudiantil. En general, mientras que materias como ética, química y biología mostraron un desempeño académico sólido y relativamente consistente, materias como física y ciencias sociales indicaron una mayor necesidad de atención para mejorar la uniformidad en el rendimiento académico de los estudiantes.

Tabla 4.16: Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 9° durante el año 2021 de una institución de Medellín.

Variable	FIS	QUI	BIO	CSO	DIB	ETI	EFI	REL	LCA	ING	MAT	TEC	FIL	EMP
Media	3.87	4.11	4.15	3.74	3.99	4.08	4.06	4.00	3.80	4.09	3.85	4.09	3.93	4.14
Varianza	0.121	0.088	0.116	0.061	0.073	0.042	0.086	0.081	0.099	0.060	0.081	0.096	0.114	0.088
Desv.Estándar	0.347	0.296	0.340	0.247	0.270	0.206	0.293	0.284	0.314	0.246	0.285	0.309	0.337	0.297
Coef.Variación	0.090	0.072	0.082	0.066	0.068	0.051	0.072	0.071	0.083	0.060	0.074	0.076	0.086	0.072

En la figura 4.4, se observa entre 2019 y 2021, el rendimiento académico de los estudiantes de noveno grado mostró tanto consistencia como variabilidad en distintas materias. En 2019, las materias con mayor promedio fueron química (4.16), biología (4.08) y educación física (4.17), destacándose educación física por su baja variabilidad (coeficiente de variación de 0.036). En 2020, los estudiantes lograron altos promedios en ética (4.56), ciencias sociales (4.42) y religión (4.37), siendo ética la materia con menor variabilidad (coeficiente de variación de 0.047). En 2021, las materias biología (4.15), química (4.11) y ética (4.08) mostraron buenos promedios, aunque con una mayor variabilidad en comparación con años anteriores, especialmente en biología (coeficiente de variación de 0.082), ver tablas 4.14, 4.15 y 4.16.

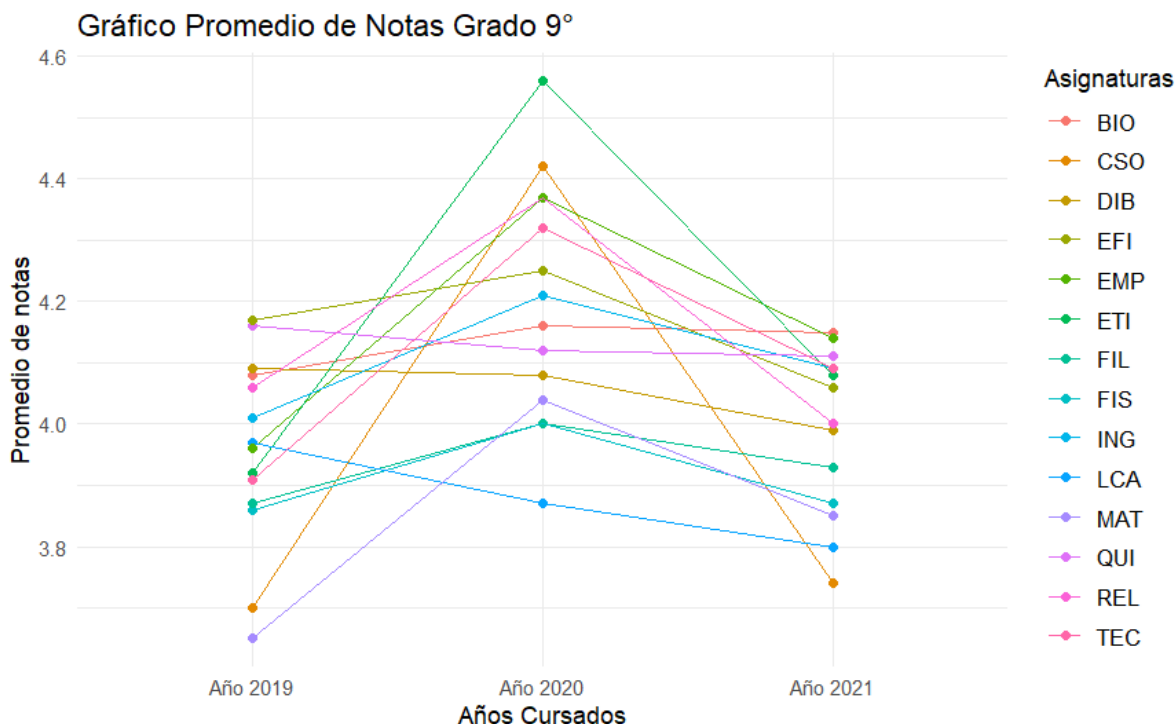


Figura 4.4: Promedio de Notas por Asignatura del 9° Grado: Años 2019-2021 (Fuente: Elaboración Propia)

A lo largo de estos tres años, materias como ética y religión se destacaron consistentemente por su alto rendimiento y baja variabilidad, mientras que materias como inglés y matemáticas mostraron mayor dispersión en las calificaciones, indicando áreas para enfocar esfuerzos en mejorar la uniformidad del rendimiento estudiantil.

Este análisis destaca la influencia del entorno de aprendizaje en el rendimiento de los estudiantes y resalta la importancia de considerar las fortalezas y desafíos de ambos enfoques, virtual y presencial, para optimizar la educación y garantizar un rendimiento académico sólido en cualquier contexto educativo. Continuar monitoreando y evaluando estos resultados es fundamental para abordar de manera efectiva las necesidades de los estudiantes en ambas modalidades de enseñanza.

4.2.5. Análisis de notas del grado 10°

En la tabla 4.17 se observan las calificaciones generadas durante los años 2019 a 2021 del grado 10°, con un total de 364 observaciones. Este conjunto de datos proporciona una visión detallada del rendimiento académico de los estudiantes a lo largo de tres años consecutivos, permitiendo así un análisis comparativo de su evolución y progreso en distintas asignaturas. El análisis de estas calificaciones es crucial para identificar áreas de mejora en el proceso educativo y desarrollar estrategias pedagógicas que fomenten un aprendizaje más efectivo. Además, estos datos pueden servir como referencia para futuras investigaciones sobre la evolución del rendimiento académico en contextos similares.

Tabla 4.17: Calificaciones obtenidas por los estudiantes de grado 10° durante los años escolares 2019, 2020 y 2021 en una institución de Medellín.

FIS: Física		DIB: Dibujo		LCA: Lengua Castellana		FIL: Filosofía								
QUI: Química		ETI: Ética		ING: Ingles		CEP: Ciencia Política								
BIO: Biología		EFI: Educación Física		TRI: Trigonometría										
CSO: Sociales		REL: Religión		TEC: Tecnología										
AÑO	FIS	QUIM	BIO	CSO	DIB	ETI	EFI	REL	LCA	ING	TRI	TEC	FIL	EMP
2019	3,6	3,7	4,4	3,6	4,2	3,9	4,1	3,6	4,1	4,0	3,7	4,3	3,6	3,6
2019	3,3	3,4	4,0	3,6	3,6	2,8	3,9	2,9	3,3	3,8	3,5	3,8	3,6	3,6
2019	3,9	3,7	4,3	3,5	3,7	4,1	4,2	3,9	4,0	4,5	3,6	4,5	4,2	3,7
2019	4,5	4,1	4,5	4,0	4,5	4,4	4,3	4,3	4,4	4,3	4,2	4,5	4,2	4,2
2019	3,5	3,6	4,4	3,8	3,8	3,8	3,9	3,5	3,8	3,6	3,5	4,0	3,9	3,7
.
.
.

En la tabla 4.18 se muestran los promedios de notas de estudiantes de décimo grado de un colegio en Medellín para el año 2019, abarcando diversas materias. Biología tiene la calificación promedio más alta con (4.16), mientras que trigonometría presenta la más baja con (3.64). La variabilidad en las notas es más alta en dibujo, con una desviación estándar de (0.408) y un coeficiente de variación de (0.106), indicando grandes diferencias en el desempeño de los estudiantes en esta materia. En contraste, educación física muestra la menor variabilidad, con una desviación estándar de (0.143) y un coeficiente de variación de (0.035), sugiriendo un desempeño más consistente entre los estudiantes. Estos datos permiten identificar áreas donde los estudiantes destacan o podrían necesitar apoyo adicional.

Tabla 4.18: Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 10° durante el año 2019 de una institución de Medellín.

Variable	FIS	QUI	BIO	CSO	DIB	ETI	EFI	REL	LCA	ING	TRI	TEC	FIL	CEP
Media	3.80	3.69	4.16	3.76	3.86	3.90	4.15	3.83	3.78	4.11	3.64	3.98	3.86	3.91
Varianza	0.158	0.082	0.069	0.079	0.167	0.100	0.021	0.119	0.154	0.094	0.139	0.112	0.100	0.101
Desv.Estándar	0.397	0.286	0.263	0.282	0.408	0.316	0.143	0.345	0.393	0.307	0.373	0.334	0.316	0.318
Coef.Variación	0.104	0.078	0.063	0.075	0.106	0.081	0.035	0.090	0.104	0.075	0.103	0.084	0.082	0.081

En la tabla 4.19 para el año 2020, los promedios de notas de los estudiantes de décimo grado de un colegio en Medellín muestran algunas variaciones con respecto al año anterior. La media de las calificaciones varía entre (3.83) en ética y (4.39) en religión. Religión tiene el promedio más alto con (4.39), mientras que ética tiene el más bajo con (3.83). En términos de variabilidad, dibujo presenta la mayor varianza, (0.109), y desviación estándar, (0.330), lo que sugiere una alta dispersión en las calificaciones de esta materia. La menor variabilidad relativa se observa en biología con un coeficiente de variación de (0.046), indicando un desempeño más uniforme entre los estudiantes. En general, las calificaciones del 2020 muestran una tendencia a tener menos dispersión en comparación con 2019, reflejando una mejora en la consistencia del desempeño estudiantil.

Tabla 4.19: Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 10° durante el año 2020 de una institución de Medellín.

Variable	FIS	QUI	BIO	CSO	DIB	ETI	EFI	REL	LCA	ING	TRI	TEC	FIL	CEP
Media	4.03	3.89	4.22	3.92	4.01	3.83	4.21	4.39	3.86	4.29	4.04	4.35	4.06	3.95
Varianza	0.093	0.051	0.037	0.061	0.109	0.055	0.074	0.042	0.084	0.059	0.097	0.045	0.073	0.052
Desv.Estándar	0.305	0.227	0.193	0.247	0.330	0.234	0.272	0.204	0.290	0.243	0.311	0.212	0.270	0.228
Coef.Variación	0.076	0.058	0.046	0.063	0.082	0.061	0.065	0.047	0.075	0.057	0.077	0.049	0.067	0.058

En la tabla 4.20 para el año 2021, tecnología destaca con el promedio de calificaciones más alto (4.26), mientras que ciencias sociales tiene el promedio más bajo (3.83). Lengua castellana muestra la mayor dispersión y variabilidad relativa en las notas, con una varianza de (0.129) y un coeficiente de variación de (0.093). Por otro lado, química presenta la menor dispersión y variabilidad relativa, sugiriendo una mayor consistencia en el desempeño estudiantil en esta materia. Estos datos ayudan a identificar áreas con altos y bajos rendimientos, así como la consistencia del desempeño académico de los estudiantes en diversas materias.

Tabla 4.20: Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 10° durante el año 2021 de una institución de Medellín.

Variable	FIS	QUI	BIO	CSO	DIB	ETI	EFI	REL	LCA	ING	TRI	TEC	FIL	CEP
Media	3.99	3.84	4.14	3.83	4.09	3.90	4.19	4.15	3.85	4.22	3.89	4.26	4.00	3.91
Varianza	0.091	0.031	0.053	0.094	0.115	0.074	0.083	0.101	0.129	0.097	0.104	0.096	0.073	0.105
Desv.Estándar	0.301	0.175	0.230	0.307	0.339	0.272	0.289	0.318	0.359	0.311	0.323	0.310	0.270	0.323
Coef.Variación	0.075	0.046	0.056	0.080	0.083	0.070	0.069	0.077	0.093	0.074	0.083	0.073	0.067	0.083

En la figura 4.5, se observa entre 2019 y 2021, indican que las materias de ciencias naturales, como biología y química, tienden a tener promedios de calificaciones más altos y una mayor consistencia en el desempeño de los estudiantes. Las materias de ciencias sociales y matemáticas, como ciencias sociales y trigonometría, presentan más desafíos, reflejados en promedios más bajos y una mayor variabilidad en las calificaciones. La identificación de estas tendencias puede ayudar a los educadores a focalizar sus esfuerzos en mejorar el rendimiento en áreas específicas y apoyar a los estudiantes de manera más efectiva, ver tablas 4.18, 4.19 y 4.20.

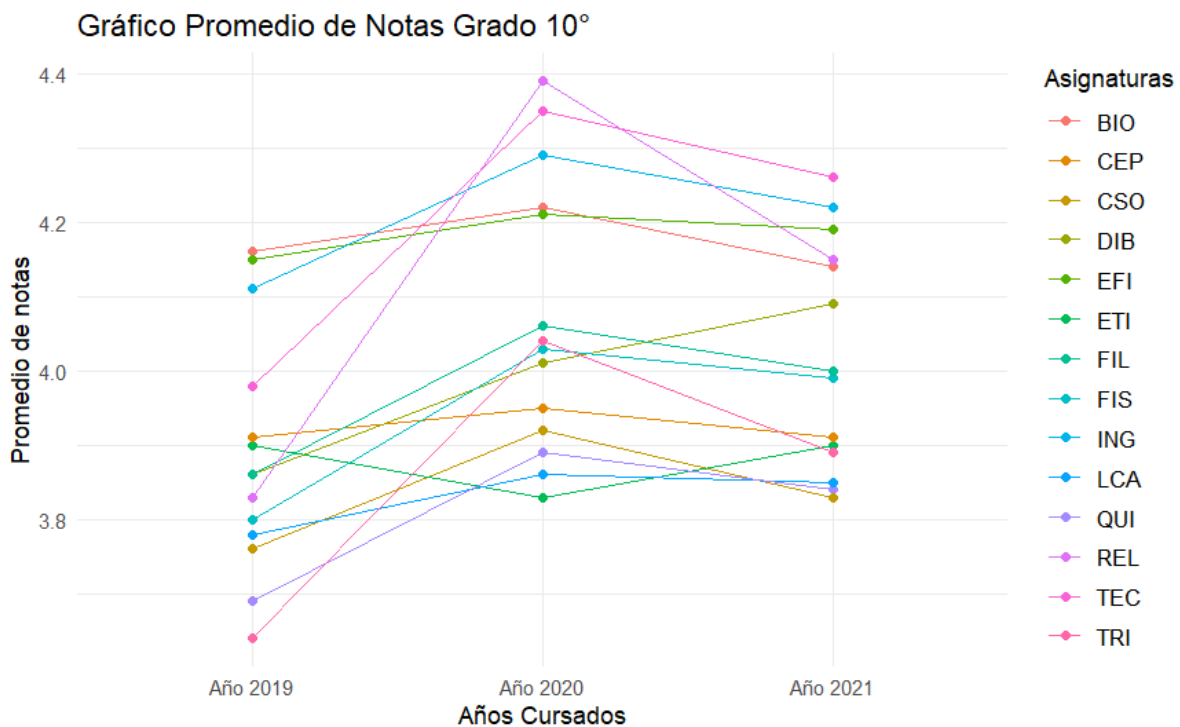


Figura 4.5: Promedio de Notas por Asignatura del 10° Grado: Años 2019-2021 (Fuente: Elaboración Propia)

En general, se observa cierta variabilidad en los promedios de notas a lo largo de los tres años, con aumentos en algunas asignaturas en 2020 y disminuciones en otras en 2021. Es importante continuar monitoreando y analizando estos datos para identificar áreas de mejora y asegurar un crecimiento constante en el futuro.

4.2.6. Análisis de notas del grado 11°

En la tabla 4.21 se observan las calificaciones generadas durante los años 2019 a 2021 del grado 11°, con un total de 353 observaciones. Este conjunto de datos proporciona una visión detallada del rendimiento académico de los estudiantes a lo largo de tres años consecutivos, permitiendo así un análisis comparativo de su evolución y progreso en distintas asignaturas. El análisis de estas calificaciones es crucial para identificar áreas de mejora en el proceso educativo y desarrollar estrategias pedagógicas que fomenten un aprendizaje más efectivo. Además, estos datos pueden servir como referencia para futuras investigaciones sobre la evolución del rendimiento académico en contextos similares.

Tabla 4.21: Calificaciones obtenidas por los estudiantes de grado 11° durante los años escolares 2019, 2020 y 2021 en una institución de Medellín.

FIS: Física		DIB: Dibujo		LCA: Lengua Castellana		FIL: Filosofía								
QUI: Química		ETI: Ética		ING: Inglés		CEP: Ciencia Política								
BIO: Biología		EFI: Educación Física		CAL: cálculo										
CSO: Sociales		REL: Religión		TEC: Tecnología										
AÑO	FIS	QUIM	BIO	CSO	DIB	ETI	EFI	REL	LCA	ING	CAL	TEC	FIL	EMP
2019	4,2	4,1	4,5	4,4	3,9	4,1	4,4	4,0	4,2	4,8	4,2	4,5	4,2	4,2
2019	3,7	3,5	4,1	3,7	3,6	3,7	4,1	3,5	3,6	3,9	3,9	4,2	3,8	3,8
2019	4,4	4,0	4,6	4,0	4,4	4,2	4,3	4,1	3,8	4,2	4,3	4,4	4,1	4,3
2019	4,4	4,3	4,8	4,4	4,1	4,1	4,4	4,0	4,2	4,1	4,3	4,4	4,3	4,3
2019	3,6	3,3	4,2	3,6	3,7	3,9	4,3	3,7	3,6	3,6	3,5	4,3	3,8	4,2
.
.
.

En la tabla 4.22, se muestra que en 2019, los estudiantes de undécimo grado presentaron un promedio de notas que varió entre (3.73) en química y (4.31) en biología, siendo esta última la materia con el promedio más alto. La mayor variabilidad en las calificaciones se observó en inglés, con una varianza de (0.154), mientras que la menor varianza se encontró en educación física, con (0.033), indicando una mayor consistencia en esta última. El coeficiente de variación fue más bajo en educación física, con (0.042), sugiriendo una uniformidad notable en las calificaciones, mientras que fue más alto en química, con (0.095), señalando mayor dispersión en los resultados.

Tabla 4.22: Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 11° durante el año 2019 de una institución de Medellín.

Variable	FIS	QUI	BIO	CSO	DIB	ETI	EFI	REL	LCA	ING	CAL	TEC	FIL	CEP
Media	3.92	3.73	4.31	3.88	3.96	4.01	4.25	3.83	3.81	4.17	3.85	4.25	3.96	4.07
Varianza	0.121	0.124	0.057	0.092	0.125	0.093	0.033	0.105	0.117	0.154	0.102	0.041	0.074	0.081
Desv.Estándar	0.348	0.353	0.239	0.304	0.353	0.305	0.180	0.325	0.343	0.393	0.320	0.203	0.272	0.285
Coef.Variación	0.089	0.095	0.055	0.078	0.089	0.076	0.042	0.085	0.090	0.094	0.083	0.048	0.069	0.070

En la tabla 4.23, se muestra que en 2020, las calificaciones promedio de los estudiantes de undécimo grado presentan una mejora general, con promedios que van de (4.04) en ciencias sociales y química a (4.44) en ética. Dibujo muestra la mayor variabilidad con

una varianza de (0.143), mientras que ética tiene la menor variabilidad con una varianza de (0.047). El coeficiente de variación más bajo se observa en ética con (0.049), indicando un desempeño consistente, y el más alto en dibujo con (0.091), reflejando una mayor dispersión en las calificaciones.

Tabla 4.23: Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 11° durante el año 2020 de una institución de Medellín.

Variable	FIS	QUI	BIO	CSO	DIB	ETI	EFI	REL	LCA	ING	CAL	TEC	FIL	CEP
Media	4.06	4.04	4.21	4.04	4.14	4.44	4.27	4.36	4.05	4.34	4.13	4.42	4.11	4.06
Varianza	0.099	0.072	0.062	0.065	0.143	0.047	0.066	0.045	0.099	0.079	0.079	0.037	0.071	0.068
Desv.Estándar	0.315	0.269	0.249	0.256	0.379	0.216	0.256	0.213	0.314	0.282	0.282	0.193	0.267	0.260
Coef.Variación	0.078	0.067	0.059	0.063	0.091	0.049	0.060	0.049	0.078	0.065	0.068	0.044	0.065	0.064

En la tabla 4.24, se muestra que en 2021, los promedios de notas oscilan entre (3.84) en lengua castellana y (4.25) en inglés. La mayor variabilidad en las calificaciones se encuentra en dibujo y lengua castellana, con una varianza de (0.080), y la menor en química, con una varianza de (0.042). El coeficiente de variación es más bajo en química, con (0.053), indicando menor dispersión, y más alto en lengua castellana, con (0.074), reflejando mayor variabilidad en el desempeño estudiantil.

Tabla 4.24: Medidas de dispersión de las calificaciones obtenidas por los estudiantes de grado 11° durante el año 2021 de una institución de Medellín.

Variable	FIS	QUI	BIO	CSO	DIB	ETI	EFI	REL	LCA	ING	CAL	TEC	FIL	CEP
Media	3.90	3.85	4.22	4.00	4.22	3.99	4.23	4.15	3.84	4.25	4.23	4.17	4.02	3.97
Varianza	0.061	0.042	0.047	0.047	0.080	0.050	0.056	0.060	0.080	0.091	0.070	0.061	0.081	0.083
Desv.Estándar	0.246	0.205	0.216	0.218	0.283	0.224	0.237	0.245	0.283	0.302	0.265	0.247	0.284	0.288
Coef.Variación	0.063	0.053	0.051	0.054	0.067	0.056	0.056	0.059	0.074	0.071	0.063	0.059	0.071	0.072

En la figura 4.6, se observa que entre 2019 y 2021, hay una tendencia general de mejora en los promedios de calificaciones de los estudiantes de undécimo grado. En 2019, la dispersión de calificaciones es más pronunciada en materias como química y dibujo, mientras que en 2020 se observa una reducción en la variabilidad de las calificaciones en la mayoría de las materias, con una mejora general en los promedios, destacándose ética como la materia con mejor desempeño. En 2021, aunque las calificaciones promedio se mantienen altas, la variabilidad en ciertas materias como dibujo y lengua castellana sigue siendo notable. En general, las calificaciones de materias de ciencias naturales como biología y química tienden a ser altas y consistentes, mientras que materias como dibujo y lengua castellana muestran mayor dispersión, lo cual sugiere áreas de posible enfoque para mejorar el desempeño estudiantil en el futuro, ver tablas 4.22, 4.23 y 4.24.

Este análisis puede ayudar a identificar áreas de mejora y fortalezas en el desempeño académico de los estudiantes del grado undécimo, y es importante continuar monitoreando y analizando estos datos para garantizar un crecimiento constante en el futuro.

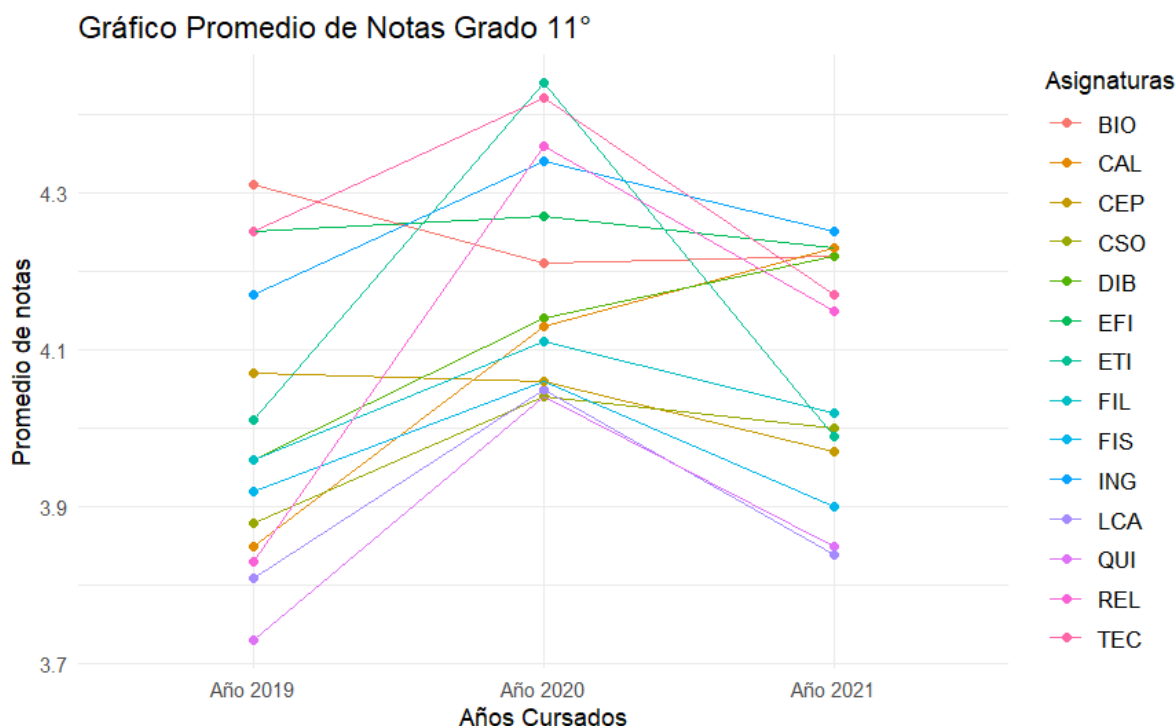


Figura 4.6: Promedio de Notas por Asignatura del 11° Grado: Años 2019-2021 (Fuente: Elaboración Propia)

Por último, este análisis subraya la importancia de considerar las ventajas y desventajas de diferentes modalidades de enseñanza y cómo pueden afectar el rendimiento académico. Continuar monitoreando y evaluando estos resultados es fundamental para adaptar estrategias educativas y garantizar un rendimiento sólido en el futuro.

4.2.7. Análisis de Resultados Prueba Saber 11°

El Instituto colombiano para la evaluación de la educación (ICFES) emplea diferentes modelos psicométricos para la medición de trazos latentes de los evaluados. En la mayoría de las pruebas, los trazos latentes corresponden a las habilidades y a los factores asociados al aprendizaje de los estudiantes. Cada uno de estos modelos tiene características propias de la medición: el tipo de escala de respuesta de los ítems (pregunta cerrada, es decir, de selección múltiple con única respuesta donde una respuesta correcta tiene 1 y una respuesta incorrecta tiene 0, pregunta abierta o escala tipo Likert) o el tamaño poblacional de los evaluados, por último el ICFES cada asignatura la califica en una escala de 0 a 100. Estas calificaciones se utilizan para evaluar el desempeño de los estudiantes en diversas áreas del conocimiento durante las pruebas de estado que se realizan en Colombia, ver (ICFES, 2021)

En la tabla 4.25 se observan las calificaciones generadas durante los años 2019 a 2021 de las pruebas saber 11°, con un total de 353 observaciones. Este conjunto de datos proporciona una visión detallada del rendimiento académico de los estudiantes a lo largo de tres años consecutivos, permitiendo así un análisis comparativo de su evolución y progreso en distintas asignaturas. El análisis de estas calificaciones es crucial para identificar áreas

de mejora en el proceso educativo y desarrollar estrategias pedagógicas que fomenten un aprendizaje más efectivo. Además, estos datos pueden servir como referencia para futuras investigaciones sobre la evolución del rendimiento académico en contextos similares.

Tabla 4.25: Calificaciones obtenidas por los estudiantes de grado 11° en las Pruebas Saber durante los años escolares 2019, 2020 y 2021 en una institución de Medellín.

LCA: Lengua Castellana		CN: Ciencias Naturales			
MAT: Matemáticas		ING: Ingles			
CSO: Sociales					
AÑO	LCA	MAT	CSO	CN	ING
2019	58	60	55	57	51
2019	63	61	62	66	77
2019	59	66	65	58	79
2019	48	42	43	47	29
2019	61	70	69	72	77
.
.
.

En la tabla 4.26, se muestra que en 2019, los resultados de las pruebas saber 11° indican que los estudiantes obtuvieron el puntaje promedio más alto en inglés con (65.23) y el más bajo en ciencias sociales con (57.50). La mayor variabilidad se observa también en inglés con una varianza de (154.712), lo que indica diferencias significativas en los puntajes de los estudiantes. La menor variabilidad se encuentra en lengua castellana con una varianza de (71.606). El coeficiente de variación más alto se presenta en ciencias sociales con (0.213), lo que sugiere una mayor dispersión relativa en los puntajes, mientras que lengua castellana tiene el coeficiente de variación más bajo con (0.137), indicando mayor consistencia en los resultados.

Tabla 4.26: Medidas de dispersión de los resultados obtenidos por los estudiantes de grado 11° en las pruebas saber durante el año 2019 de una institución de Medellín.

Variable	LCA	MAT	SOC	CN	ING
Media	61.68	61.19	57.50	58.24	65.23
Varianza	71.606	101.318	149.849	103.023	154.712
Desv.Estándar	8.462	10.066	12.241	10.150	12.438
Coef.Variación	0.137	0.164	0.213	0.174	0.191

En la tabla 4.27, Se presentan los resultados del año 2020, donde las calificaciones promedio muestran mejoras en la mayoría de las áreas evaluadas. El puntaje promedio más alto se obtuvo en inglés con (64.17), mientras que el más bajo se registró en ciencias naturales con (57.92). La mayor variabilidad en los puntajes se observa en matemáticas con una varianza de (140.675), mientras que lengua castellana presenta la menor variabilidad con una varianza de (60.215). Además, el coeficiente de variación más alto se registra nuevamente en ciencias sociales con (0.170), lo que indica una mayor dispersión en los puntajes en esta área. Por otro lado, lengua castellana tiene el coeficiente de variación más bajo con (0.126), lo que sugiere una mayor uniformidad en los puntajes obtenidos en esta materia.

Tabla 4.27: Medidas de dispersión de los resultados obtenidos por los estudiantes de grado 11° en las pruebas saber durante el año 2020 de una institución de Medellín.

Variable	LCA	MAT	SOC	CN	ING
Media	61.48	62.79	60.34	57.92	64.17
Varianza	60.215	140.675	105.009	87.184	133.489
Desv.Estándar	7.760	11.861	10.247	9.337	11.554
Coef.Variación	0.126	0.189	0.170	0.161	0.180

En la tabla 4.28, se presentan los resultados del año 2021. Se observa una notable mejora en los promedios de calificaciones, especialmente en inglés que alcanza (71.44), el promedio más alto registrado en los tres años. El puntaje promedio más bajo es en ciencias naturales con (59.35). La mayor variabilidad se encuentra en ciencias sociales con una varianza de (96.265), mientras que la menor se presenta en lengua castellana con una varianza de (48.494). El coeficiente de variación más alto se encuentra en ciencias sociales con (0.162), sugiriendo mayor dispersión relativa, y el más bajo en lengua castellana con (0.107), indicando mayor consistencia en los resultados.

Tabla 4.28: Medidas de dispersión de los resultados obtenidos por los estudiantes de grado 11° en las pruebas saber durante el año 2021 de una institución de Medellín.

Variable	LCA	MAT	SOC	CN	ING
Media	65.04	60.88	60.44	59.35	71.44
Varianza	48.494	88.743	96.265	70.347	94.652
Desv.Estándar	6.964	9.420	9.811	8.387	9.729
Coef.Variación	0.107	0.155	0.162	0.141	0.136

En la figura 4.7, se observa que entre 2019 y 2021, hay una mejora general en los puntajes promedio de las pruebas saber 11°, especialmente en inglés, donde los estudiantes consistentemente obtienen los puntajes más altos. La variabilidad en los puntajes tiende a disminuir con el tiempo, reflejando una mejora en la uniformidad de los resultados. Lengua castellana se destaca por tener la menor variabilidad en todos los años, indicando un desempeño más consistente en esta área. En contraste, ciencias sociales y matemáticas muestran mayores dispersiones en los puntajes, sugiriendo que estas áreas pueden beneficiarse de un enfoque educativo más uniforme. En general, los resultados indican que los estudiantes han mejorado su desempeño, especialmente en inglés, y muestran una tendencia hacia una mayor consistencia en los resultados de las pruebas a lo largo del tiempo.

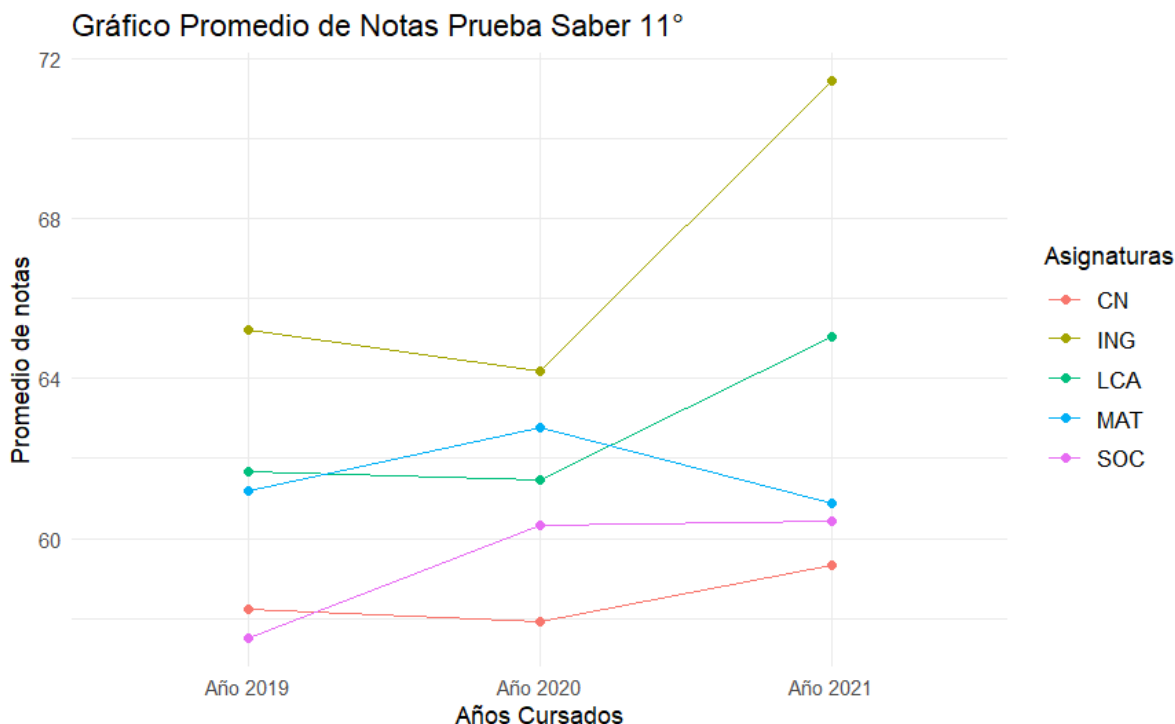


Figura 4.7: Promedio de Notas por Asignatura de la Prueba Saber : Años 2019-2021 (Fuente: Elaboración Propia)

Después de analizar los resultados presentados, como insumo se sugiere que la institución preste atención a las tendencias identificadas y reconozca la importancia de fortalecer las asignaturas que mostraron un descenso en sus resultados durante el año 2021. Para abordar esta situación, se proponen varias estrategias educativas:

Uso de herramientas digitales mejoradas: La implementación de herramientas digitales más avanzadas podría facilitar el proceso de enseñanza y aprendizaje, permitiendo a los estudiantes acceder a recursos educativos de alta calidad y mejorar su comprensión de los temas.

Realización de simulacros internos específicos: Se podrían realizar simulacros internos adaptados a las necesidades específicas de los estudiantes en áreas donde se identificaron deficiencias. Estos simulacros podrían ayudar a los estudiantes a practicar y mejorar en los temas en los que tienen dificultades, preparándolos mejor para futuras evaluaciones.

Implementación de estrategias de apoyo individualizado: Se podría ofrecer apoyo adicional a los estudiantes que necesiten ayuda extra en áreas específicas. Esto podría incluir tutorías individuales, grupos de estudio o actividades de refuerzo diseñadas para abordar las necesidades individuales de cada estudiante.

Evaluación continua y ajustes en el plan de estudios: Es importante realizar una evaluación continua de los resultados académicos y ajustar el plan de estudios según sea necesario. Esto permitirá a la institución adaptarse a las necesidades cambiantes de los estudiantes y garantizar un crecimiento continuo en el futuro.

Al implementar estas estrategias, la institución educativa puede trabajar para mejorar

el rendimiento académico de los estudiantes en las áreas identificadas y garantizar un desarrollo integral y continuo en su educación.

Capítulo 5

Métodos Estadísticos Multivariados

En este capítulo, se presentará la técnica estadística multivariada conocida como Análisis de Componentes Principales (ACP). El ACP permite reducir la dimensionalidad de los datos, facilitando su interpretación y visualización, al tiempo que conserva la mayor cantidad de variabilidad presente en los datos originales. A continuación, se describe el proceso detallado que se sigue para construir instrumentos estadísticos utilizando el ACP.

Normalizar los datos, este paso es crucial cuando las variables tienen diferentes escalas. La normalización ajusta los datos para que tengan una media de cero y una desviación estándar de uno, garantizando que todas las variables contribuyan equitativamente al análisis.

Calcular la matriz de covarianza, la matriz de covarianza proporciona una medida de cómo varían conjuntamente las variables. Es fundamental para entender las relaciones lineales entre las variables y es la base para calcular los componentes principales.

Calcular los valores y vectores propios de la matriz de covarianza, los valores propios indican la cantidad de variabilidad explicada por cada componente principal, mientras que los vectores propios muestran las direcciones de máxima varianza en el espacio de los datos.

Seleccionar los componentes principales con los mayores valores propios, se eligen los componentes principales que tienen los mayores valores propios, ya que estos explican la mayor parte de la variabilidad en los datos. Esto permite una reducción de dimensionalidad efectiva.

Transformar los datos originales utilizando estos componentes, finalmente, se proyectan los datos originales en el espacio definido por los componentes principales seleccionados, obteniendo así una nueva representación de los datos que facilita su análisis y visualización.

En este análisis se llevará a cabo un ACP para los grados de sexto a undécimo, utilizando la base de datos correspondiente a los resultados académicos de dichos grados. Cada grado cursa un promedio de 12 a 14 asignaturas; específicamente, del grado sexto y séptimo tienen 12 asignaturas, mientras que del grado octavo a undécimo tienen 14 asignaturas. Se presentarán los resultados del análisis de componentes principales, eligiendo las 30 mejores observaciones y las 6 variables con mayor calidad de contribución.

En los gráficos expuestos se observará que los puntos representan observaciones o estudiantes, mientras que las flechas representan cada una de las seis variables ilustradas. Cada punto en el gráfico corresponde a un estudiante en el plano bidimensional definido por las componentes uno y dos. Los puntos cercanos entre sí indican estudiantes con un desempeño similar en las asignaturas, mientras que los puntos más alejados representan estudiantes más heterogéneos en cuanto a las asignaturas evaluadas. Se espera que la representación biplot sea una buena aproximación de los datos originales en baja dimensión, en este caso, en dos dimensiones.

Por último, dado que este trabajo está enfocado al análisis de las pruebas Saber del grado 11°, se llevarán a cabo tres ACP donde se representará un año en el gráfico, mientras que los otros dos años restantes se considerarán como individuos suplementarios. Se incluyeron estas observaciones suplementarias en el análisis para determinar si era justificable considerarlas como activas.

5.1. Análisis de Componentes Principales (ACP) Grado 6°

A continuación, se presenta la base de datos que detalla los resultados académicos del grado sexto, ofreciendo una visión detallada del rendimiento estudiantil en las materias cursadas durante cada año. La unión de los datos de estos tres años en una sola base se debe a que el currículo no varía según el grado, aunque las temáticas abordadas en las diferentes materias pueden cambiar con el tiempo. Sin embargo, los resultados académicos mantienen unos promedios muy similares a lo largo de estos años.

Para el grado sexto se incluyen 369 observaciones, como se muestra en la tabla 5.1. Utilizando el Análisis de Componentes Principales (ACP) en la mencionada base de datos, se evaluó el desempeño académico de los estudiantes de grado sexto en una institución educativa de Medellín durante el periodo de 2019 a 2021. En la figura 5.1 se presenta el biplot de los resultados del ACP con las 30 mejores observaciones y las seis variables de mayor calidad de contribución en el primer plano factorial, según el criterio de \cos^2 .

Tabla 5.1: Calificaciones obtenidas por los estudiantes de grado 6° durante los años escolares 2019, 2020 y 2021 en una institución de Medellín.

BIO: Biología			ETI: Ética			LCA: Lengua Castellana			TEC: Tecnología		
CSO: Sociales			EFI: Educación Física			ING: Inglés			FIL: Filosofía		
DIB: Dibujo			REL: Religión			MAT: Matemáticas			EMP: Emprendimiento		
BIO	CSO	DIB	ETI	EFI	REL	LCA	ING	MAT	TEC	FIL	EMP
4,3	4,1	4,2	4,5	4,1	4,7	4,1	4,3	4,2	4,5	4,4	4,4
3,8	3,5	3,5	4,0	3,9	3,5	3,5	3,6	3,6	3,7	3,8	4,0
4,2	3,7	4,0	4,5	4,2	4,5	3,8	4,1	3,7	4,4	4,1	4,6
3,8	3,6	3,5	3,5	4,0	3,8	3,5	3,5	3,5	4,0	3,7	3,6
3,9	4,0	3,7	4,0	4,2	4,5	3,5	3,5	3,8	4,1	3,9	4,3
.
.
.

Al observar la tabla de valores propios del grado 6°, se puede notar que las primeras tres componentes (CP1, CP2, CP3) explican aproximadamente el 74.92 % de la varianza total, mientras que las seis componentes explican alrededor del 87.41 % ver tabla 5.2.

Tabla 5.2: Tabla de valores propios asociada a cada una de las CP del Grado 6°

CP	Valor. Propio	Var. Explicada	Acumulado
CP1	7.6187476	63.48956	63.48956
CP2	0.7492153	6.243461	69.73302
CP3	0.6226380	5.188650	74.92167
CP4	0.5808318	4.840265	79.76194
CP5	0.4778875	3.982396	83.74434
CP6	0.439802	3.665016	87.40935

Tabla 5.3: Tabla de vectores propios de la matriz de correlación grado 6°

Asignaturas	CP1	CP2
bio	0.2988	0.2204
cso	0.3140	0.0814
dib	0.2625	0.4979
eti	0.2955	-0.2636
efi	0.2724	-0.2245
rel	0.2752	-0.5488
lca	0.3308	-0.0276
ing	0.2454	0.4682
mat	0.3143	0.1059
tec	0.2846	-0.0089
fil	0.2923	-0.0472
emp	0.2664	-0.2058

Al analizar la tabla 5.3, se observa que en la componente principal uno (CP1), asignaturas como lengua castellana, matemáticas y sociales tienen las mayores contribuciones. Esto sugiere que CP1 podría estar relacionado con una combinación de habilidades en ciencias y humanidades. Por otro lado, las asignaturas que más contribuyen en la componente principal dos CP2 son religión, dibujo e inglés. Esto sugiere que CP2 podría estar capturando una dimensión que distingue entre áreas más prácticas y expresivas frente a las académicas.

Se ilustran los coeficientes de cada variable en la primera y segunda componente principal de los doce vectores.

Sea

$$CP_1 = e_{11} \cdot z_1 + e_{12} \cdot z_2 + e_{13} \cdot z_3 + \dots + e_{112} \cdot z_{12}$$

$$CP_2 = e_{21} \cdot z_1 + e_{22} \cdot z_2 + e_{23} \cdot z_3 + \dots + e_{212} \cdot z_{12}$$

donde,

CP_1 es la primera componente principal.

CP_2 es la segunda componente principal.

$z_1, z_2, z_3, \dots, z_p$ son las variables estandarizadas (asignaturas).

$e_{11}, e_{12}, e_{13}, \dots, e_{1p}$ y $e_{21}, e_{22}, e_{23}, \dots, e_{2p}$ son los coeficientes asociados a cada variable en la primera y segunda componente principal.

Al ser un análisis normado se tiene que $z = \frac{x-\mu}{\sigma}$

Según la tabla 5.3 así se observa las dos primeras componentes principales CP1 y CP2.

$$CP_1 = 0.2988 \cdot bio + 0.3140 \cdot cso + 0.2625 \cdot dib + 0.2955 \cdot eti + 0.2724 \cdot efi + 0.2752 \cdot rel + 0.3308 \cdot lca + 0.2454 \cdot ing + 0.3143 \cdot mat + 0.2846 \cdot tec + 0.2923 \cdot fil + 0.2664 \cdot emp$$

$$CP_2 = 0.2204 \cdot bio + 0.0814 \cdot cso + 0.4979 \cdot dib - 0.2636 \cdot eti - 0.2245 \cdot efi - 0.5488 \cdot rel - 0.0276 \cdot lca + 0.4682 \cdot ing + 0.1059 \cdot mat - 0.0089 \cdot tec - 0.0472 \cdot fil - 0.2058 \cdot emp$$

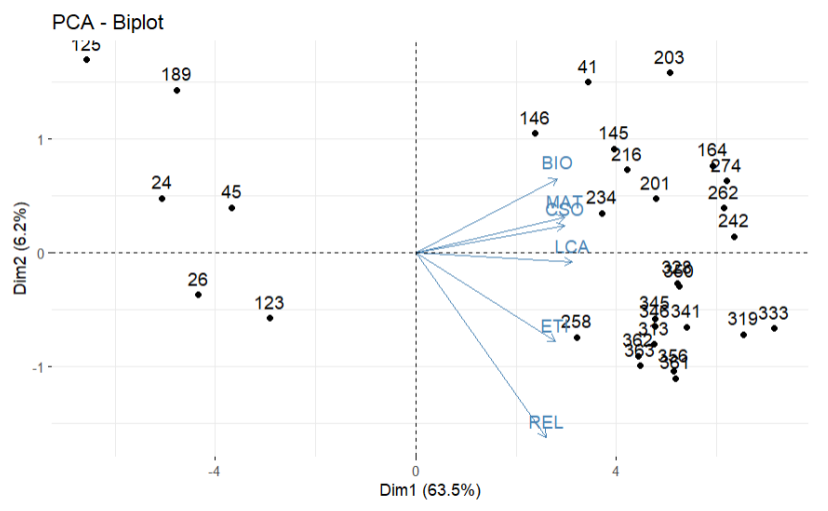


Figura 5.1: Relación entre las 6 mejores variables y las 30 mejores observaciones del Grado 6° (Fuente: Elaboración Propia)

En la figura 5.1 se observa que el primer eje, conocido como componente principal uno CP_1 , captura el 63.5% de la variabilidad total de los datos, mientras que el segundo eje, o componente principal dos CP_2 , abarca el 6.2%. La variable que más contribuye a la varianza retenida por CP_1 es lengua castellana, mientras que la variable que más contribuye a la varianza retenida por CP_2 es religión, como se muestra en el gráfico. Además, se visualizan observaciones que, independientemente del eje, pueden aparecer negativas en uno y positivas en otro. Si se reemplazan los datos de cada observación en la fórmula de CP_1 , se obtienen valores positivos, mientras que algunos valores en CP_2 son negativos debido a que se está trabajando con datos estandarizados.

En resumen, tanto lengua castellana para la CP_1 como religión para la CP_2 tienen contribuciones similares según los coeficientes asociados a cada componente.

5.2. Análisis de Componentes Principales (ACP) Grado 7°

A continuación, se presenta la base de datos que detalla los resultados académicos del grado séptimo, ofreciendo una visión detallada del rendimiento estudiantil en las materias cursadas durante cada año. La unión de los datos de estos tres años en una sola base se debe a que el currículo no varía según el grado, aunque las temáticas abordadas en las diferentes materias pueden cambiar con el tiempo. Sin embargo, los resultados académicos mantienen unos promedios muy similares a lo largo de estos años.

Para el grado séptimo se incluyen 378 observaciones, como se muestra en la tabla 5.4. Utilizando el ACP en la mencionada base de datos, se evaluó el desempeño académico

de los estudiantes de grado séptimo en una institución educativa de Medellín durante el periodo de 2019 a 2021. En la figura 5.2 se presenta el biplot de los resultados del ACP con las 30 mejores observaciones y las seis variables de mayor calidad de contribución en el primer plano factorial, según el criterio de \cos^2 .

Tabla 5.4: Calificaciones obtenidas por los estudiantes de grado 7° durante los años escolares 2019, 2020 y 2021 en una institución de Medellín.

BIO: Biología			ETI: Ética			LCA: Lengua Castellana			TEC: Tecnología		
CSO: Sociales			EFI: Educación Física			ING: Inglés			FIL: Filosofía		
DIB: Dibujo			REL: Religión			MAT: Matemáticas			EMP: Emprendimiento		
BIO	CSO	DIB	ETI	EFI	REL	LCA	ING	MAT	TEC	FIL	EMP
3.8	3.8	3.6	4.1	4.1	3.7	3.8	3.9	3.7	4.1	4	4.3
3.6	3.5	3.8	3.6	4.1	3.5	3.5	3	3.5	4	3.8	4.2
3.5	3.6	3.7	3.7	4.2	4	3.9	3.8	3.5	4.1	4	4.3
3.8	3.9	4.1	3.9	4.1	3.8	4.3	4.1	3.9	4.6	4	4.3
4.1	4.2	4.1	3.9	4.2	4.1	4.2	4	3.7	4.3	4	4.5
.
.
.

Tabla 5.5: Tabla de Valores Propios Grado 7°

CP	Valor. Propio	Var. Explicada	Acumulado
CP1	8.0159368	66.79947	66.79947
CP2	0.9249334	7.707779	74.50725
CP3	0.5607606	4.673005	79.18026
CP4	0.4506721	3.755601	82.93586
CP5	0.3633823	3.028186	85.96404
CP6	0.3300906	2.750755	88.71480

Al observar la tabla de valores propios del grado 7°, se puede notar que las primeras tres componentes (CP1, CP2, CP3) explican aproximadamente el 79.18% de la varianza total, mientras que las seis componentes explican alrededor del 88.71% ver tabla 5.5.

Tabla 5.6: Tabla de vectores propios de la matriz de correlación grado 7°

Asignaturas	CP1	CP2
bio	0.3015	-0.1305
cso	0.3169	-0.1800
dib	0.2831	0.1724
eti	0.2779	-0.4202
efi	0.2793	0.2222
rel	0.2977	-0.3088
lca	0.2916	0.1294
ing	0.2729	-0.1518
mat	0.3076	-0.2264
tec	0.2886	0.3891
fil	0.2892	0.0161
emp	0.2521	0.5973

Al analizar la tabla 5.6, se aprecia que las asignaturas que más contribuyen a CP1 son sociales, matemáticas, biología y religión. Esto sugiere que este componente está relacionado con las habilidades académicas fundamentales y generales de los estudiantes.

Por otro lado, las asignaturas que más contribuyen a CP2 (en ambas direcciones) son ética y emprendimiento. Este componente podría estar capturando la variabilidad relacionada con las habilidades éticas y emprendedoras de los estudiantes, lo que indica una dimensión diferente en el rendimiento académico.

Según la tabla 5.6 así se observa las dos primeras componentes principales CP1 y CP2.

$$CP_1 = 0.3015 \cdot bio + 0.3169 \cdot cso + 0.2831 \cdot dib + 0.2779 \cdot eti + 0.2793 \cdot efi + 0.2977 \cdot rel + 0.2916 \cdot lca + 0.2729 \cdot ing + 0.3076 \cdot mat + 0.2886 \cdot tec + 0.2892 \cdot fil + 0.2521 \cdot emp$$

$$CP_2 = -0.1305 \cdot bio - 0.1800 \cdot cso + 0.1724 \cdot dib - 0.4202 \cdot eti + 0.2222 \cdot efi - 0.3088 \cdot rel + 0.1294 \cdot lca - 0.1518 \cdot ing - 0.2264 \cdot mat + 0.3891 \cdot tec + 0.0161 \cdot fil + 0.5973 \cdot emp$$

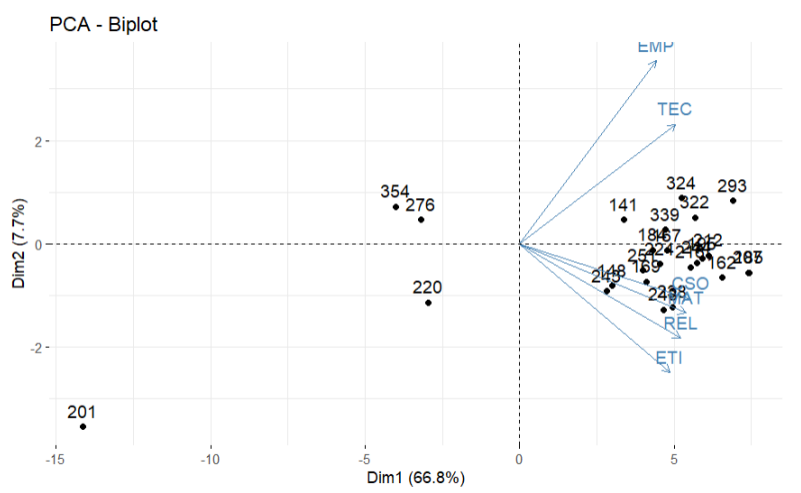


Figura 5.2: Relación entre las 6 mejores variables y las 30 mejores observaciones del Grado 7° (Fuente: Elaboración Propia)

En la figura 5.2 se observa que el primer eje, conocido como CP1, revela la captura del 66.8% de la variabilidad total de los datos, mientras que el segundo eje, o Componente dos CP2, abarca el 7.7%. Se observa que la variable que más contribuye a la varianza retenida por la CP1 es sociales, mientras que la variable que más contribuye a la varianza retenida por la CP2 es emprendimiento. Además, se visualizan observaciones que, independientemente del eje, pueden aparecer negativas en uno y positivas en otro. Si se reemplazan los datos de cada observación en la fórmula de CP1, se obtienen valores positivos, mientras que algunos valores en CP2 son negativos debido a que se está trabajando con datos estandarizados

En resumen, tanto sociales para la CP1 como emprendimiento para la CP2 tienen contribuciones similares según los coeficientes asociados a cada componente.

5.3. Análisis de Componentes Principales (ACP) Grado 8°

A continuación, se presenta la base de datos que detalla los resultados académicos del grado octavo, ofreciendo una visión detallada del rendimiento estudiantil en las materias

cursadas durante cada año. La unión de los datos de estos tres años en una sola base se debe a que el currículo no varía según el grado, aunque las temáticas abordadas en las diferentes materias pueden cambiar con el tiempo. Sin embargo, los resultados académicos mantienen unos promedios muy similares a lo largo de estos años.

Para el grado octavo se incluyen 391 observaciones, como se muestra en la tabla 5.7. Utilizando el Análisis de Componentes Principales (ACP) en la mencionada base de datos, se evaluó el desempeño académico de los estudiantes de grado octavo en una institución educativa de Medellín durante el periodo de 2019 a 2021. En la figura 5.3 se presenta el biplot de los resultados del ACP con las 30 mejores observaciones y las seis variables de mayor calidad de contribución en el primer plano factorial, según el criterio de \cos^2 .

Tabla 5.7: Calificaciones obtenidas por los estudiantes de grado 8° durante los años escolares 2019, 2020 y 2021 en una institución de Medellín.

FIS: Física				DIB: Dibujo				LCA: Lengua Castellana			FIL: Filosofía		
QUI: Química				ETI: Ética				ING: Ingles			EMP: Emprendimiento		
BIO: Biología				EFI: Educación Física				MAT: Matemáticas					
CSO: Sociales				REL: Religión				TEC: Tecnología					
FIS	QUI	BIO	CSO	DIB	ETI	EFI	REL	LCA	ING	MAT	TEC	FIL	EMP
4	4	4	4	4,1	3,8	4,1	4	4,1	4,5	4,2	4,1	4,2	4,3
4,1	3,6	3,9	3,9	4,5	3,8	4,1	4,4	4	3,9	3,7	3,9	4,4	4,3
4	3,9	4,1	4	4,1	3,8	4,2	4,4	4,2	4,6	4,4	4,3	4,2	4,4
4	3,9	3,8	3,7	4,2	3,7	4,2	4,5	4	4,4	4,4	4,1	4	3,7
4,2	3,9	4,3	3,8	3,8	4,1	4,4	4,5	4,2	4,6	4,3	4	4,3	4,1
.
.
.

Tabla 5.8: Tabla de Valores Propios Grado 8°

CP	Valor. Propio	Var. Explicada	Acumulado
CP1	8.9620729	64.01481	64.01481
CP2	0.9694139	6.924385	70.93919
CP3	0.7249643	5.178316	76.11751
CP4	0.5817875	4.155625	80.27313
CP5	0.4718219	3.370157	83.64329
CP6	0.4382476	3.13034	86.77363

Al observar la tabla de valores propios del grado 8°, se puede notar que las primeras tres componentes (CP1, CP2, CP3) explican aproximadamente el 76.11 % de la varianza total, mientras que las seis componentes explican alrededor del 86.77 % ver tabla 5.8.

Tabla 5.9: Tabla de vectores propios de la matriz de correlación grado 8°

Asignaturas	CP1	CP2
fis	0.2954	-0.0621
quim	0.2917	0.0692
bio	0.2964	0.0851
cso	0.2703	0.0178
dib	0.2491	-0.0644
eti	0.2678	0.0431
efi	0.2354	-0.3024
rel	0.2700	-0.2185
lca	0.2725	-0.3257
ing	0.2406	0.1018
mat	0.2815	-0.3468
tec	0.2662	0.4661
fil	0.2646	-0.0052
emp	0.2290	0.6202

Al analizar la tabla 5.9, se observa que las asignaturas que más contribuyen a CP1 son física, biología, lengua castellana y matemáticas. Esto sugiere que este componente está relacionado con las habilidades académicas fundamentales y generales de los estudiantes. Por otro lado, las asignaturas que más contribuyen a CP2 (en ambas direcciones) son tecnología y emprendimiento. Este componente podría estar capturando la variabilidad relacionada con habilidades técnicas y emprendedoras, lo que indica una dimensión diferente en el rendimiento académico.

Según la tabla 5.9 así se observa las dos primeras componentes principales CP1 y CP2.

$$CP_1 = 0.2954 \cdot fis + 0.2917 \cdot quim + 0.2964 \cdot bio + 0.2703 \cdot cso + 0.2491 \cdot dib + 0.2678 \cdot eti + 0.2354 \cdot efi + 0.2700 \cdot rel + 0.2725 \cdot lca + 0.2406 \cdot ing + 0.2815 \cdot mat + 0.2662 \cdot tec + 0.2646 \cdot fil + 0.2290 \cdot emp$$

$$CP_2 = -0.0621 \cdot fis + 0.0692 \cdot quim + 0.0851 \cdot bio + 0.0178 \cdot cso - 0.0644 \cdot dib + 0.0431 \cdot eti - 0.3024 \cdot rel - 0.2185 \cdot lca - 0.3257 \cdot ing + 0.1018 \cdot ing - 0.3468 \cdot mat + 0.4661 \cdot tec - 0.0052 \cdot fil + 0.6202 \cdot emp$$

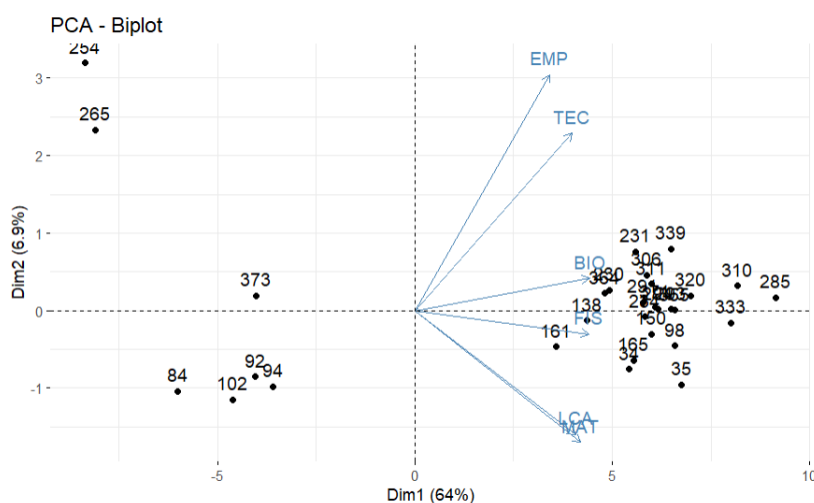


Figura 5.3: Relación entre las 6 mejores variables y las 30 mejores observaciones del Grado 8° (Fuente: Elaboración Propia)

En la figura 5.3 se observa que el primer eje, conocido como CP1, revela la captura del 64 % de la variabilidad total de los datos, mientras que el segundo eje, o Componente dos CP2, abarca el 6.9 %. Se observa que la variable que más contribuye a la varianza retenida por la CP1 es biología, mientras que la variable que más contribuye a la varianza retenida por la CP2 es emprendimiento. Además, se visualizan observaciones que, independientemente del eje, pueden aparecer negativas en uno y positivas en otro. Si se reemplazan los datos de cada observación en la fórmula de CP1, se obtienen valores positivos, mientras que algunos valores en CP2 son negativos debido a que se está trabajando con datos estandarizados.

En resumen, tanto biología para la componente uno CP1 como emprendimiento para la componente dos CP2 tienen contribuciones similares según los coeficientes asociados a cada componente.

5.4. Análisis de Componentes Principales (ACP) Grado 9°

A continuación, se presenta la base de datos que detalla los resultados académicos del grado noveno, ofreciendo una visión detallada del rendimiento estudiantil en las materias cursadas durante cada año. La unión de los datos de estos tres años en una sola base se debe a que el currículo no varía según el grado, aunque las temáticas abordadas en las diferentes materias pueden cambiar con el tiempo. Sin embargo, los resultados académicos mantienen unos promedios muy similares a lo largo de estos años

Para el grado noveno se incluyen 378 observaciones, como se muestra en la tabla 5.10. Utilizando el Análisis de Componentes Principales (ACP) en la mencionada base de datos, se evaluó el desempeño académico de los estudiantes de grado noveno en una institución educativa de Medellín durante el periodo de 2019 a 2021. En la figura 5.4 se presenta el biplot de los resultados del ACP con las 30 mejores observaciones y las seis variables de mayor calidad de contribución en el primer plano factorial, según el criterio de \cos^2 .

Tabla 5.10: Calificaciones obtenidas por los estudiantes de grado 9° durante los años escolares 2019, 2020 y 2021 en una institución de Medellín.

FIS: Física				DIB: Dibujo				LCA: Lengua Castellana			FIL: Filosofía		
QUI: Química				ETI: Ética				ING: Ingles			EMP: Emprendimiento		
BIO: Biología				EFI: Educación Física				MAT: Matemáticas					
CSO: Sociales				REL: Religión				TEC: Tecnología					
FIS	QUI	BIO	CSO	DIB	ETI	EFI	REL	LCA	ING	MAT	TEC	FIL	EMP
4	4,5	3,9	3,6	4,1	4,1	4	3,7	3,9	4,4	3,5	3,8	3,5	3,7
3,6	4	3,9	3,5	3,7	3,7	4,1	4,1	3,6	3,6	3,5	3,7	3,7	3,7
4,1	4,6	4,4	3,8	4,1	4,3	4,2	4,4	4,2	3,7	3,9	4,1	4,2	4,1
3,7	4,1	3,7	3,7	3,9	3,7	3,9	3,9	3,7	3,7	3,5	3,6	3,5	3,5
3,9	4,5	4,4	3,7	4,3	3,9	4,2	4,1	4,2	4,4	3,8	4	4,2	4,2
.
.
.

Al observar la tabla de valores propios del grado 9°, se puede notar que las primeras tres componentes (CP1, CP2, CP3) explican aproximadamente el 77.99 % de la varianza total, mientras que las seis componentes explican alrededor del 87.93 % ver tabla 5.11.

Tabla 5.11: Tabla de Valores Propios Grado 9°

CP	Valor. Propio	Var. Explicada	Acumulado
CP1	8.7873974	62.767120	62.76712
CP2	1.4515304	10.368074	73.13520
CP3	0.6809342	4.863816	77.99901
CP4	0.5740154	4.100110	82.09912
CP5	0.4443817	3.174155	85.27328
CP6	0.3720574	2.657553	87.93083

Tabla 5.12: Tabla de vectores propios de la matriz de correlación grado 9°

Asignaturas	CP1	CP2
fis	0.2914	0.1345
quim	0.2576	0.3712
bio	0.2811	0.2209
cso	0.2499	-0.4412
dib	0.2273	0.3111
eti	0.2606	-0.4386
efi	0.2640	0.1473
rel	0.2804	-0.1698
lca	0.2656	0.3606
ing	0.2055	-0.0634
mat	0.2998	-0.0924
tec	0.2790	-0.2100
fil	0.2804	0.1313
emp	0.2833	-0.2388

Al analizar la tabla 5.12, se observa que en la CP1, las asignaturas de matemáticas, física y emprendimiento tienen las contribuciones más significativas, lo que sugiere que estos temas son fundamentales para explicar la variabilidad general en el rendimiento académico de los estudiantes en el grado 9°. Por otro lado, en CP2, las asignaturas de sociales, lengua castellana y química emergen como las principales contribuyentes, lo que indica que estas áreas del currículo tienen un impacto considerable en la variabilidad observada en este componente.

Según la tabla 5.12 así se observa las dos primeras componentes principales CP1 y CP2.

$$CP_1 = 0.2914 \cdot fis + 0.2576 \cdot quim + 0.2811 \cdot bio + 0.2499 \cdot cso + 0.2273 \cdot dib + 0.2606 \cdot eti + 0.2640 \cdot efi + 0.2804 \cdot rel + 0.2656 \cdot lca + 0.2055 \cdot ing + 0.2998 \cdot mat + 0.2790 \cdot tec + 0.2804 \cdot fil + 0.2833 \cdot emp$$

$$CP_2 = 0.1345 \cdot fis + 0.3712 \cdot quim + 0.2209 \cdot bio - 0.4412 \cdot cso + 0.3111 \cdot dib - 0.4386 \cdot eti + 0.1473 \cdot efi - 0.1698 \cdot rel + 0.3606 \cdot lca - 0.0634 \cdot ing - 0.0924 \cdot mat - 0.2100 \cdot tec + 0.1313 \cdot fil - 0.2388 \cdot emp$$

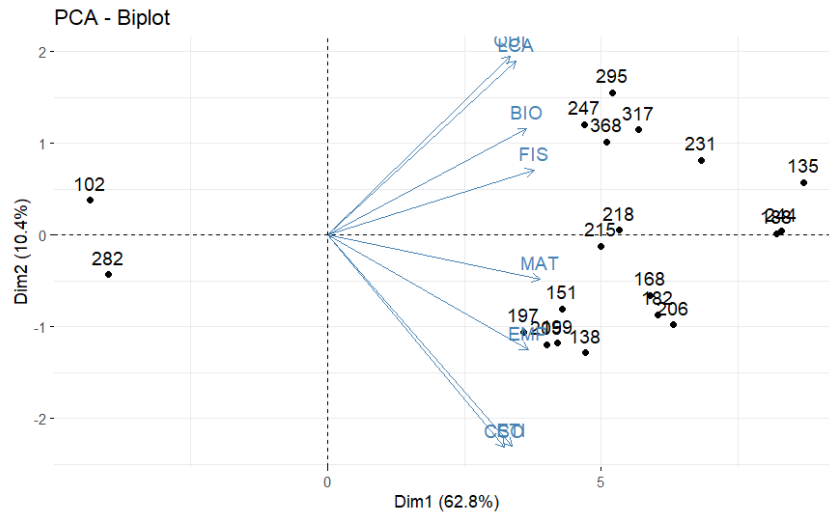


Figura 5.4: Relación entre las 6 mejores variables y las 30 mejores observaciones del grado 9° (Fuente: Elaboración Propia)

En la figura 5.3 se observa que el primer eje, conocido como Componente Principal uno CP1, revela la captura del 62.8% de la variabilidad total de los datos, mientras que el segundo eje, o Componente dos CP2, abarca el 10.4%. Se observa que la variable que más contribuye a la varianza retenida por la CP1 es matemáticas, mientras que la variable que más contribuye a la varianza retenida por la CP2 es sociales. Además, se visualizan observaciones que, independientemente del eje, pueden aparecer negativas en uno y positivas en otro. Si se reemplazan los datos de cada observación en la fórmula de CP1, se obtienen valores positivos, mientras que algunos valores en CP2 son negativos debido a que se está trabajando con datos estandarizados.

En resumen, tanto matemáticas para la componente uno CP1 como sociales para la componente dos CP2 tienen contribuciones similares según los coeficientes asociados a cada componente.

5.5. Análisis de Componentes Principales (ACP) Grado 10°

A continuación, se presenta la base de datos que detalla los resultados académicos del grado decimo, ofreciendo una visión detallada del rendimiento estudiantil en las materias cursadas durante cada año. La unión de los datos de estos tres años en una sola base se debe a que el currículo no varía según el grado, aunque las temáticas abordadas en las diferentes materias pueden cambiar con el tiempo. Sin embargo, los resultados académicos mantienen unos promedios muy similares a lo largo de estos años.

Para el grado décimo se incluyen 364 observaciones, como se muestra en la tabla 5.13. Utilizando el Análisis de Componentes Principales (ACP) en la mencionada base de datos, se evaluó el desempeño académico de los estudiantes de grado decimo en una institución educativa de Medellín durante el periodo de 2019 a 2021. En la figura 5.5 se presenta el biplot de los resultados del ACP con las 30 mejores observaciones y las seis variables de mayor calidad de contribución en el primer plano factorial, según el criterio de \cos^2 .

Tabla 5.13: Calificaciones obtenidas por los estudiantes de grado 10° durante los años escolares 2019, 2020 y 2021 en una institución de Medellín.

FIS: Física				DIB: Dibujo				LCA: Lengua Castellana			FIL: Filosofía		
QUI: Química				ETI: Ética				ING: Ingles			CEP: Ciencia Política		
BIO: Biología				EFI: Educación Física				TRI: Trigonometría					
CSO: Sociales				REL: Religión				TEC: Tecnología					
FIS	QUI	BIO	CSO	DIB	ETI	EFI	REL	LCA	ING	TRI	TEC	FIL	CEP
3,6	3,7	4,4	3,6	4,2	3,9	4,1	3,6	4,1	4,0	3,7	4,3	3,6	3,6
3,3	3,4	4,0	3,6	3,6	2,8	3,9	2,9	3,3	3,8	3,5	3,8	3,6	3,6
3,9	3,7	4,3	3,5	3,7	4,1	4,2	3,9	4,0	4,5	3,6	4,5	4,2	3,7
4,5	4,1	4,5	4,0	4,5	4,4	4,3	4,3	4,4	4,3	4,2	4,5	4,2	4,2
3,5	3,6	4,4	3,8	3,8	3,8	3,9	3,5	3,8	3,6	3,5	4,0	3,9	3,7
.
.
.

Tabla 5.14: Tabla de Valores Propios Grado 10°

CP	Valor. Propio	Var. Explicada	Acumulado
CP1	9.4041971	67.17284	67.17284
CP2	0.8560826	6.114876	73.28771
CP3	0.6644330	4.745950	78.03366
CP4	0.5797719	4.141228	82.17489
CP5	0.4048637	2.891884	85.06677
CP6	0.3826217	2.733012	87.79979

Al observar la tabla de valores propios del grado 10°, se puede notar que las primeras tres componentes (CP1, CP2, CP3) explican aproximadamente el 78.03 % de la varianza total, mientras que las seis componentes explican alrededor del 87.79 % ver tabla 5.14.

Tabla 5.15: Tabla de vectores propios de la matriz de correlación grado 10°

Asignaturas	CP1	CP2
fis	0.2922	0.1231
quim	0.2846	0.1875
bio	0.2664	-0.2525
cso	0.2831	-0.2662
dib	0.2622	-0.0698
eti	0.2573	-0.5685
efi	0.2100	0.0873
rel	0.2670	0.3075
lca	0.2790	-0.1685
ing	0.2086	0.3177
tri	0.2851	0.2444
tec	0.2687	0.3957
fil	0.2866	0.1197
cep	0.2743	-0.4091

Al analizar la tabla 5.15, se observa que las asignaturas de física, filosofía, química, trigonometría y sociales son las asignaturas que más contribuyen a la variabilidad en CP1, indicando que estas materias tienen una alta variabilidad en las calificaciones de los estudiantes de décimo grado. Por otro lado, tecnología, inglés, trigonometría, religión y ética son las asignaturas que más contribuyen a la variabilidad en CP2.

Según la tabla 5.15 así se observa las dos primeras componentes principales CP1 y CP2.

$$CP_1 = 0.2922 \cdot fis + 0.2846 \cdot quim + 0.2664 \cdot bio + 0.2831 \cdot cso + 0.2622 \cdot dib + 0.2573 \cdot eti + 0.2100 \cdot efi + 0.2670 \cdot rel + 0.2790 \cdot lca + 0.2086 \cdot ing + 0.2851 \cdot tri + 0.2687 \cdot tec + 0.2866 \cdot fil + 0.2743 \cdot cep$$

$$CP_2 = 0.115 \cdot fis - 0.175 \cdot quim + 0.236 \cdot bio + 0.249 \cdot cso + 0.065 \cdot dib + 0.532 \cdot eti - 0.082 \cdot efi - 0.288 \cdot rel + 0.158 \cdot lca - 0.297 \cdot ing - 0.229 \cdot tri + 0.370 \cdot tec - 0.112 \cdot fil + 0.383 \cdot emp$$

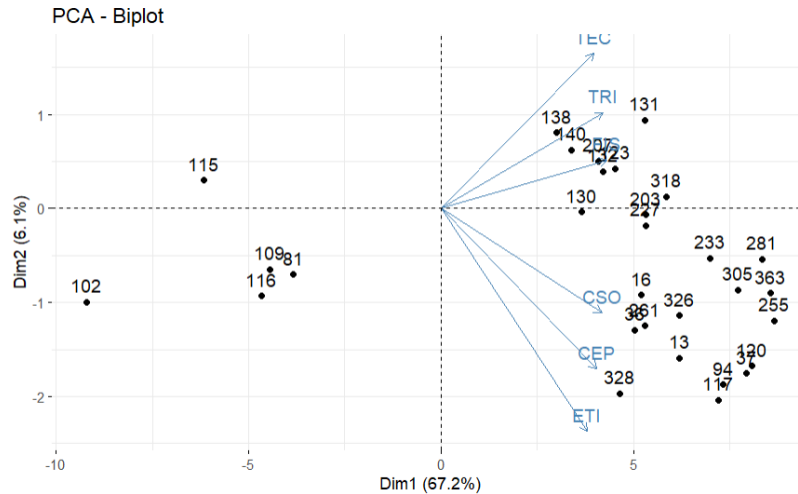


Figura 5.5: Relación entre las 6 mejores variables y las 30 mejores observaciones del grado 10° (Fuente: Elaboración Propia)

En la figura 5.5 se observa que el primer eje, conocido como Componente Principal uno CP1, revela la captura del 67.2% de la variabilidad total de los datos, mientras que el segundo eje, o Componente dos CP2, abarca el 6.1%. Se observa que la variable que más contribuye a la varianza retenida por la CP1 es física, mientras que la variable que más contribuye a la varianza retenida por la CP2 es ética. Además, se visualizan observaciones que, independientemente del eje, pueden aparecer negativas en uno y positivas en otro. Si se reemplazan los datos de cada observación en la fórmula de CP1, se obtienen valores positivos, mientras que algunos valores en CP2 son negativos debido a que se está trabajando con datos estandarizados.

En resumen, tanto física para la componente uno CP1 como ética para la componente dos CP2 tienen contribuciones similares según los coeficientes asociados a cada componente

5.6. Análisis de Componentes Principales (ACP) Grado 11°

En este contexto, se llevarán a cabo tres análisis de Componentes Principales (ACP) con el objetivo de trabajar en la mejora de los resultados de los estudiantes de grado undécimo en las Pruebas Saber y asegurar la continuidad de buenos desempeños académicos. Estos

análisis permitirán una representación detallada de los datos y ofrecerán una visión clara de las tendencias y patrones a lo largo del tiempo. Para cada uno de los análisis de ACP, se representará un año específico en el gráfico principal, mientras que los datos de los otros dos años se considerarán como individuos suplementarios. Esta metodología permitirá evaluar la consistencia y relevancia de las observaciones de los años suplementarios en relación con el año activo representado en el gráfico.

La inclusión de estas observaciones suplementarias en el análisis es crucial para determinar si es justificable considerarlas como activas. Al hacerlo, podremos evaluar si las tendencias observadas son consistentes a lo largo de los años y si las estrategias implementadas han tenido un impacto sostenido en los resultados de las pruebas Saber para los estudiantes de grado undécimo. Esta información es vital para realizar ajustes y mejoras continuas en las estrategias educativas.

El análisis de componentes principales es una herramienta poderosa que nos permitirá descomponer la variabilidad de los datos en componentes principales, facilitando la identificación de los factores que más influyen en los resultados de las pruebas. Al identificar estos factores clave, podemos diseñar intervenciones específicas y focalizadas para mejorar el rendimiento académico de los estudiantes de grado undécimo. En particular, al centrarnos en este grupo clave de estudiantes, podemos implementar estrategias educativas que aborden sus necesidades específicas y fortalezcan las áreas en las que se requiere mayor atención. Esto no solo contribuirá a mejorar sus resultados en las pruebas saber, sino que también preparará mejor a los estudiantes para sus futuros desafíos académicos y profesionales.

En resumen, estos análisis no solo nos proporcionarán una visión profunda de los datos históricos, sino que también serán fundamentales para seguir cosechando buenos resultados en las pruebas saber. La capacidad de identificar, entender y actuar sobre las tendencias y patrones educativos nos permitirá mantener y mejorar la calidad de la educación, asegurando que los estudiantes de grado undécimo continúen alcanzando altos niveles de desempeño académico y estén bien preparados para sus futuras etapas educativas.

ACP grado 11° año 2019

Para el grado undécimo, se incluyen 125 observaciones, de las cuales 123 corresponden a los resultados del año 2019, y las observaciones 124 y 125 corresponden a los promedios de los años 2020 y 2021, respectivamente. En la figura 5.6, se presenta el biplot de los resultados del ACP con las 30 mejores observaciones y las seis variables de mayor calidad de contribución en el primer plano factorial, según el criterio de \cos^2 .

Tabla 5.16: Tabla de Valores Propios Grado 11° 2019

CP	Valor. Propio	Var. Explicada	Acumulado
CP1	9,5396633	68,140452	68,14045
CP2	0,9253176	6,609412	74,74986
CP3	0,6663420	4,759586	79,50945
CP4	0,5422591	3,873279	83,38273
CP5	0,4684742	3,346245	86,72897
CP6	0,3341888	2,387063	89,11604

Al observar la tabla de valores propios del grado 11°, se puede notar que las primeras tres componentes (CP1, CP2, CP3) explican aproximadamente el 79.50 % de la varianza total, mientras que las seis componentes explican alrededor del 89.11 % ver tabla 5.16.

Tabla 5.17: Tabla de vectores propios de la matriz de correlación grado 11° 2019

Asignaturas	CP1	CP2
fis	0.2925	0.0313
quim	0.2613	0.0421
bio	0.2536	-0.1803
cso	0.2884	0.0044
dib	0.2678	-0.1887
eti	0.2733	-0.0627
efi	0.1525	0.8526
rel	0.2874	-0.0440
lca	0.2869	-0.2000
ing	0.2468	0.2974
cal	0.2771	0.1013
tec	0.2524	-0.0345
fil	0.2796	-0.2367
cep	0.2907	0.0186

Al analizar la tabla 5.17, se observa que física, ciencia política, sociales, religión y lengua castellana son las asignaturas que más contribuyen a la variabilidad en CP1, indicando que estas materias tienen una alta variabilidad en las calificaciones de los estudiantes de undécimo grado. Por otro lado, educación física, inglés, y filosofía son las asignaturas que más contribuyen a la variabilidad en CP2, con educación física destacándose de manera significativa.

Según la tabla 5.17 así se observa las dos primeras componentes principales CP1 y CP2.

$$CP_1 = 0.2925 \cdot fis + 0.2613 \cdot quim + 0.2536 \cdot bio + 0.2884 \cdot cso + 0.2678 \cdot dib + 0.2733 \cdot eti + 0.1525 \cdot efi + 0.2874 \cdot rel + 0.2869 \cdot lca + 0.2468 \cdot ing + 0.2771 \cdot cal + 0.2524 \cdot tec + 0.2796 \cdot fil + 0.2907 \cdot cep$$

$$CP_2 = 0.0313 \cdot fis + 0.0421 \cdot quim - 0.1803 \cdot bio + 0.0044 \cdot cso - 0.1887 \cdot dib - 0.0627 \cdot eti + 0.8526 \cdot efi - 0.0440 \cdot rel - 0.2000 \cdot lca + 0.2974 \cdot ing + 0.1013 \cdot cal - 0.0345 \cdot tec - 0.2367 \cdot fil + 0.0186 \cdot cep$$

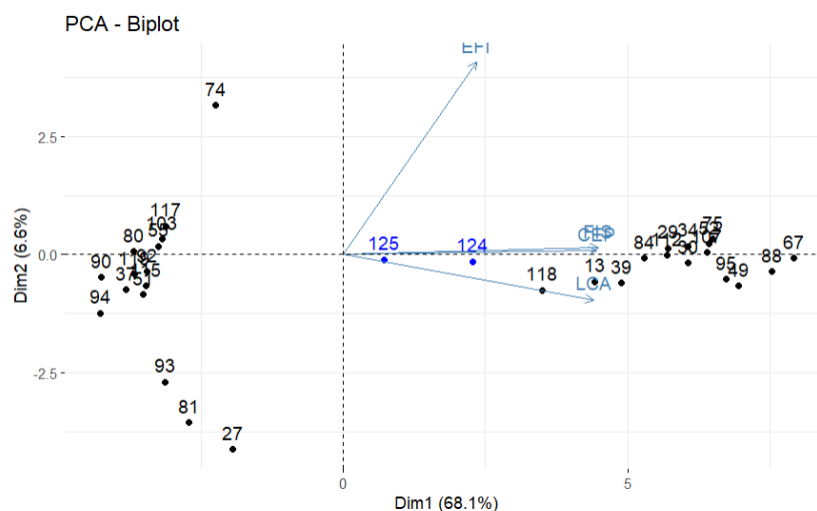


Figura 5.6: Relación entre las 6 mejores variables y las 30 mejores observaciones del grado 11° 2019 (Fuente: Elaboración Propia)

El primer eje, conocido como Componente Principal uno CP1, revela la captura del 68.9% de la variabilidad total de los datos, mientras que el segundo eje, o Componente dos CP2, abarca el 6.6%. Se observa que la variable que más contribuye a la varianza retenida por la CP1 es física, mientras que la variable que más contribuye a la varianza retenida por la CP2 es educación física, como se muestra en la figura 5.6.

En resumen, tanto física para la componente uno CP1 como educación física para la componente dos CP2 tienen contribuciones similares según los coeficientes asociados a cada componente.

Por último, en la figura 5.6 se incluyeron las observaciones 124 y 125, correspondientes a los promedios de notas del 2020 y 2021, respectivamente. Estas se agregaron como observaciones suplementarias en el análisis con el fin de evaluar si debían considerarse como activas. Sin embargo, tras observar que no ejercían una influencia significativa en los resultados, debido a que se encuentran cerca del centroide o se pueden considerar como individuos promedio.

ACP grado 11° año 2020

Para el grado undécimo, se incluyen 113 observaciones, de las cuales 111 corresponden a los resultados del año 2020, y las observaciones 112 y 113 corresponden a los promedios de los años 2019 y 2021, respectivamente. En la figura 5.7, se presenta el biplot de los resultados del ACP con las 30 mejores observaciones y las seis variables de mayor calidad de contribución en el primer plano factorial, según el criterio de \cos^2 .

Tabla 5.18: Tabla de Valores Propios Grado 11° 2020

CP	Valor. Propio	Var. Explicada	Acumulado
CP1	9,6640488	69,02892	69,02892
CP2	0,9403753	6,716967	75,74589
CP3	0,5835896	4,168497	79,91438
CP4	0,5158165	3,684404	83,59879
CP5	0,4811949	3,437107	87,03589
CP6	0,3327754	2,376967	89,41286

Al observar la tabla de valores propios del grado 11°, se puede notar que las primeras tres componentes (CP1, CP2, CP3) explican aproximadamente el 79.91% de la varianza total, mientras que las seis componentes explican alrededor del 89.41% ver tabla 5.18.

Al analizar la tabla 5.19, se observa que física, religión, lengua castellana, ciencia política, sociales y química son las asignaturas que más contribuyen a la variabilidad en CP1, indicando que estas materias tienen una alta variabilidad en las calificaciones de los estudiantes de undécimo grado. Por otro lado, educación física, inglés y filosofía son las asignaturas que más contribuyen a la variabilidad en CP2, con educación física destacándose de manera significativa.

Tabla 5.19: Tabla de vectores propios de la matriz de correlación grado 11° 2020

Asignaturas	CP1	CP2
fis	0.2935	0.0017
quim	0.2807	0.0238
bio	0.2520	-0.2353
cso	0.2879	0.0101
dib	0.2638	-0.1921
eti	0.2726	-0.0727
efi	0.1458	0.9551
rel	0.2915	-0.0206
lca	0.2890	-0.1924
ing	0.2440	0.3522
cal	0.2739	0.0965
tec	0.2445	-0.0259
fil	0.2789	-0.2681
cep	0.2880	0.0374

Según la tabla 5.19 así se observa las dos primeras componentes principales CP1 y CP2.

$$CP_1 = 0.2935 \cdot fis + 0.2807 \cdot quim + 0.2520 \cdot bio + 0.2879 \cdot cso + 0.2638 \cdot dib + 0.2726 \cdot eti + 0.1458 \cdot efi + 0.2915 \cdot rel + 0.2890 \cdot lca + 0.2440 \cdot ing + 0.2739 \cdot cal + 0.2445 \cdot tec + 0.2789 \cdot fil + 0.2880 \cdot cep$$

$$CP_2 = 0.0017 \cdot fis + 0.0238 \cdot quim - 0.2353 \cdot bio + 0.0101 \cdot cso - 0.1921 \cdot dib - 0.0727 \cdot eti + 0.9551 \cdot efi - 0.0206 \cdot rel - 0.1924 \cdot lca + 0.3522 \cdot ing + 0.0965 \cdot cal - 0.0259 \cdot tec - 0.2681 \cdot fil + 0.0374 \cdot cep$$

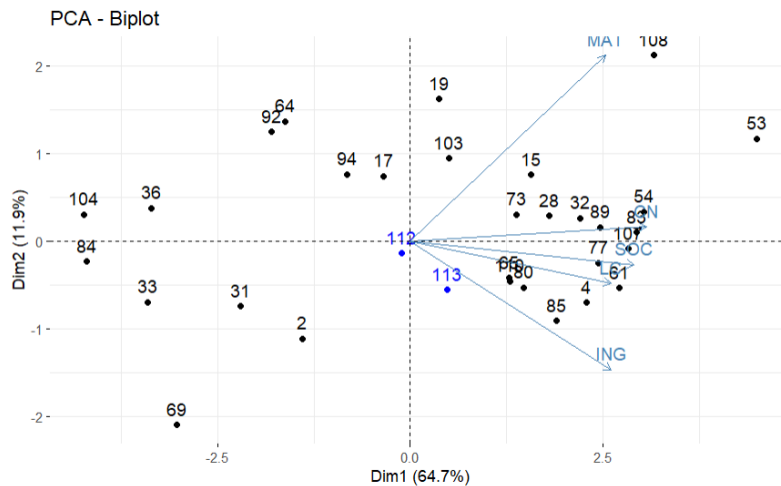


Figura 5.7: Relación entre las 6 mejores variables y las 30 mejores observaciones del grado 11° 2020 (Fuente: Elaboración Propia)

El primer eje, conocido como componente principal uno CP1, revela la captura del 64.7% de la variabilidad total de los datos, mientras que el segundo eje, o componente dos CP2, abarca el 11.9%. Se observa que la variable que más contribuye a la varianza retenida por la CP1 es física, mientras que la variable que más contribuye a la varianza retenida por la CP2 es educación física, como se muestra en la figura 5.7.

En resumen, tanto física para la componente uno CP1 como educación física para la

componente dos CP2 tienen contribuciones similares según los coeficientes asociados a cada componente.

Por último, en la figura 5.7 se incluyeron las observaciones 112 y 113, correspondientes a los promedios de notas del 2019 y 2021, respectivamente. Estas se agregaron como observaciones suplementarias en el análisis con el fin de evaluar si debían considerarse como activas. Sin embargo, tras observar que no ejercían una influencia significativa en los resultados, debido a que se encuentran cerca del centroide o se pueden considerar como individuos promedio.

ACP grado 11° año 2021

Para el grado undécimo, se incluyen 121 observaciones, de las cuales 119 corresponden a los resultados del año 2021, y las observaciones 120 y 121 corresponden a los promedios de los años 2019 y 2020, respectivamente. En la figura 5.8, se presenta el biplot de los resultados del ACP con las 30 mejores observaciones y las seis variables de mayor calidad de contribución en el primer plano factorial, según el criterio de \cos^2 .

Tabla 5.20: Tabla de Valores Propios Grado 11° 2021

CP	Valor. Propio	Var. Explicada	Acumulado
CP1	9,0453863	64,609902	64,6099
CP2	0,9297668	6,641191	71,25109
CP3	0,7632266	5,451619	76,70271
CP4	0,5814994	4,153567	80,85628
CP5	0,5262231	3,758736	84,61502
CP6	0,4702702	3,359073	87,97409

Al observar la tabla de valores propios del grado 11°, se puede notar que las primeras tres componentes (CP1, CP2, CP3) explican aproximadamente el 76.70% de la varianza total, mientras que las seis componentes explican alrededor del 87.97% ver tabla 5.20.

Tabla 5.21: Tabla de vectores propios de la matriz de correlación grado 11° 2021

Asignaturas	CP1	CP2
fis	0.3011	-0.0047
quim	0.2817	-0.0454
bio	0.2915	0.0064
cso	0.2694	-0.3997
dib	0.2070	0.4632
eti	0.2529	-0.5220
efi	0.2408	0.2271
rel	0.2674	0.0767
lca	0.2904	0.0288
ing	0.1989	0.2835
cal	0.2839	0.1732
tec	0.2642	0.1521
fil	0.2824	0.1252
cep	0.2864	-0.3789

Al analizar la tabla 5.21, se observa que física, biología, lengua castellana, ciencia política y cálculo son las asignaturas que más contribuyen a la variabilidad en CP1. Esto sugiere que estas materias tienen una alta variabilidad en las calificaciones de los estudiantes de undécimo grado. Por otro lado, dibujo, sociales, inglés y ciencia política

son las asignaturas que más influyen en la variabilidad en CP2, siendo dibujo destacada en este aspecto.

Según la tabla 5.21 así se observa las dos primeras componentes principales CP1 y CP2.

$$CP_1 = 0.3011 \cdot fis + 0.2817 \cdot quim + 0.2915 \cdot bio + 0.2694 \cdot cso + 0.2070 \cdot dib + 0.2529 \cdot eti + 0.2408 \cdot efi + 0.2674 \cdot rel + 0.2904 \cdot lca + 0.1989 \cdot ing + 0.2839 \cdot cal + 0.2642 \cdot tec + 0.2824 \cdot fil + 0.2864 \cdot cep$$

$$CP_2 = -0.0047 \cdot fis - 0.0454 \cdot quim + 0.0064 \cdot bio - 0.3997 \cdot cso + 0.4632 \cdot dib - 0.5220 \cdot eti + 0.2271 \cdot efi + 0.0767 \cdot rel + 0.0288 \cdot lca + 0.2835 \cdot ing + 0.1732 \cdot cal + 0.1521 \cdot tec + 0.1252 \cdot fil - 0.3789 \cdot cep$$

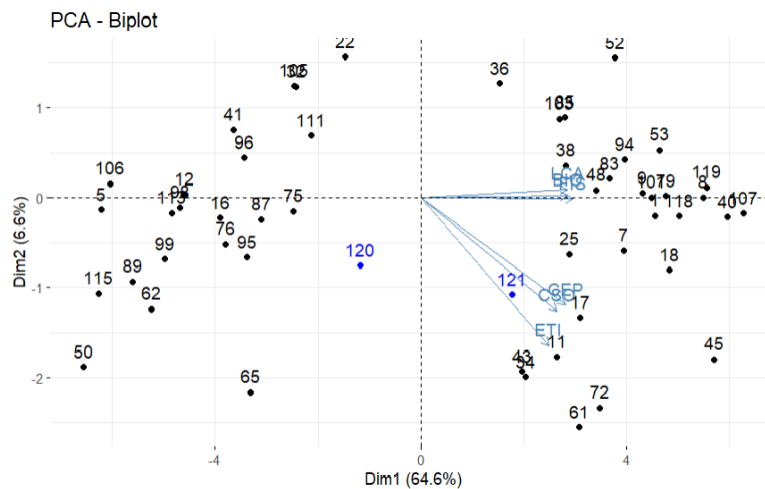


Figura 5.8: Relación entre las 6 mejores variables y las 30 mejores observaciones del grado 11° 2021 (Fuente: Elaboración Propia)

El primer eje, conocido como Componente Principal uno (CP1), revela la captura del 64.6% de la variabilidad total de los datos, mientras que el segundo eje, o Componente dos CP2, abarca el 6.6%. Se observa que la variable que más contribuye a la varianza retenida por la CP1 es física, mientras que la variable que más contribuye a la varianza retenida por la CP2 es ética, como se muestra en la figura 5.8.

En resumen, tanto física para la componente uno CP1 como ética para la componente dos CP2 tienen contribuciones similares según los coeficientes asociados a cada componente.

Por último, en la figura 5.8 se incluyeron las observaciones 120 y 121, correspondientes a los promedios de notas del 2019 y 2020, respectivamente. Estas se agregaron como observaciones suplementarias en el análisis con el fin de evaluar si debían considerarse como activas. Sin embargo, tras observar que no ejercían una influencia significativa en los resultados, debido a que se encuentran cerca del centroide o se pueden considerar como individuos promedio.

5.7. Análisis de Componentes Principales (ACP) Prueba Saber 11°

ACP Prueba Saber 11° 2019

A continuación, se presenta la base de datos que detalla los resultados de Pruebas Saber 11° del grado undécimo, ofreciendo una visión detallada del rendimiento estudiantil en las asignaturas evaluadas durante cada año.

Para el grado undécimo, se incluyen 125 observaciones, de las cuales 123 corresponden a los resultados del año 2019, y las observaciones 124 y 125 corresponden a los promedios de los años 2020 y 2021, respectivamente. En la figura 5.9, se presenta el biplot de los resultados del ACP con las 30 mejores observaciones y la calidad de contribución de las cinco variables en el primer plano factorial, según el criterio de \cos^2 .

Tabla 5.22: Tabla de Valores Propios Prueba Saber 11° 2019

CP	Valor. Propio	Var. Explicada	Acumulado
CP1	3,7404531	74,809062	74,80906
CP2	0,4533547	9,067094	83,87616
CP3	0,4018193	8,036385	91,91254
CP4	0,2302873	4,605745	96,51829
CP5	0,1740857	3,481713	100

Al observar la tabla de valores propios de las pruebas saber 11° 2019, se puede notar que las primeras tres componentes (CP1, CP2, CP3) explican aproximadamente el 91.91 % de la varianza total, mientras que las cinco componentes ya explican el 100 % ver tabla 5.22.

Tabla 5.23: Tabla de vectores propios de la matriz de correlación Prueba Saber 11° 2019

Asignaturas	CP1	CP2
<i>lca</i>	0.4329	-0.3148
<i>mat</i>	0.4510	-0.2768
<i>cso</i>	0.4730	-0.0213
<i>cn</i>	0.4645	-0.2010
<i>ing</i>	0.4119	0.8851

Al analizar la tabla 5.23, CP1 muestra una contribución positiva y significativa de asignaturas fundamentales como lengua castellana, matemáticas, sociales y ciencias naturales. Esto sugiere que CP1 captura un aspecto integral del rendimiento académico relacionado con habilidades básicas y conocimientos en áreas clave del currículo. Por otro lado, CP2 muestra una contribución positiva significativa de la asignatura de inglés. Esto sugiere que CP2 podría representar una dimensión específica del rendimiento académico relacionada con habilidades lingüísticas en inglés.

Según la tabla 5.23 así se observa las dos primeras componentes principales CP1 y CP2

$$CP_1 = 0.4329 \cdot lca + 0.4510 \cdot mat + 0.4730 \cdot cso + 0.4645 \cdot cn + 0.4119 \cdot ing$$

$$CP_2 = -0.3148 \cdot lca - 0.2768 \cdot mat - 0.0213 \cdot cso - 0.2010 \cdot cn + 0.8851 \cdot ing$$

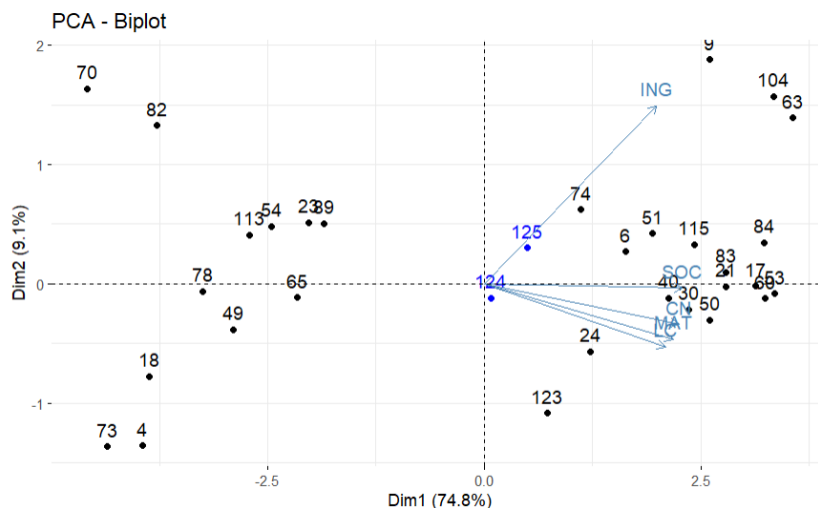


Figura 5.9: Relación entre las 5 mejores variables y las 30 mejores observaciones de la Prueba Saber 11° 2019 (Fuente: Elaboración Propia)

El primer eje, conocido como componente principal uno CP1, revela la captura del 74.8 % de la variabilidad total de los datos, mientras que el segundo eje, o componente dos CP2, abarca el 9.1 %. Se observa que la variable que más contribuye a la varianza retenida por la CP1 es sociales, mientras que la variable que más contribuye a la varianza retenida por la CP2 es inglés, como se muestra en la figura 5.9.

En resumen, tanto sociales para la componente uno CP1 como inglés para la componente dos CP2 tienen contribuciones similares según los coeficientes asociados a cada componente.

Por último, en la figura 5.9 se incluyeron las observaciones 124 y 125, correspondientes a los promedios de notas del 2020 y 2021, respectivamente. Estas se agregaron como observaciones suplementarias en el análisis con el fin de evaluar si debían considerarse como activas. Sin embargo, tras observar que no ejercían una influencia significativa en los resultados, debido a que se encuentran cerca del centroide o se pueden considerar como individuos promedio.

ACP Prueba Saber 11° 2020

A continuación, se presenta la base de datos que detalla los resultados de Pruebas Saber 11° del grado undécimo, ofreciendo una visión detallada del rendimiento estudiantil en las asignaturas evaluadas durante cada año.

Para el grado undécimo, se incluyen 113 observaciones, de las cuales 111 corresponden a los resultados del año 2020, y las observaciones 112 y 113 corresponden a los promedios de los años 2019 y 2021, respectivamente. En la figura 5.10, se presenta el biplot de los resultados del ACP con las 30 mejores observaciones y la calidad de contribución de las cinco variables en el primer plano factorial, según el criterio de \cos^2 .

Al observar la tabla de valores propios de las pruebas saber 11° 2020, se puede notar que las primeras dos componentes (CP1, CP2) explican aproximadamente el 87.05 % de la varianza total, mientras que las cinco componentes ya explican el 100 % ver tabla 5.24.

Tabla 5.24: Tabla de Valores Propios Prueba Saber 11° 2020

CP	Valor. Propio	Var. Explicada	Acumulado
CP1	3,2359691	64,719381	64,71938
CP2	0,5948562	11,897124	76,6165
CP3	0,5219867	10,439735	87,05624
CP4	0,4190371	8,380741	95,43698
CP5	0,228151	4,563019	100

Tabla 5.25: Tabla de vectores propios de la matriz de correlación grado 11° 2020

Asignaturas	CP1	CP2
lca	0.4242	-0.1820
mat	0.4116	0.8021
cso	0.4718	-0.1025
cn	0.4985	0.0623
ing	0.4237	-0.5560

Al analizar la tabla 5.25, CP1 muestra una contribución positiva y significativa de asignaturas fundamentales como lengua castellana, matemáticas, sociales, ciencias naturales e inglés. Esto sugiere que CP1 captura un aspecto integral del rendimiento académico relacionado con habilidades básicas y conocimientos en áreas clave del currículo. Por otro lado, CP2 muestra una contribución significativa y positiva en cálculo y negativa en la asignatura de inglés. Esto sugiere que CP2 podría representar una dimensión específica del rendimiento académico relacionada con habilidades lingüísticas en inglés y cuantitativas en matemáticas.

Según la tabla 5.25 así se observa las dos primeras componentes principales CP1 y CP2.

$$CP_1 = 0.4242 \cdot lca + 0.4116 \cdot mat + 0.4718 \cdot cso + 0.4985 \cdot cn + 0.4237 \cdot ing$$

$$CP_2 = -0.1820 \cdot lca + 0.8021 \cdot mat - 0.1025 \cdot cso + 0.0623 \cdot cn - 0.5560 \cdot ing$$

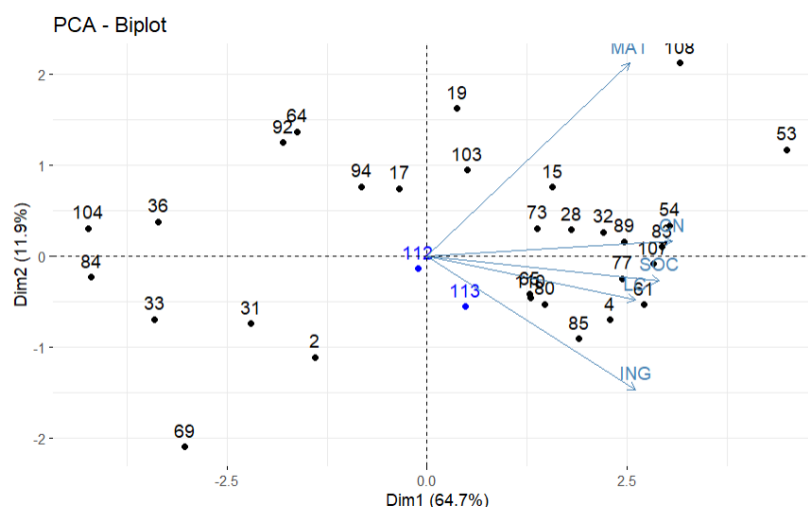


Figura 5.10: Relación entre las 5 mejores variables y las 30 mejores observaciones de la Prueba Saber 11° 2020 (Fuente: Elaboración Propia)

El primer eje, conocido como Componente Principal uno CP1, revela la captura del 64.7% de la variabilidad total de los datos, mientras que el segundo eje, o Componente dos CP2, abarca el 11.9%. Se observa que la variable que más contribuye a la varianza retenida por la CP1 es ciencias naturales, mientras que la variable que más contribuye a la varianza retenida por la CP2 es matemáticas, como se muestra en la figura 5.10.

En resumen, tanto ciencias naturales para la componente uno CP1 como matemáticas para la componente dos CP2 tienen contribuciones similares según los coeficientes asociados a cada componente.

Por último, en la figura 5.10 se incluyeron las observaciones 112 y 113, correspondientes a los promedios de notas del 2019 y 2021, respectivamente. Estas se agregaron como observaciones suplementarias en el análisis con el fin de evaluar si debían considerarse como activas. Sin embargo, tras observar que no ejercían una influencia significativa en los resultados, debido a que se encuentran cerca del centroide o se pueden considerar como individuos promedio.

ACP Prueba Saber 11° 2021

A continuación, se presenta la base de datos que detalla los resultados de Pruebas Saber 11° del grado undécimo, ofreciendo una visión detallada del rendimiento estudiantil en las asignaturas evaluadas durante cada año.

Para el grado undécimo, se incluyen 121 observaciones, de las cuales 119 corresponden a los resultados del año 2021, y las observaciones 120 y 121 corresponden a los promedios de los años 2019 y 2020, respectivamente. En la figura 5.11, se presenta el biplot de los resultados del ACP con las 30 mejores observaciones y la calidad de contribución de las cinco variables en el primer plano factorial, según el criterio de \cos^2 .

Tabla 5.26: Tabla de Valores Propios Prueba Saber 11° 2021

CP	Valor. Propio	Var. Explicada	Acumulado
CP1	3,5055534	70,111068	70,11107
CP2	0,6209017	12,418034	82,5291
CP3	0,4402601	8,805202	91,3343
CP4	0,2268873	4,537747	95,87205
CP5	0,2063975	4,12795	100

Al observar la tabla de valores propios de las pruebas saber 11° 2021, se puede notar que las primeras dos componentes (CP1, CP2) explican aproximadamente el 82.59% de la varianza total, mientras que las cinco componentes ya explican el 100% ver tabla 5.26.

Tabla 5.27: Tabla de vectores propios de la matriz de correlación grado 11° 2021

Asignaturas	CP1	CP2
lca	0.4713	0.1734
mat	0.4329	-0.4796
cso	0.4621	-0.1256
cn	0.4786	-0.2654
ing	0.3846	0.8085

Al analizar la tabla 5.27, CP1 muestra una contribución positiva y significativa de asignaturas fundamentales como lengua castellana, matemáticas, sociales, ciencias naturales e inglés. Esto sugiere que CP1 captura un aspecto integral del rendimiento académico relacionado con habilidades básicas y conocimientos en áreas clave del currículo. Por otro lado, CP2 muestra una contribución significativa y negativa en matemáticas y positiva en la asignatura de inglés. Esto sugiere que CP2 podría representar una dimensión específica del rendimiento académico relacionada con habilidades lingüísticas en inglés y habilidades cuantitativas en matemáticas.

Según la tabla 5.27 así se observa las dos primeras componentes principales CP1 y CP2.

$$CP_1 = 0.4713 \cdot lca + 0.4329 \cdot mat + 0.4621 \cdot cso + 0.4786 \cdot cn + 0.3846 \cdot ing$$

$$CP_2 = 0.1734 \cdot lca - 0.4796 \cdot mat - 0.1256 \cdot cso - 0.2654 \cdot cn + 0.8085 \cdot ing$$

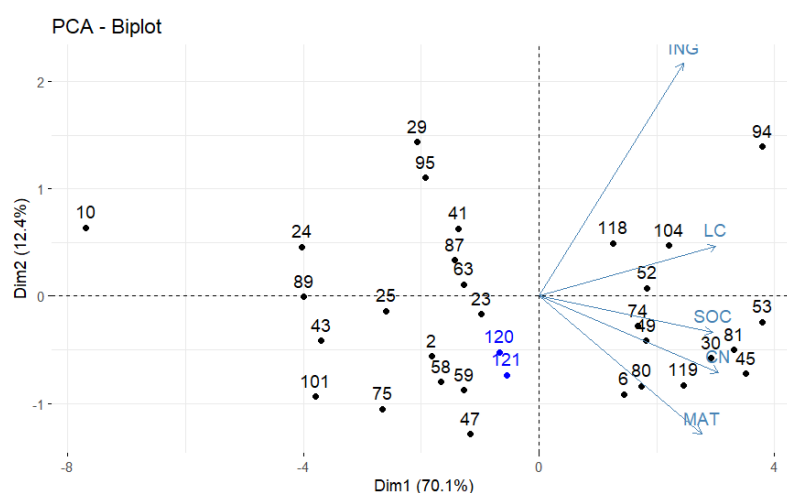


Figura 5.11: Relación entre las 5 mejores variables y las 30 mejores observaciones de la Prueba Saber 11° 2021 (Fuente: Elaboración Propia)

El primer eje, conocido como Componente Principal uno CP1, revela la captura del 70.1% de la variabilidad total de los datos, mientras que el segundo eje, o Componente dos CP2, abarca el 12.4%. Se observa que la variable que más contribuye a la varianza retenida por la CP1 es ciencias naturales, mientras que la variable que más contribuye a la varianza retenida por la CP2 es matemáticas, como se muestra en la figura 5.11.

En resumen, tanto ciencias naturales para la componente uno CP1 como matemáticas para la componente dos CP2 tienen contribuciones similares según los coeficientes asociados a cada componente.

Por último, en la figura 5.11 se incluyeron las observaciones 120 y 121, correspondientes a los promedios de notas del 2019 y 2020, respectivamente. Estas se agregaron como observaciones suplementarias en el análisis con el fin de evaluar si debían considerarse como activas. Sin embargo, tras observar que no ejercían una influencia significativa en los resultados, debido a que se encuentran cerca del centroide o se pueden considerar como individuos promedio.

Basándose en el análisis realizado, es notable observar que las contribuciones de las áreas en las componentes principales no se mantuvieron constantes a lo largo de los años,

e incluso, no conservaron el mismo signo entre un año y otro. Este fenómeno sugiere que hubo cambios significativos en el patrón de desempeño académico de los estudiantes de grado 11 en las pruebas Saber durante el período analizado. Tales fluctuaciones podrían atribuirse a una variedad de factores, incluyendo cambios en el currículo escolar, metodologías de enseñanza, niveles de preparación de los estudiantes, entre otros. Se hace necesario profundizar en el análisis de estos cambios para comprender mejor los factores que influyen en el rendimiento académico de los estudiantes y poder implementar estrategias efectivas para mejorar su desempeño en el futuro.

Capítulo 6

Conclusiones

Durante el análisis descriptivo y la aplicación del Análisis de Componentes Principales (ACP) en los grados sexto al undécimo, se emplearon diversas herramientas estadísticas para comprender en profundidad los procesos académicos en la institución educativa. A través de la construcción de gráficos y la interpretación de los resultados obtenidos, fue posible identificar los enfoques pedagógicos implementados, así como las estrategias metodológicas utilizadas por los docentes. Se evidenció, por ejemplo, el uso de plataformas tecnológicas y libros de texto en áreas clave del conocimiento, así como el impacto del trabajo en clase en el desempeño académico de los estudiantes.

La variedad de calificaciones obtenidas refleja la diversidad de enfoques educativos y la efectividad de las estrategias implementadas en el proceso de enseñanza-aprendizaje. Es evidente que el uso de estas herramientas y metodologías contribuye al desarrollo y preparación de los estudiantes para enfrentar con éxito la Prueba Saber 11° año tras año.

Al analizar detalladamente los resultados académicos de la institución en Medellín, se destaca la relevancia de la variabilidad observada en los resultados de las Pruebas Saber 11°. Esta variabilidad ofrece una base fundamental para futuras investigaciones y acciones de mejora educativa. Identificar las tendencias y comprender las razones subyacentes detrás de estos resultados es crucial para poder implementar estrategias específicas que fortalezcan áreas de oportunidad y promuevan el crecimiento académico continuo entre los estudiantes.

Además, estas observaciones brindan la oportunidad de extraer conclusiones adicionales que amplíen la comprensión de los resultados obtenidos. Esto puede incluir la identificación de factores influyentes en el desempeño académico, la evaluación de la efectividad de programas y políticas educativas, y el diseño de intervenciones específicas dirigidas a mejorar el rendimiento estudiantil y la calidad educativa en general.

6.1. Generales

Después de analizar detalladamente los resultados y construir los instrumentos del ACP, se pueden extraer las siguientes conclusiones generales.

El ACP nos ha permitido identificar patrones subyacentes en los datos, lo que nos ayuda a comprender mejor la estructura de la información y las relaciones entre las variables estudiadas.

Las componentes principales identificadas nos brindan una visión clara de las dimensiones más relevantes del rendimiento académico de los estudiantes. Estas dimensiones pueden estar relacionadas con habilidades específicas o áreas del conocimiento que influyen en el desempeño general.

Se observa que las contribuciones de las variables a las componentes principales pueden variar significativamente entre diferentes años o períodos de tiempo. Esto sugiere que existen cambios en el perfil de rendimiento de los estudiantes a lo largo del tiempo, lo cual es importante considerar al desarrollar estrategias educativas y de mejora del desempeño.

Es crucial interpretar los resultados del ACP en el contexto específico de la institución educativa y el entorno escolar. Factores como el currículo, las políticas educativas y las características de los estudiantes pueden influir en los patrones identificados, por lo que es importante considerar estos aspectos al realizar análisis e interpretaciones.

En resumen, el análisis mediante ACP proporciona una herramienta poderosa para comprender la estructura subyacente de los datos y extraer información relevante sobre el rendimiento académico de los estudiantes. Estas conclusiones pueden orientar la toma de decisiones educativas y contribuir a la implementación de intervenciones efectivas para mejorar el aprendizaje y el éxito estudiantil

Si se observa la variabilidad capturada por las componentes principales en cada año, las componentes principales CP1 y CP2 explican una parte significativa de la variabilidad en los datos, destacando las dimensiones más relevantes del rendimiento académico en las diferentes asignaturas.

En la prueba saber del año 2019, todas las asignaturas presentan altas contribuciones a CP1, indicando que CP1 captura una combinación general de habilidades académicas. Inglés contribuye significativamente a CP2, sugiriendo una dimensión de habilidades lingüísticas específicas, en 2020, se observa una estructura similar con todas las asignaturas mostrando contribuciones altas a CP1. Matemáticas e Inglés presentan altos valores en CP2, reflejando dimensiones diferenciadas en razonamiento lógico y habilidades en lengua extranjera, en 2021, todas las asignaturas mantienen una alta contribución a CP1. Matemáticas e Inglés nuevamente destacan en CP2, sugiriendo la persistencia de la dimensión de habilidades específicas.

Si se observan los cambios en las contribuciones a lo largo de los años, la comparación muestra que, aunque las contribuciones a CP1 se mantienen relativamente estables, las contribuciones a CP2 varían significativamente. Por ejemplo, Inglés cambia de tener una contribución positiva en 2019 a negativa en 2020, y vuelve a ser positiva en 2021. Por otro lado, matemáticas también muestra cambios en su contribución a CP2, pasando de una contribución negativa en 2019 a positiva en 2020, y nuevamente negativa en 2021.

Al mirar implicaciones de los cambios observados, los cambios en las contribuciones de las asignaturas a las componentes principales pueden reflejar variaciones en el enfoque curricular, métodos de enseñanza, o incluso en las características de los cohortes de estudiantes, adicional, la variabilidad en las contribuciones a CP2 podría estar indicando cambios en la importancia relativa de ciertas habilidades específicas (como el razonamiento lógico o habilidades en lengua extranjera) en las diferentes cohortes.

Si se observa la interpretación de las dimensiones académicas, CP1 en todos los años parece capturar una dimensión general de rendimiento académico en diversas asignaturas,

reflejando una combinación de habilidades científicas, sociales y lingüísticas. Por otro lado, CP2 capta dimensiones más específicas, que varían año a año, y que pueden estar relacionadas con habilidades particulares en matemáticas e inglés, así como en otras áreas como sociales y ciencias naturales.

En resumen, el ACP de las Pruebas Saber de 2019, 2020 y 2021 revela patrones consistentes en la estructura del rendimiento académico de los estudiantes, pero también muestra variabilidad en cómo ciertas habilidades específicas contribuyen a las diferentes dimensiones año tras año. Estos hallazgos pueden ser utilizados para ajustar estrategias educativas y políticas curriculares con el fin de mejorar el rendimiento estudiantil de manera más focalizada y efectiva.

6.2. Recomendaciones Futuras

Es relevante destacar que en este estudio se realizó un Análisis de Componentes Principales (ACP) para cada grado y área temática. Esta técnica permitió identificar patrones y relaciones significativas entre las variables, resaltando aquellas más relevantes para cada componente académico. Sería beneficioso complementar este análisis con técnicas adicionales, como clasificación y agrupamiento, para identificar patrones de desempeño académico similares entre estudiantes o grados. La aplicación de estas técnicas podría facilitar la identificación de grupos con características académicas comunes, lo que a su vez permitiría personalizar enfoques de enseñanza según las necesidades específicas de cada grupo identificado. Este enfoque personalizado puede contribuir a mejorar la efectividad de las estrategias educativas y promover un aprendizaje más adaptado a las particularidades de cada conjunto de estudiantes.

En este trabajo de grado, sería pertinente considerar la implementación de Modelos de Ecuaciones Estructurales (SEM) con el propósito de evaluar las relaciones causales entre las variables latentes y observadas. La aplicación de SEM puede proporcionar una comprensión más profunda de cómo diversos aspectos del desempeño académico están interconectados, lo que permite un análisis más detallado de las relaciones subyacentes entre las variables.

Otro aspecto importante a considerar para futuros trabajos sería el empleo de técnicas de validación cruzada para evaluar la robustez de los modelos estadísticos. Esta práctica resulta especialmente útil al construir modelos predictivos, ya que garantiza su generalización a datos no vistos y proporciona una medida más precisa de su capacidad predictiva.

Apéndice A

Apéndice: Soportes Algebraicos

A.1. Primera sección

A.1.1. Teorema

Sea Σ la matriz de covarianzas del vector aleatorio $X = (X_1, \dots, X_p)'$ y $(\lambda_1, e_1), \dots, (\lambda_p, e_p)$ los pares de vectores y valores propios asociados a Σ , donde $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. Entonces, la i -ésima componente principal de Σ está dada por:

$$\begin{aligned} Y_i &= e_i'X = e_{i1}X_1 + \dots + e_{ip}X_p, \text{ donde } i = 1, 2, \dots, p, \\ \text{Var}[Y_i] &= e_i'\Sigma e_i = \lambda_i, \text{ donde } i = 1, 2, \dots, p, \text{ y} \\ \text{Cov}[Y_i, Y_k] &= e_i'\Sigma e_k = 0, \text{ donde } i, k = 1, 2, \dots, p; i \neq k. \end{aligned}$$

Haciendo $B = \Sigma$, se tiene que:

$$\begin{aligned} \max_{X \neq 0} \frac{X^T \Sigma X}{X^T X} &= \lambda_1 = e_1' \Sigma e_1 = \text{Var}[Y_1], \\ \max_{X \perp e_1, \dots, e_k} \frac{X^T \Sigma X}{X^T X} &= \lambda_{k+1} = e_{k+1}' \Sigma e_{k+1} = \text{Var}[Y_{k+1}]. \end{aligned}$$

Finalmente,

$$\text{Cov}[Y_i, Y_k] = \text{Cov}[e_i'X, e_k'X] = e_i'\Sigma e_k = e_i'e_k \lambda_k = 0, \text{ donde } i \neq k.$$

A.1.2. Proposición 1:

Sea $X = (X_1, \dots, X_p)'$ un vector aleatorio con matriz de covarianzas Σ y pares de valores y vectores propios $(\lambda_1, e_1), \dots, (\lambda_p, e_p)$, tales que $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. Sean Y_1, \dots, Y_p las componentes principales asociadas a Σ . Entonces:

$$\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \sum_{i=1}^p \text{Var}[X_i] = \lambda_1 + \lambda_2 + \dots + \lambda_p = \sum_{i=1}^p \text{Var}[Y_i].$$

Prueba:

Se sabe que $\sigma_{11} + \dots + \sigma_{pp} = \text{tr}(\Sigma)$. Ahora, como $\Sigma = P\Delta P'$, donde P es una matriz ortogonal formada por los vectores propios de Σ y Δ es una matriz diagonal que contiene los valores propios de Σ , se tiene que:

$$\sum_{i=1}^p \sigma_{ii} = \text{tr}(\Sigma) = \text{tr}(\Delta P'P) = \text{tr}(\Delta P P') = \text{tr}(\Delta) = \sum_{i=1}^p \lambda_i.$$

La cantidad $\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp}$ se conoce como Varianza poblacional total. La proporción de varianza total explicada por la i -ésima componente principal se calcula como

$$\frac{\lambda_i}{\lambda_1 + \lambda_2 + \dots + \lambda_p}, \text{ para } i = 1, 2, \dots, p.$$

En relación con la i -ésima componente principal, suponga que $e_i = (e_{i1}, e_{i2}, \dots, e_{ip})'$. La magnitud de e_{ik} mide la importancia de la k -ésima variable X_k sobre la i -ésima componente principal. Más aún, e_{ik} es proporcional a la correlación entre X_k y Y_i

A.1.3. Proposición 2:

Sea $X = (X_1, \dots, X_p)'$ un vector aleatorio con matriz de covarianzas Σ y pares de valores y vectores propios $(\lambda_1, e_1), \dots, (\lambda_p, e_p)$, tales que $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. Sean Y_1, \dots, Y_p las componentes principales asociadas a Σ . Entonces:

$$\rho(Y_i, X_k) = \frac{e_{ik}}{\sqrt{\lambda_i \sigma_{kk}}}, \text{ donde } i, k = 1, 2, \dots, p.$$

Sea a un vector en \mathbb{R}^p con 1 en la k -ésima entrada y cero en las demás, es decir, $a = (0, \dots, 0, 1, 0, \dots, 0)'$. De esta manera, se tiene que $X_k = a'X$.

Como $Y_i = e_i'X$, entonces

$$\text{Cov}[Y_i, X_k] = \text{Cov}[a'X, e_i'X] = a'\Sigma e_i = a'\lambda_i e_i = \lambda_i e_{ik}.$$

$$\rho(Y_i, X_k) = \frac{\text{Cov}[Y_i, X_k]}{\sqrt{\text{Var}[Y_i]\text{Var}[X_k]}} = \frac{\lambda_i e_{ik}}{\sqrt{\lambda_i \sigma_{kk} \sigma_{kk}}} = \frac{e_{ik}}{\sqrt{\lambda_i \sigma_{kk}}}.$$

Aunque las correlaciones entre las componentes principales y las variables sean útiles para interpretar las componentes principales, solo miden la contribución individual de las variables aleatorias sobre una componente particular. Esto es, no indican la importancia de una variable X sobre una componente principal Y en presencia de las demás variables X 's. Por esta razón se recomienda que solo los coeficientes e_{ik} , y no las correlaciones, sean usados para interpretar las componentes principales.

A.1.4. Componentes principales de variables estandarizadas

Es posible también obtener componentes principales a partir de variables estandarizadas. Sea $X = (X_1, \dots, X_p)'$ un vector aleatorio con vector de medias μ y matriz de covarianzas Σ .

Sea $Z = (Z_1, \dots, Z_p)'$ otro vector aleatorio tal que $Z_i = \frac{X_i - \mu_i}{\sqrt{\sigma_{ii}}}$, para $i = 1, 2, \dots, p$.

$$\text{Entonces } Z = (V^{1/2})^{-1}(X - \mu), \text{ donde } V^{1/2} = \begin{bmatrix} \sqrt{\sigma_{11}} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sqrt{\sigma_{pp}} \end{bmatrix}$$

y, $E[Z] = 0$ y $\text{Cov}[Z] = (V^{1/2})^{-1}\Sigma(V^{1/2})^{-1} = \rho$.

Las componentes principales de Z pueden ser obtenidas de los vectores propios de ρ , la matriz de correlaciones de X . En general, las componentes principales obtenidas con Σ son diferentes de las obtenidas con ρ . Cuando las escalas de medición de las variables son muy diferentes, se sugiere estandarizarlas.

A.1.5. Proposición 3:

Sea $X = (X_1, \dots, X_p)'$ un vector aleatorio con vector de medias μ y matriz de covarianzas Σ , y sea $Z = (V^{1/2})^{-1}(X - \mu)$, donde $V^{1/2} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & \dots & 0 \\ 0 & \sqrt{\sigma_{22}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{\sigma_{pp}} \end{bmatrix}$.

La i -ésima componente principal de las variables estandarizadas está dada por:

$$Y_i = e_i'Z = e_i'(V^{1/2})^{-1}(X - \mu), \text{ para } i = 1, 2, \dots, p.$$

Más aún,

$$\sum_{i=1}^p \text{Var}[Y_i] = \sum_{i=1}^p \text{Var}[Z_i] = p$$

y

$$\rho(Y_i, Z_k) = \frac{e_{ik}}{p\lambda_i}, \text{ para } i, k = 1, 2, \dots, p,$$

donde $(\lambda_1, e_1), \dots, (\lambda_p, e_p)$ son los pares de valores y vectores propios de ρ , con $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$.

A.1.6. Componentes principales muestrales

Suponga que se tiene una muestra aleatoria de una distribución p -dimensional representada por las variables aleatorias X_1, \dots, X_n , que están asociadas al vector $X = (X_1, \dots, X_p)'$. Este vector tiene un vector de medias μ y una matriz de covarianzas Σ desconocida. A partir de esta muestra, se puede calcular el vector de medias muestrales \bar{X} , la matriz de covarianzas muestrales S , y la matriz de correlaciones muestrales R .

El objetivo es encontrar combinaciones lineales de las características medidas que capturen la mayor cantidad de variabilidad muestral posible. Estas combinaciones se conocen como Componentes Principales Muestrales. La siguiente proposición describe cómo encontrar estas componentes.

A.1.7. Proposición 4:

Suponga que se tiene una muestra aleatoria de una distribución p -dimensional representada por las variables aleatorias X_1, \dots, X_n , que están asociadas al vector $X = (X_1, \dots, X_p)'$. La muestra tiene un vector de medias μ y una matriz de covarianzas desconocida Σ .

Se calcula la matriz de covarianzas muestral S y denote los pares de valores y vectores propios de S como $(\hat{\lambda}_1, \hat{e}_1), \dots, (\hat{\lambda}_p, \hat{e}_p)$, donde $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p \geq 0$.

Entonces, la i -ésima componente principal asociada a S se define como:

$$\hat{Y}_i = \hat{e}'_i X = \hat{e}_{i1} X_1 + \dots + \hat{e}_{ip} X_p, \text{ para } i = 1, 2, \dots, p.$$

Esta proposición describe cómo encontrar las componentes principales muestrales a partir de una muestra de variables p -dimensionales y su matriz de covarianzas muestral S . Estas componentes capturan la variabilidad muestral de manera eficaz.

$$s_{\hat{Y}_k}^2 = \hat{\lambda}_k \quad \text{y} \quad \sigma_{\hat{Y}_i, \hat{Y}_k} = 0; \quad i, k = 1, \dots, p, \quad i \neq k.$$

$$\sum_{i=1}^p s_{ii} = \sum_{i=1}^p \hat{\lambda}_i \quad \text{y} \quad r_{\hat{Y}_i, X_k} = \frac{\hat{e}_{ik} \sqrt{\hat{\lambda}_i}}{\sqrt{s_{kk}}}; \quad i, k = 1, \dots, p.$$

Si las observaciones X_i están centradas alrededor de \bar{X} , esta operación no tiene efecto sobre la matriz de covarianzas muestral. En ese caso, la i -ésima componente principal está dada por:

$$\hat{Y}_i = \hat{e}'_i (X - \bar{X}), \quad \text{para } i = 1, \dots, p.$$

Haciendo $\hat{Y}_{ij} = \hat{e}'_i (X_j - \bar{X})$, para $i = 1, \dots, p$, se tiene que:

$$\hat{Y}_i = \frac{1}{n} \sum_{j=1}^n \hat{e}'_i (X_j - \bar{X}) = \frac{1}{n} \hat{e}'_i \left(\sum_{j=1}^n (X_j - \bar{X}) \right) = \frac{1}{n} \hat{e}'_i 0 = 0.$$

Bibliografía

- Acosta Solano, J., Lancheros Cuesta, D. J., Umaña Ibáñez, S. F., and Coronado Hernández, J. R. (2022). Predictive models assessment based on crisp-dm methodology for students performance in colombia saber 11° test. *Corporación Universitaria Rafael Núñez, Universidad Cooperativa de Colombia, Departamento de Productividad e Innovación Universidad de la Costa CUC*, 198(10):512–517.
- Allaire, J., Xie, Y., Dervieux, C., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., and Iannone, R. (2024). *rmarkdown: Dynamic Documents for R*. R package version 2.26, <https://pkgs.rstudio.com/rmarkdown/>.
- Cerón Benavides, S. M., Romero Ospina, M. F., and Fonseca Gómez, L. R. (2020). *Análisis Estadístico Multivariado De Los Resultados En Las Pruebas Saber Pro Del Programa De Ingeniería En Producción Acuícola En La Universidad De Nariño 2016 – 2019*. Fundación Universitaria Los Libertadores.
- Chan, D. and Galli, M. G. (2020). Aplicación de técnicas estadísticas multivariadas con el lenguaje de programación r en investigaciones educativas del nivel superior. *Revista Argentina de Educación Superior (RAES)*, 12(20):123–136.
- Chue Gallardo, J., Barreno, E., and Millones Rivalles, R. (2007). Sistema para el análisis estadístico con técnicas multivariadas del rendimiento académico de los estudiantes de una institución de enseñanza superior. *Universidad Nacional Agraria La Molina, The University of San Martín de Porres y Universidad de Lima*, 2(164).
- Congreso de la República, d. C. (1991). Constitución política de 1991, asamblea constituyente. http://www.secretariasenado.gov.co/senado/basedoc/constitucion_politica_1991.html.
- Corredor Rivera, D. A. and Pineda Ríos, W. (2017). *Estudio De La Incidencia Actitudinal De Los Estudiantes En Pruebas De Matemáticas Tipo ICFES: Una Aproximación Semiparamétrica*. Universidad Santo Tomas.
- Garanto, J., Mateo, J., and Rodríguez, S. (2002). *Modelos Y Técnicas De Análisis Del Rendimiento Académico*. Dialnet.
- Ibarra, M. d. C. and Michalus, J. C. (2010). Análisis del rendimiento académico mediante un modelo logit. *Universidad Nacional de Misiones*, 9(2):47–56.
- ICFES (2016). *Instituto Colombiano Para La Evaluación de la Educación, Resolución 457 del 14 de julio del 2016*. normograma.icfes.gov.co.

- ICFES (2021). *Saber al detalle: ¿Cuáles Son Los Modelos De Calificación De Las Pruebas Saber?*, volume 8, ISSN: 2590 - 4663. Mineducación.
- Jhonson, R. and Wichern, D. (2007). *Applications of Multivariate Technique*. Pearson Education.
- Madrid Echeverry, J. I. (2017). *Propuesta De Un Modelo Estadístico Para Caracterizar Y Predecir La Deserción Estudiantil Universitaria*. Universidad Nacional de Colombia, Facultad de Minas, Medellín.
- Marquín Triana, M. J. and Romero, M. (2017). *Predicción Del Rendimiento Académico Mediante Técnicas Del Análisis Multivariado En La Asignatura De Algebra Lineal*. Universidad Tecnológica de Pereira (UTP).
- McArdle, J. J., Paskus, T. S., and Boker, S. M. (2013). A multilevel multivariate analysis of academic performances in college based on ncaa student-athletes. *Multivariate Behavioral Research*, 48(1):57–95.
- MEN (1994). Congreso de la República de Colombia, Ley 115 de Febrero 8 de 1994, artículo 80. <https://www.medellin.gov.co/es/centro-documental/ley-115-de-1994-ley-general-de-educacion/>.
- MEN (2001). Congreso de la República de Colombia, Ley 715 de 2001. <https://www.mineducacion.gov.co/portal/normativa/Decretos/86070:Decreto-1621-de-Agosto-2-de-2002>.
- MEN (2008). Al tablero: Periódico de un país que educa y que se educa. <https://www.mineducacion.gov.co/1621/article-162342.html>.
- MEN (2015). Ministerio de Educación Nacional, Decreto no. 1075 del 26 de mayo de 2015. https://www.mineducacion.gov.co/1780/w3-article-351080.html_noredirect=1.
- Monroy, L. G. D. (2007). *Estadística multivariada: inferencia y métodos*. Universidad Nacional de Colombia.
- Okeke, E. N., Okeke, J. U., and Daniel, A. (2018). Multivariate analysis of variance of university students' academic performance. *La Federal University Wukari (FUW)*, 7(1):13–20.
- Peña Méndez, D. P. and Gutiérrez Sánchez, R. (2014). *Análisis de componentes principales en la estimación de índices de empoderamiento en mujeres de Colombia*. Universidad de Granada.
- Pustovalova, N. and Avdeenko, T. (2022). Multivariate analysis of the influence of students' characteristics on academic performance. *State Technical University Novosibirsk*, 978(1):09.
- Pérez Pulido, M., Aguilar Galvis, F., Orlandoni Merli, G. R. P., and J. (2016). Análisis estadístico de los resultados de las pruebas de estado para el ingreso a la educación superior en la universidad de santander. *Universidad de Santander*, 27(27):328–339.

- Rodríguez Manrique, J. A., Ruiz Escorcía, R. R., and Cohen Manrique, C. S. (2022). Análisis multivariado aplicado a la evaluación de competencias saber-pro en el departamento de sucre, colombia. *Corporación Universitaria del Caribe (CECAR)*, 16(16):19–21.
- Rodríguez Rodríguez, A. K. and Hernández Cano, C. Y. (2021). *Análisis de las Pruebas Saber 11, años 2017 y 2018, identificando las variables determinantes en los bajos resultados y en la brecha existente entre los estudiantes de colegios categoría A y A+ versus los D en el sector público*. Disponible en https://ciencia.lasalle.edu.co/ing_industrial/170.
- RStudio Team (2015). *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA.
- Ruiz Escorcía, R. R., Arevalo Medrano, J. B., Morillo, G. P., and Acosta-Humanéz, P. B. (2016). Principal component analysis applied to the state colombian test icfes saber 11°. *Universidad Nacional de las lomas, Corporación Universitaria Minuto de Dios, Universidad Simón Bolívar*, 39(10):3.
- Xie, Y. (2024). *knitr: A General-Purpose Package for Dynamic Report Generation in R*. R package version 1.46.