



**Análisis del rendimiento académico de la educación primaria y secundaria
en el departamento Uruguay del año 2023**

Tecnicatura Universitaria en Procesamiento y Explotación de Datos

Exploración de Datos Multivariados

Docente:

Melisa Fernández

Alumnos:

Elias Coradini, Daiana Gareis

Oro Verde - 5 de junio del 2024

Resumen

Este estudio proporciona un análisis detallado del rendimiento académico de los estudiantes de educación primaria y secundaria en el departamento de Uruguay, Entre Ríos, durante el año 2023. A pesar de los desafíos encontrados, como la carga incorrecta de datos por parte de varias escuelas y la imposibilidad de tener notas inferiores a 6 en primaria se consideró que los datos podrían llevarnos a resolver los problemas planteados. Se identificó una preferencia por las escuelas públicas y el turno mañana entre los estudiantes de ambos niveles. Además, se observó una diferencia en el rendimiento académico entre las materias y los niveles educativos: los estudiantes de primaria tienen un mejor desempeño en Matemática que en Lengua, mientras que en secundaria el rendimiento es mejor en Lengua que en Matemática. Estos hallazgos son fundamentales para la implementación de programas complementarios que buscan mejorar el desempeño académico.

Introducción

La educación es un pilar fundamental para el desarrollo social y económico del país. Por lo tanto, es crucial monitorear el nivel de aprendizaje de los estudiantes para identificar qué materias se les dificultan más y así poder garantizar el correcto aprendizaje. En este informe, se aborda precisamente esta problemática. Para ello, se analizará una base de datos de las escuelas del departamento de Uruguay, dicha base cuenta con un total de 18,920,746 registros distribuidos en 727,721 filas y 26 columnas. Se identificaron 14 variables cuantitativas y 12 variables cualitativas, algunas de estas son: las asignaturas, documento de los alumnos, turno de cursado, modalidad de enseñanza, año de cursado.

Se abordarán aspectos como la calidad de los datos, las características generales de la educación en el departamento asignado y el rendimiento de los estudiantes de nivel primario y secundario en las diferentes asignaturas que realizaron en el año 2023. Este análisis permitirá identificar las materias que más se les dificulta a los estudiantes, examinando su rendimiento durante los tres trimestres, para asignar acompañamientos en las asignaturas que están teniendo más dificultades. Este apoyo permitirá formular estrategias efectivas para garantizar un sistema educativo de calidad que promueva el desarrollo de los estudiantes y contribuya al crecimiento del país.

A lo largo del trabajo, se encontraron deficiencias en la base de datos analizada. Se identificaron numerosos datos mal cargados y duplicados, lo que generó inexactitudes en el análisis. Además, se observaron columnas cuyo significado era desconocido, lo que dificultó su interpretación. Algo a tener en cuenta es que séptimo año en el nivel secundario es únicamente de las escuelas técnicas, por lo que no será tomado en cuenta para los análisis.

Metodología

Para llevar adelante esta investigación decidimos usar el lenguaje de programación R junto a Jupyter Notebook implementado a través de la plataforma Anaconda. Empleamos numerosas librerías que son específicas de la distribución de Anaconda por ejemplo: readxl, readr, 'stringr', y otras que no son propias de esta, por ejemplo: dplyr, ggplot2, moments, nortest, entre otras.

En el dataset, encontramos algunas columnas que requirieron modificaciones en sus nombres para una mejor comprensión de la información que contienen, tales como 'ddivision' que se renombró como 'IDdivision', también se realizó una normalización en las variables de asignatura, por ejemplo nos encontramos con "Matematica" y "MATEMATICA". En cuanto a la calidad de los datos, se observó la presencia de valores faltantes en algunas variables, así como datos sucios en la variable 'Nota'. Asimismo, se identificaron registros duplicados en el conjunto de datos. Para resolver este problema, decidimos quedarnos solo con registros únicos utilizando la función unique().

Además, se encontraron columnas cuyo significado desconocemos, como: X, orden y Nivel. También, habían algunas que consideramos irrelevantes para nuestro análisis, como:CodigoPostal, anioLectivo, observaciones, localidad; y una que resultó ser redundante: anioLectivo.

Para realizar un análisis más detallado, se agregaron diferentes variables que permitieron realizar el análisis del desempeño de los alumnos, estas fueron NotasPrimerT, NotasSegundoT, NotasTercerT, Año_cursado, Promedio y Resultado. Las nuevas columnas de Notas son el resultado de la combinación de las columnas "periodoEvaluatorio" y "nota".

También se decidió agregar Desempeño, que categoriza los alumnos dependiendo su promedio, en primaria estas categorías se definieron de la siguiente manera: "Insuficiente" si el promedio es NA y todas las notas trimestrales son "EPAC", "Regular" si el promedio es NA y la tercera nota trimestral es "EPAC" o si el promedio es NA y dos de las tres notas trimestrales son "EPAC", "Aprobado" si el promedio es menor a 7 o si una de las tres notas trimestrales es "EPAC" y la nota de tercer trimestre no es "EPAC", "Bueno" si el promedio está entre 7 y menor a 8, "Muy Bueno" si el promedio está entre 8 y menor a 9 y "Distinguido" si el promedio es 9 o mayor.

Mientras que en secundaria las categorías para la variable "Desempeño" se definieron de esta manera: "Insuficiente", si el promedio es menor que 4, "Regular", si el promedio es mayor o igual a 4 y menor que 6, "Aprobado", si el promedio es mayor o igual a 6 y menor que 7, "Bueno" si el promedio es mayor o igual a 7 y menor que 8, "Muy Bueno" si el promedio es mayor o igual a 8 y menor que 9, y Distinguido si el promedio es mayor o igual a 9 y menor o igual a 10.

También se decidió agregar la columna "cantidad_rinde", la cual contabiliza la cantidad de materias que el alumno no aprueba. Para ello, se cuenta el valor "En proceso" en la columna "Resultado" para cada valor único en la columna "Documento" en el caso de primaria, y se cuenta el valor "Rinde" en la columna "Resultado" en el caso de secundaria.

Se dividió el dataset en dos subconjuntos de datos según el nivel educativo, resultando dos datasets: primaria y secundaria. Para comparar las matrículas totales de alumnos que había en cada nivel se emplearon otras bases de datos proveídas por el ministerio de educación, estas son: primaria 2023, secundaria 2023, primaria por departamentos y secundaria por departamentos.

En el caso de primaria, se eliminó la columna IDModEnsenianza ya que no era relevante para el análisis. Luego para las variables de notas se renombraron PD a EPAC, se verificó que no hayan notas mayores a 10 ni menores a 6 y se eliminaron los que eran NULL. También se filtró ModEnsenianza Común y Escuelas NINA, no Multianio y Modalidad Primaria.

Para limpiar el dataset de secundaria se realizaron varios pasos, una vez creadas las nuevas columnas se eliminaron las columnas curso y modalidad porque su información no contribuía con el análisis que se llevó a cabo o tenían información irrelevante. También, se borraron los "EPAC, PA, PD" ya que no pueden estar en el nivel secundario y otros datos borrados fueron los que tenían registros "NULL" que dificultan el análisis. Por último, se filtró el dataset de manera que no aparezcan datos de alumnos multiaños y que solo queden aquellos con estas modalidades de enseñanza: "Comun", "Agrotécnica", "Técnica", "EGB3".

Notamos una discrepancia en los números de documento de los estudiantes tanto en primaria como en secundaria, encontramos casos especiales de números de documento extremadamente bajos y altos. Por ello, el último paso de limpieza que llevamos a cabo fue quedarnos con un rango de números de documento que se consideraron más coherentes y representativos de la población estudiantil. De esta manera, en primaria se optó por considerar los documentos mayores a 48 millones y menores a 100 millones, y en secundaria solo se tuvieron en cuenta los documentos a partir de 43 millones hasta aquellos menores a 100 millones.

Para analizar si los datos eran normales se realizaron histogramas de los promedios, ya que si presentamos la distribución de una variable en un histograma de frecuencias, es posible fijar los límites entre los que se encuentra la mayoría de los datos. También se

emplearon gráficos Q-Q plots que examinan la distribución de los datos y la compara con la esperada para una distribución normal que tiene la misma media y varianza. Si estos datos se ajustan a una medida normal, el gráfico de probabilidad mostrará que la mayoría de los puntos están lo más cercano posible a la recta trazada, esto indica su posible normalidad. Para verificar si lo que los gráficos mostraban corresponden realmente a una distribución normal, se realizaron test de Kolmogorov-Smirnov cuando la cantidad de promedios a analizar era mayor a 50, y test de Shapiro-Wilk cuando la cantidad de promedios era menor a 50.

Realizamos una estandarización de las variables de promedio en Matemática y Lengua. Al estandarizar los datos, podemos ver fácilmente cuánto se aleja cada valor de la media en términos de desviaciones estándar, lo que proporciona una medida más precisa de su posición relativa dentro de la distribución. También, se puede analizar las puntuaciones en diferentes asignaturas de manera más objetiva, ya que se eliminan las diferencias de escala y se centran los datos alrededor de un punto de referencia común.

Se realizaron transformaciones por variables, que se aplican con el objeto de hacer comparables los valores asignados a los distintos individuos u objetos de análisis. A estas transformaciones suele denominarse z-scores o puntuaciones Z..

Se analizó tanto en secundaria como en primaria si existía alguna correlación entre las diferentes asignaturas. Para visualizarlas se realizaron gráficos de dispersión, ya que permiten identificar relaciones no lineales y outliers. Además, se utilizó el coeficiente de correlación de Spearman para verificar exactamente si la correlación era débil, moderada o fuerte. No se podía utilizar el test de correlación de Pearson ya que los datos de cada año no seguían una distribución normal. Luego se realizaron matrices de correlación para cada año de primaria y secundaria, ya que informa como son las correlaciones entre dos variables de a dos, no permite saber cómo son las correlaciones de todas las variables, además para asegurarnos de que existían esas correlaciones, realizamos correlaciones parciales, estas a diferencia de las correcciones normales, los valores no son afectados por las otras variables.

Para realizar comparaciones entre dos poblaciones basadas en información de dos muestras, una de cada población, se realizaron diversos tests. Para el análisis univariado, se utilizó el test no paramétrico de Wilcoxon debido a que los datos no cumplían con los supuestos de normalidad.

En el caso del análisis bivariado, primero se debía determinar si los grupos eran independientes o dependientes. Si los grupos eran independientes, se comparó la

homogeneidad de las varianzas mediante la prueba de Levene y al no contar con distribuciones normales, se utilizó el test de Wilcoxon. Si los grupos eran dependientes, se aplicó el test no paramétrico de muestras dependientes llamado Sign Test.

Cuando se trataba de más de dos grupos, al igual que en el análisis bivariado, se debía comprobar si los grupos eran dependientes o independientes. Si eran independientes, se analizaba la normalidad y la varianza. Dado que los datos no eran normales, se utilizó el test de Kruskal-Wallis. En el caso de grupos dependientes, se debía comprobar la linealidad y la normalidad de los datos. Se pudo utilizar el ANOVA de dos factores repetidos para muestras dependientes en muestras de estudiantes más pequeñas. También se realizó la versión no paramétrica de este último, el test de Friedman.

Se utilizó el test de Cochran para variables cualitativas. Por otro lado, no se pudieron realizar el ANCOVA y el MANOVA unidireccional debido a la falta de normalidad en los datos.

Un aspecto importante a tener en cuenta con los resultados de los tests es la potencia de estos. Si la potencia del test es baja, sugiere que podría no ser lo suficientemente sensible para detectar diferencias si existen. En estos casos, es recomendable considerar un tamaño de muestra mayor o mejorar el diseño del estudio para aumentar la potencia en futuros análisis.

Realizamos tablas de contingencia tanto para primaria, de primer y cuarto grado para las asignaturas de Lengua y Matemática, como para secundaria, de segundo y tercer año, en la asignatura de Matemática según la variable Desempeño, considerando el tipo de gestión, los turnos y también comparando el desempeño de Lengua con Matemática. En cada nivel educativo se optó por escoger estos grados/años y estas asignaturas porque en estas los estudiantes tuvieron más dificultades.

Se realizaron varios análisis estadísticos para evaluar la asociación entre dos variables categóricas en una tabla de contingencia. Se calculó el coeficiente de chi-cuadrado para determinar si existe una relación significativa entre las variables. Además, se examinaron los residuos, donde los valores positivos indican una mayor frecuencia de casos observados de lo esperado, mientras que los valores negativos indican una menor frecuencia. Se utilizó el coeficiente de V de Cramer para medir la fuerza de la asociación entre las variables, donde valores cercanos a 0 sugieren una asociación débil y valores cercanos a 1 indican una asociación fuerte.

Se realizó un análisis de datos multivariantes para llevar a cabo un estudio estadístico de varias variables medidas en un subconjunto de elementos de una población.

En este análisis, se calcularon los vectores de media, los cuales son vectores de dimensión p cuyos componentes representan las medias de cada una de las variables. Además, se obtuvieron las matrices de Varianza y Covarianza, donde la diagonal principal está compuesta por las varianzas muestrales de cada variable y los elementos fuera de la diagonal representan las covarianzas muestrales entre pares de variables.

Para detectar posibles valores atípicos (outliers), se calculó la distancia de Mahalanobis, la cual permite determinar similitudes entre dos variables aleatorias multidimensionales. Además, se utilizaron diferentes métodos como el Minimum Covariance Determinant (MCD), el Máximo Valor Atípico Esperado (MVE) y el método clásico. Es crucial identificar estos outliers, ya que si no son detectados, pueden alterar significativamente el análisis descriptivo de las variables originales.

Se realizaron algunos gráficos multivariados, como gráficos de coordenadas paralelas, perfiles paralelos, diagramas radiales, entre otros, para facilitar la visualización profunda y detallada de los datos, identificando patrones y relaciones entre las variables.

Por último, se realizaron test multivariantes como el Manova y el test de Hotelling, que no pudieron ser tomados en cuenta, ya que los datos no seguían los supuestos.

Resultados

Del dataset completo se identificaron un total de 57 escuelas en el nivel primario y 144 escuelas en el nivel secundario. Se realizó un conteo de la matrícula total y la matrícula cargada en notas de Matemática para cada grado en ambos niveles educativos, junto a su respectivo porcentaje de carga, esto se muestra en las tablas 1 y 2. En secundaria, además, se eligió no tomar en cuenta los alumnos de séptimo año, ya que este año solo está presente en las escuelas técnicas.

Tabla 1: Distribución de matrículas en Matemática por grado en primaria, departamento Uruguay.

PRIMARIA						
	1°	2°	3°	4°	5°	6°
Matrícula Total	1767	1731	1867	1790	1870	1753
Matrícula Cargada en Notas	862	880	925	917	900	1008
Porcentaje de Carga	48,78	50,84	49,54	51,23	48,13	57,50

Fuente: Elaboración propia (2024).

Tabla 2: Distribución de matrículas en Matemática por año en secundaria, departamento Uruguay.

SECUNDARIA						
	1°	2°	3°	4°	5°	6°
Matrícula Total	1972	2050	1790	1636	1425	1203
Matrícula Cargada en Notas	1203	1314	1100	1089	828	673
Porcentaje de Carga	61,00	64,10	61,45	66,56	58,11	55,94

Fuente: Elaboración propia (2024).

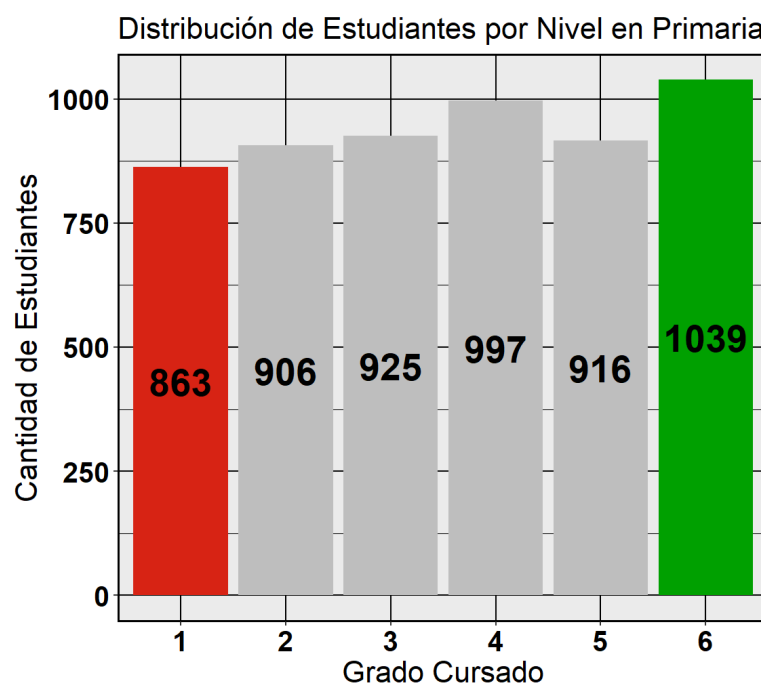
La distribución de estudiantes es diferente en primaria y secundaria. En primaria la distribución es bastante uniforme, teniendo mayor cantidad de estudiantes en sexto grado siendo en total 1039, mientras que el grado con menor cantidad es primero. En cambio en secundaria, se puede observar que luego de los dos primeros años hay una caída progresiva en la cantidad de estudiantes. El año con más estudiantes es segundo con 1404, mientras que el de menos es sexto. Las figuras 1 y 2 muestran la cantidad de estudiantes por grado/año. La barra color verde identifica el año con mayor cantidad de estudiantes y la roja, aquel que tiene menor cantidad.

El nivel secundario está preparado para recibir a los estudiantes que egresan de sexto grado de primaria, ya que en 2023 fueron menos los alumnos que hicieron sexto de primaria en comparación con los que hicieron primer año del secundario. Por esto mismo,

en 2024 no es necesario ampliar la cantidad de bancos en secundaria. Para el año siguiente, ocurriría lo mismo si se mantiene la cantidad de alumnos registrada en 2023, debido a que es menor la cantidad de alumnos en quinto de primaria que en sexto.

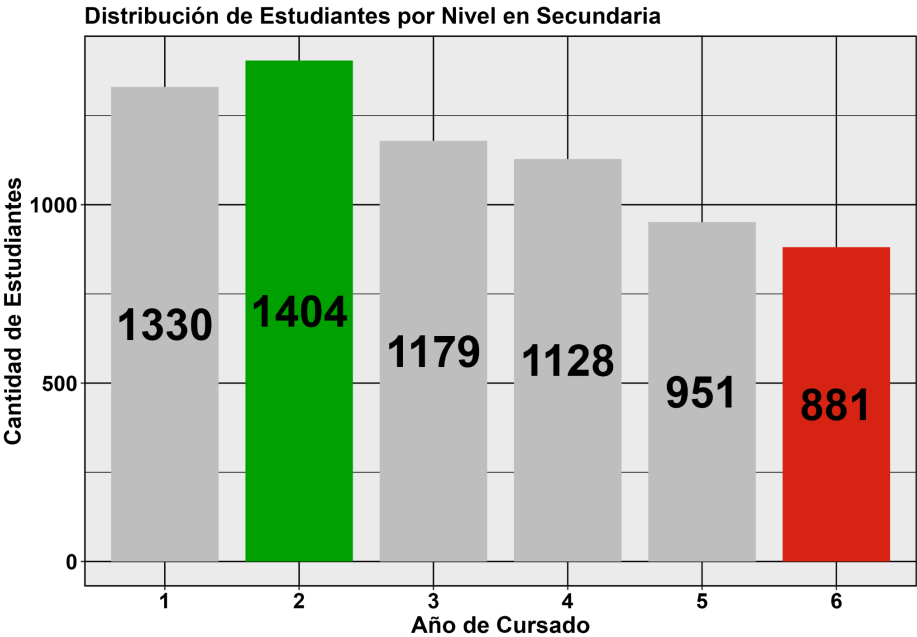
Se realizaron tests de Kruskal Wallis con los datos por departamento para contestar las siguientes preguntas: ¿Existen diferencias significativas en los conteos de estudiantes entre los diferentes ámbitos (escuela rural, urbana, etc) en primaria? La respuesta es sí, pues el p-value obtenido fue $3.15e-06$, un valor muy chico, lo que sugiere que hay diferencias significativas entre las matrículas en los distintos ámbitos. En el caso de secundaria la pregunta fue: ¿Las matrículas son iguales en todos los años en secundaria? El p-value obtenido fue 0.8235 lo que indica que no hay diferencias en las matrículas de los seis años.

Figura 1: Gráfico de barras, cantidad de alumnos de primaria por grados en el departamento Uruguay primaria.



Fuente: Elaboración propia (2024).

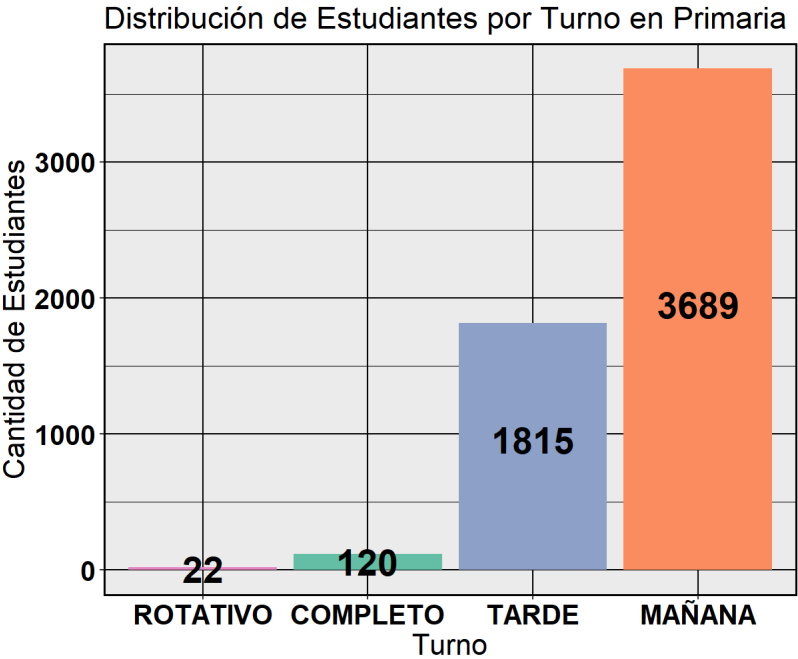
Figura 2: Gráfico de barras, cantidad de alumnos de secundaria por años en el departamento Uruguay secundaria.



Fuente: Elaboración propia (2024).

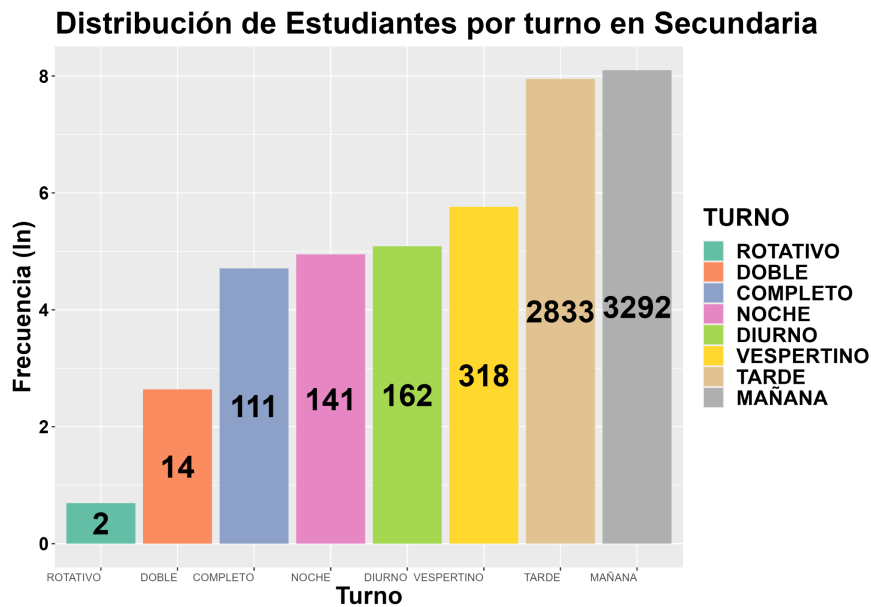
Esta cantidad de estudiantes se distribuye en diferentes turnos. Tanto en primaria como en secundaria el turno con más cantidad de alumnos es el de la mañana. Las figuras 3 y 4 muestran la distribución por turnos en cada nivel. En el caso de secundaria, para lograr una mejor visualización se optó por usar escala logarítmica(ln) en el eje y.

Figura 3: Gráfico de barras, distribución de alumnos por turno en el departamento Uruguay primaria.



Fuente: Elaboración propia (2024)

Figura 4: Gráfico de barras, distribución de alumnos por turno en el departamento Uruguay secundaria.



Fuente: Elaboración propia (2024).

Se decidió contestar la siguiente pregunta: ¿Existen diferencias significativas en las notas de los tres trimestres de Matemática de los alumnos de primer año del turno noche del secundario? Para ello se decidió realizar el test de Anova de dos factores repetidos. Para realizarlo se comprobó si los residuos del test eran normales, como el valor del p-value luego de aplicar el test de Shapiro-Wilk fue 0.1988 se confirmó la normalidad, que las varianzas eran homogéneas entre los trimestres y que no había violaciones significativas de la linealidad.

El p-value obtenido luego de realizar el test fue 0.1452. A partir de los resultados, se puede concluir que no hay evidencias suficientes para afirmar que existen diferencias significativas entre las notas de Matemática en los tres trimestres para los alumnos del primer año del turno noche. La potencia del test calculada es 0.374. Esto indica que el test tiene una baja probabilidad de detectar un efecto significativo si realmente existe.

En primaria, se quiso averiguar si existían diferencias significativas en las notas de los tres trimestres de Matemática en primer grado del turno completo, con el test de Anova de dos factores repetidos. Sin embargo, al realizar el test de Shapiro para comprobar la normalidad de los residuos, se obtuvo un p-valor de 0.0022, lo que indica que los residuos no siguen una distribución normal. Debido a esto, los resultados obtenidos del ANOVA no son correctos.

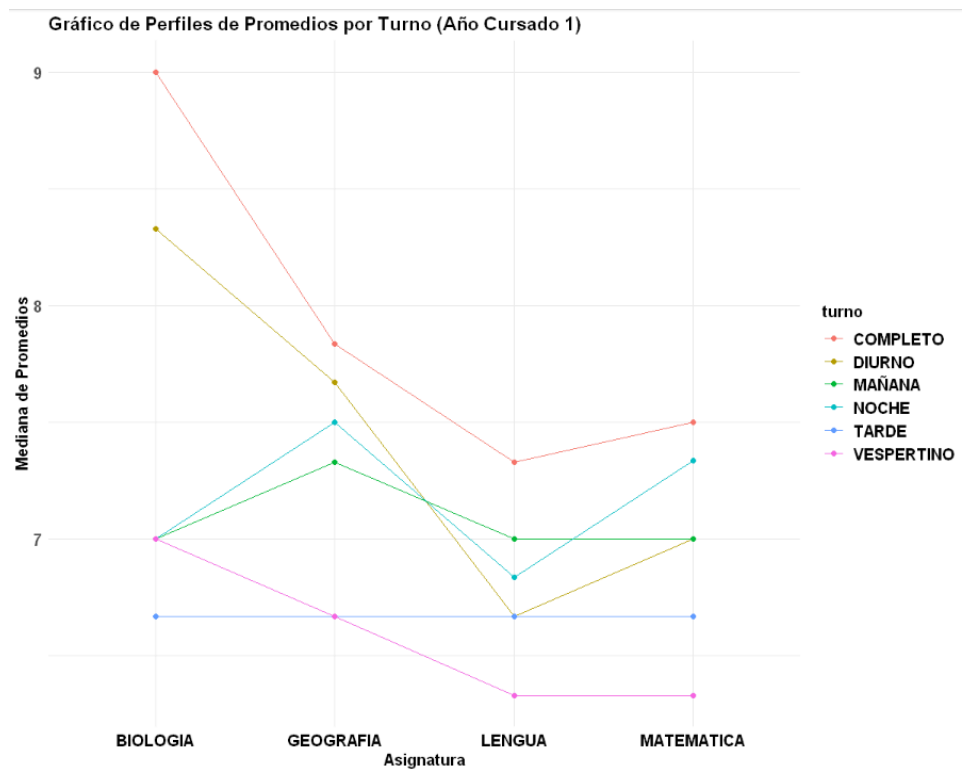
Dada esta situación se procedió a hacer la misma pregunta pero para segundo grado. En este caso, el test de Shapiro arrojó un p-valor de 0.8596, lo cual indica normalidad en los residuos. Se realizó un test de varianzas y se comprobó que tienen un p-value superior al 0.05 en todos los trimestres, lo que sugiere que no hay diferencias significativas en estas.

Los resultados del ANOVA indican que al menos una de las medias de las notas de Matemática en los tres trimestres de los alumnos de segundo grado del turno completo es diferente, con una potencia baja de 0.321.

Se quiso responder la siguiente pregunta: ¿Existe una diferencia entre los promedios de los diferentes turnos en las asignaturas troncales de primer año? Para responderla, se quería usar el test de manova dado que se requería un test multivariado, para ello se comprobó si se cumplían los supuestos. Primero se calculó la normalidad de los diferentes turnos y las asignaturas de estos dieron que en su mayoría no seguían una distribución normal en la mayoría de los casos, con la excepción del turno Noche donde todas las asignaturas cumplen con la normalidad. Luego se calculó la normalidad multivariante y el resultado del test en la prueba de Mardia-Asimetría dio un p-value= 0.0007452443 y la Prueba de Mardia - Curtosis un p-value de 4.402779e-11. Por esto se concluye que se rechaza la hipótesis nula de normalidad multivariada en términos de asimetría y en términos de curtosis. Como no se cumplió con los supuestos de normalidad y la normalidad multivariante se concluye que no se pudo aplicar el test de Manova.

El siguiente gráfico muestra cómo fue el desempeño de los alumnos de primer año de todos los turnos en las materias troncales. En el mismo, lo más llamativo es que los turnos mañana y noche parecen tener un rendimiento parecido pero en el caso del turno tarde los alumnos tuvieron el mismo rendimiento en las cuatro asignaturas.

Figura 5: Gráfico de coordenadas paralelas, departamento Uruguay secundaria.



Fuente: Elaboración propia (2024).

Tablas de contingencia según el turno en Primaria

Se crearon tablas de contingencia para la variable desempeño según el turno en la asignatura de Lengua para los estudiantes de cuarto grado.

Porcentaje Total

	insuficiente	regular	aprobado	bueno	muy bueno	distinguido	Total
Mañana	0.9	6.6	11.2	17.4	20.1	16.1	72.3
Tarde	0.8	4.0	3.7	6.6	7.3	5.2	27.6
Total	1.7	10.6	14.9	24.0	27.4	21.3	99.9

El 72.3% del total de estudiantes de cuarto grado en la asignatura de Lengua se encuentran en el turno de la mañana. En contraste el 27.6% se encuentran en el turno de la tarde.

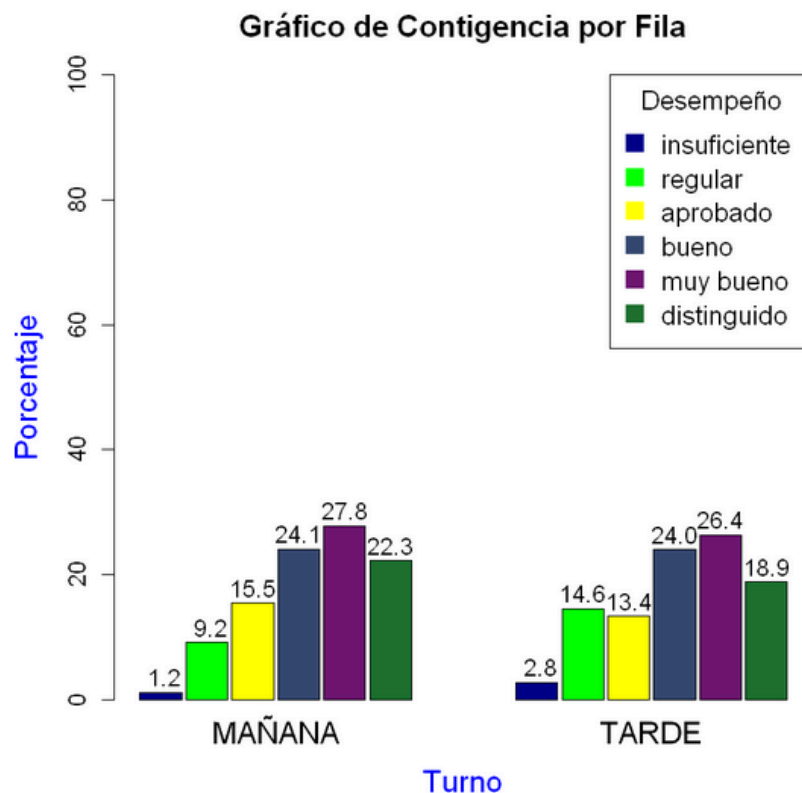
Porcentaje Fila

	insuficiente	regular	aprobado	bueno	muy bueno	distinguido	Total
Mañana	1.2	9.2	15.5	24.1	27.8	22.3	100.0
Tarde	2.8	14.6	13.4	24.0	26.4	18.9	100.0
Total	1.6	10.7	14.9	24.0	27.4	21.3	100.0

En el turno Mañana se observa que el 1.2% de los estudiantes en este turno tienen un desempeño "Insuficiente", el 9.2% tienen un desempeño "Regular", el 15.5% tienen un desempeño "Aprobado", el 24.1% tienen un desempeño "Bueno", el 27.8% tienen un desempeño "Muy Bueno", y el 22.3% tienen un desempeño "Distinguido".

Mientras que en el turno Tarde el 2.8% de los estudiantes tienen un desempeño "Insuficiente", el 14.6% tienen un desempeño "Regular", el 13.4% tienen un desempeño "Aprobado", el 24.0% tienen un desempeño "Bueno", el 26.4% tienen un desempeño "Muy Bueno", y el 18.9% tienen un desempeño "Distinguido".

Figura 6: Gráfico de barras adosadas por fila por desempeño y turno en Lengua de cuarto grado, departamento Uruguay Primaria



Fuente: Elaboración propia (2024).

Porcentaje Columna

	insuficiente	regular	aprobado	bueno	muy bueno	distinguido	Total
Mañana	53.3	62.2	75.2	72.4	73.4	75.5	72.4
Tarde	46.7	37.8	24.8	27.6	26.6	24.5	27.6
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0

En el turno Mañana tiene la mayor concentración en "Distinguido" un 75.5%, mientras que en el turno Tarde la categoría con mayor concentración es "Insuficiente" con un 46.7%.

. Podemos notar que, en general, hay diferencias significativas en la distribución del desempeño entre los turnos. Por ejemplo, el turno de la Mañana tiende a tener un mayor porcentaje de estudiantes en las categorías de "Regular", "Aprobado", "Bueno", "Muy Bueno" y "Distinguido", mientras que el turno de la Tarde tiene un mayor porcentaje de estudiantes en la categoría de "Insuficiente".

Se realizó el test de Chi-cuadrado para evaluar la relación entre el turno de asistencia (Mañana y Tarde) y el desempeño en la asignatura de Lengua para estudiantes de cuarto grado. El resultado fue un p-value de 0.0946, lo que nos dice que no hay evidencia suficiente para rechazar la hipótesis nula de independencia entre las variables.

Además, se calculó el coeficiente de asociación de Cramer V, obteniendo un valor de 0.10. Este valor sugiere una asociación débil entre las variables. También se calcularon los residuos los cuales son relativamente chicos, lo que sugiere que las frecuencias observadas y esperadas son bastante similares.

Tablas de contingencia según el turno en Secundaria

Se crearon tablas de contingencia para la variable desempeño según el turno en la asignatura de Matemática para los alumnos de segundo año.

Porcentaje Total

	Distinguido	Muy Bueno	Bueno	Aprobado	Regular	Insuficiente	Total
MAÑANA	6.2	6.0	7.0	10.9	14,5	3.0	47.6
TARDE	4.9	5.8	8.4	11.3	13.1	2.1	45.7
VESPERTINO	0.2	0.5	0.6	1.8	3.2	0.4	6.7

TOTAL	11.4	12.3	15.9	24.0	30.8	5.5	100.0
-------	------	------	------	------	------	-----	-------

El porcentaje de la población total del turno mañana es el 47.6%. El porcentaje de la población total que ocupa cada Desempeño es 11,4% Distinguido, 12,3% Muy Bueno, 15.9% Bueno, 24.0% Aprobado, 30,8% Regular y 5.5% Insuficiente. El porcentaje de la población total tiene como desempeño Muy Bueno y es de turno vespertino es el 0.5%.

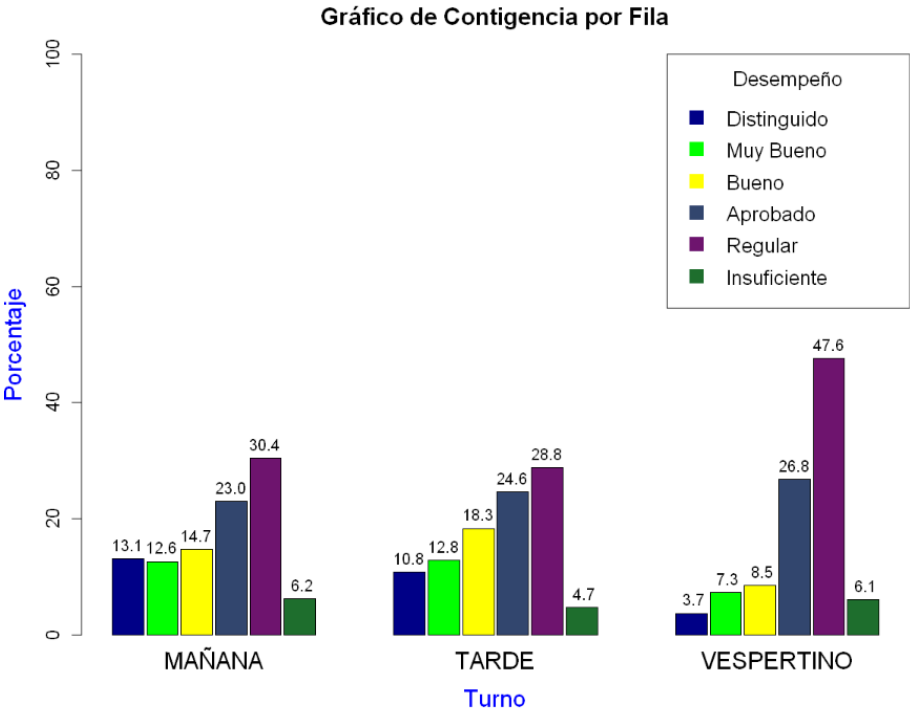
Porcentaje Fila

	Distinguido	Muy Bueno	Bueno	Aprobado	Regular	Insuficiente	Total
MAÑANA	13.1	12.6	14.7	23.0	30.4	6.2	100
TARDE	10.8	12.8	18.3	24.6	28.8	4.7	100
VESPERTINO	3.7	7.3	8.5	26.8	47.6	6.1	100
TOTAL	11.4	12.3	15.9	24.0	30.8	5.5	100

Del turno vespertino el 3,7% son de desempeño Distinguido, el 7.3 de Muy Bueno, el 8.5 de Bueno, el 26.8 de Aprobado, el 47.6 de Regular y el 6.1 de Insuficiente. Del turno tarde el 10.8% son de desempeño Distinguido, el 12.8 de Muy Bueno, el 18.3 de Bueno, el 24.6 de Aprobado, el 28.8 de Regular y el 4.7 de Insuficiente. Por último, del turno mañana, el 13.1% es de desempeño distinguido el 12.6 de Muy Bueno, el 14.7 de Bueno, el 23.0 de Aprobado, el 30.4 de Regular y el 6.2 de Insuficiente.

Como se puede apreciar en el gráfico de contingencia por fila, el turno con mayor porcentaje de alumnos "Regular" es el vespertino. También que este turno es el que menor porcentaje de alumnos con desempeño "Distinguido" tiene.

Figura 7: Gráfico de barras adosadas por fila por desempeño y turno en Matemática de segundo año, departamento Uruguay Secundaria.



Fuente: Elaboración propia (2024).

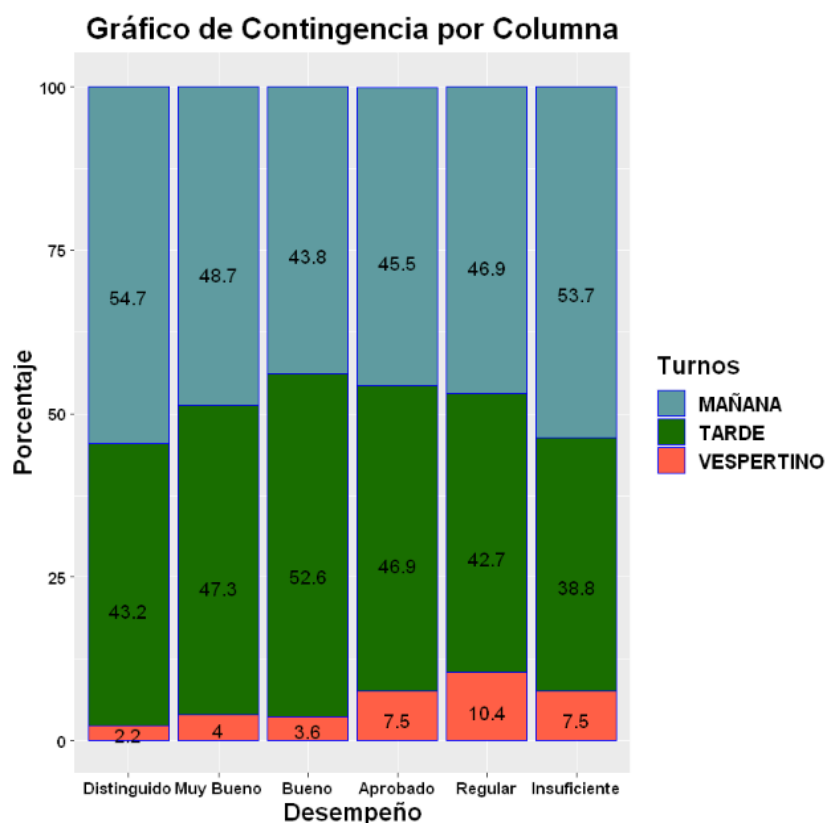
Porcentaje Columna

	Distinguido	Muy Bueno	Bueno	Aprobado	Regular	Insuficiente	Total
MAÑANA	54.7	48.7	43.8	45.5	46.9	53.7	47.6
TARDE	43.2	47.3	52.6	46.9	42.7	38.8	45.7
VESPERTINO	2.2	4.0	3.6	7.5	10.4	7.5	6.7
TOTAL	100	100	100	100	100	100	100

Como se puede apreciar en la tabla el 54.7% de los de desempeño “Distinguido” van a turno mañana, el 43.2% van al turno tarde y el 2.2% al turno vespertino. El 46.9% de los de desempeño “Regular” van al turno mañana, el 42.7 al turno tarde y el 10.4 al vespertino. El desempeño que tiene mayor proporción en turno vespertino es el “Regular” y el que tiene mayor proporción en turno mañana es el “Distinguido”.

Se realizó un gráfico de columnas, en el mismo se puede apreciar que el desempeño “Distinguido” es el que tiene mayor proporción en turno mañana y el desempeño “Regular” en el turno vespertino.

Figura 8: Gráfico de barras apiladas por columna por desempeño y turno en Matemática de segundo año, departamento Uruguay Secundaria.



Fuente: Elaboración propia (2024).

Se realizó la prueba de Chi-cuadrado de independencia, el p-valor dio 0.009523 que es un valor pequeño, también se calcularon las frecuencias observadas y esperadas y dieron que las mismas tenían diferencias notables. El desempeño de los estudiantes en Matemática parece depender del turno en el que asisten.

Se calcularon también los residuos estandarizados y los residuos ajustados que determinaron la misma conclusión del test Chi-cuadrado: existe una asociación significativa entre el los turnos y el desempeño de los estudiantes. Los mayores residuos (positivos y negativos) se encuentran en el turno vespertino, específicamente en las categorías “Regular” y “Distinguido”, indicando una discrepancia significativa entre las frecuencias observadas y esperadas para estos desempeños en ese turno.

Se calculó el coeficiente de Cramer que dio un valor de 0.09, la asociación entre el desempeño y el tipo de escuela es débil. Esto significa que, aunque hay una relación entre el turno y el desempeño, esta no es fuerte. Por último, también se realizó un análisis de diferencias de las proporciones, el valor p fue 3.753e-07. Esto sugiere que hay diferencias significativas entre las proporciones de los desempeños respecto a los turnos.

Promedios de los alumnos de primer grado del turno mañana, en las materias troncales

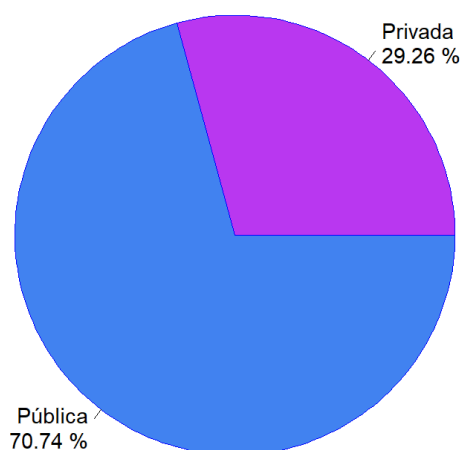
Se realizó el test de Friedman para responder a la pregunta: ¿Hay diferencias en los promedios de los alumnos de primer año de la escuela de turno mañana en Lengua, Matemática o Ciencias Sociales? El p-valor obtenido fue $2.2e-16$, lo que indica que hay diferencias significativas en los promedios de las asignaturas de Ciencias Sociales, Lengua y Matemática en la escuela de turno mañana en primer grado. La potencia del test calculado es 0.99.

Comparación de los promedios de escuelas Públicas y Privadas

Para obtener la cantidad de alumnos que asisten a escuelas públicas y privadas se decidió realizar tablas de frecuencias. En el caso de primaria, la mayoría de los niños asisten a escuelas públicas, representando el 70,74% del total de estudiantes. En contraste, el 29,26% de los estudiantes asisten a escuelas privadas. En secundaria también la cantidad de estudiantes que asisten a escuelas públicas son más que la mitad que los que asisten a la privada, siendo el 79,54% del total.

Figura 9: Gráfico de torta de las distribuciones de alumnos en escuelas públicas y privadas del nivel primario del departamento Uruguay.

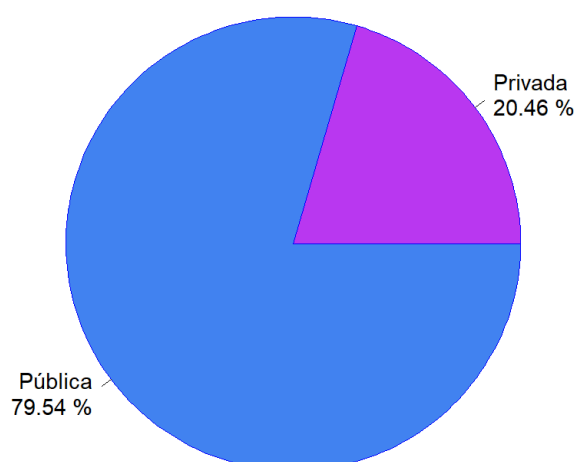
Distribución de Escuelas Públicas y Privadas Nivel Primario



Fuente: Elaboración propia (2024).

Figura 10: Gráfico de torta de las distribuciones de alumnos en escuelas públicas y privadas del nivel secundario del departamento Uruguay.

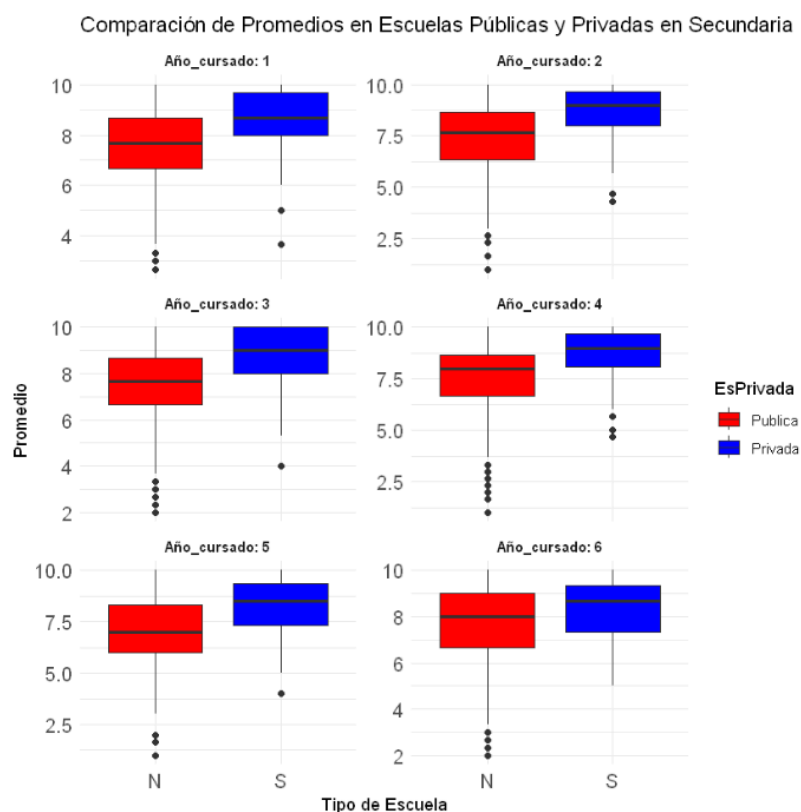
Distribución de Escuelas Públicas y Privadas Nivel Secundario



Fuente: Elaboración propia (2024).

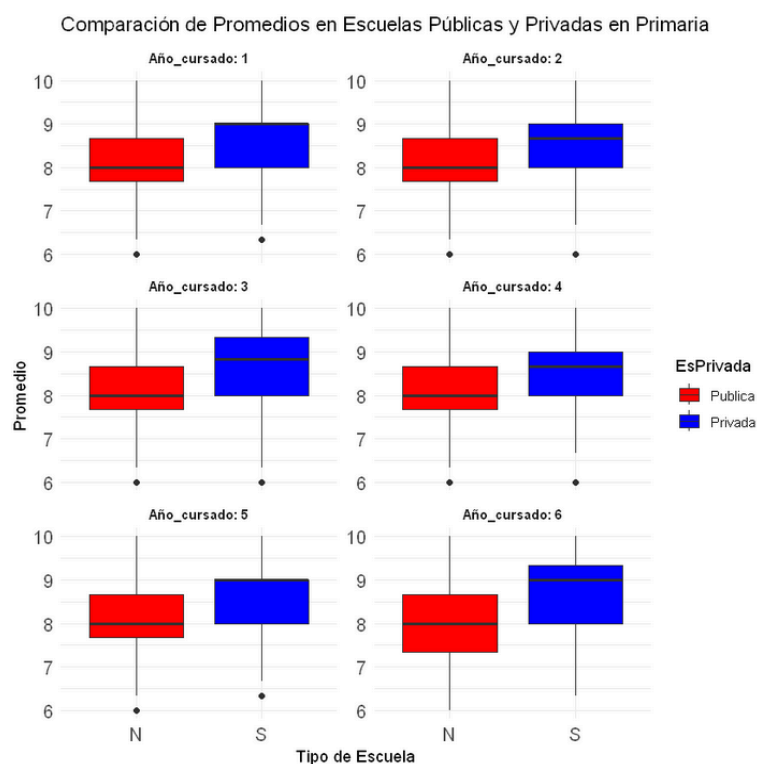
Para averiguar si las medianas de los promedios de los estudiantes en escuelas públicas y privadas son iguales, primero se compararon las varianzas de ambos grupos y se encontró que en primaria las varianzas son diferentes en primer, tercer y sexto grado, mientras que en secundaria eran iguales en todos los años. Sin embargo, al aplicar el test de Kolmogorov-Smirnov, se verificó que los datos no seguían una distribución normal. Por esta razón, se utilizó el test de Wilcoxon. Los resultados mostraron que en todos los años había diferencias significativas entre los promedios de los estudiantes de escuelas públicas y privadas. Además, los boxplots realizados para todos los años confirmaron que, en cada año, los estudiantes de las escuelas privadas tenían promedios más altos.

Figura 11: Gráfico agrupado de boxplots sobre los promedios en las escuelas públicas y privadas, departamento Uruguay Secundaria.



Fuente: Elaboración propia (2024).

Figura 12: Gráfico agrupado de boxplots sobre los promedios en las escuelas públicas y privadas, departamento Uruguay primaria.



Fuente: Elaboración propia (2024).

Tablas de contingencia según el tipo de Gestión en Primaria

Se crearon tablas de contingencia para la variable desempeño según el tipo de gestión en la asignatura de Lengua para los estudiantes de cuarto grado.

Porcentaje Total

	insuficiente	regular	aprobado	bueno	muy bueno	distinguido	Total
Privada	0.2	0.7	1.2	2.0	5.7	11.1	28.8
Pública	1.7	9.5	13.2	18.1	16.1	12.6	71.2
Total	1.9	10.7	15.2	23.8	27.2	21.2	100.0

En general se puede ver que la mayoría de los estudiantes de cuarto grado en Lengua se encuentran en "Muy Bueno" (27.2%) y "Distinguido" (21.2%), lo que indica un buen desempeño.

Las escuelas privadas tienen un porcentaje total mucho menor de estudiantes en las categorías de desempeño más bajo (0.2% insuficiente y 0.7% regular) en comparación con las escuelas públicas (1.7% insuficiente y 9.5% regular), además de tener una mayor proporción de estudiantes en la categoría de "Distinguido".

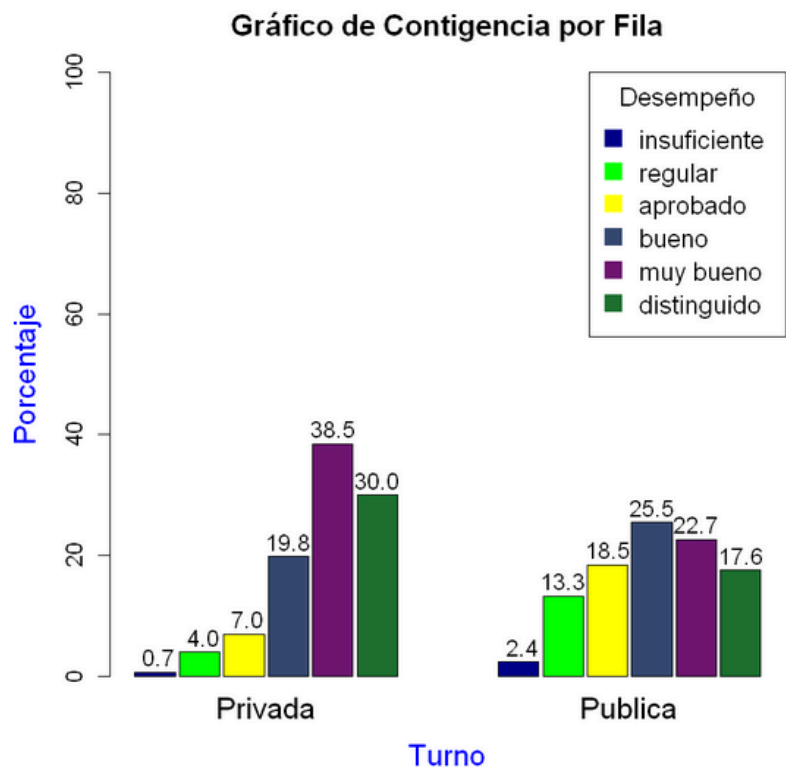
Porcentaje Fila

	insuficiente	regular	aprobado	bueno	muy bueno	distinguido	Total
Privada	0.7	4.0	7.0	19.8	38.5	30.0	100.0
Pública	2.4	13.3	18.5	25.5	22.7	17.6	100.0
Total	1.9	10.7	15.2	23.8	27.2	21.2	100.0

En las escuelas privadas el 0.7% son de categoría "Insuficiente", el 4% "Regular", 7% Aprobado, 19.8% "bueno", 38.5% "Muy Bueno" y 30% "Distinguido".

En las escuelas públicas el mayor porcentaje se encuentra en "Bueno" (25.5%), seguido por "Muy Bueno" (22.7%) y "Regular" (13.3%).

Figura 13: Gráfico de barras adosadas por fila por desempeño y sector en Lengua de cuarto grado, departamento Uruguay Primaria.



Fuente: Elaboración propia (2024).

Porcentaje Columna

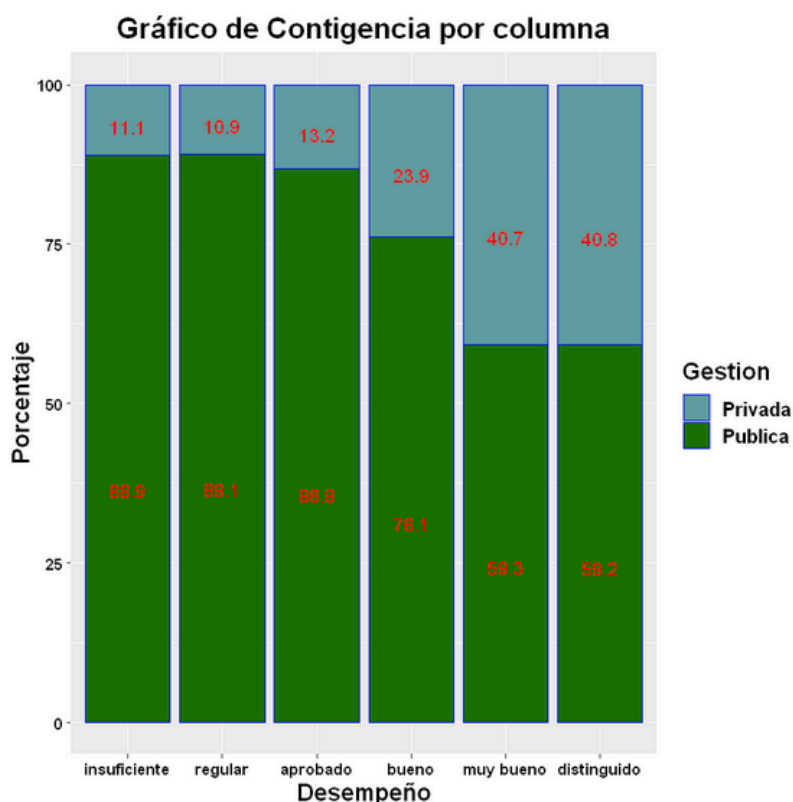
	insuficiente	regular	aprobado	bueno	muy bueno	distinguido	Total
Privada	11.1	10.9	13.2	23.9	40.7	40.8	28.8
Pública	88.9	89.1	86.8	86.8	76.1	59.3	71.2
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Se observa que en la categoría distinguido hay 40.8% en escuelas privadas y un 59.3% en escuelas públicas, en términos proporcionales, un mayor porcentaje de estudiantes en escuelas públicas son clasificados como "Distinguido" en comparación con las escuelas privadas.

Esta diferencia en la distribución sugiere que aunque las escuelas privadas tienen una proporción significativa de estudiantes en la categoría "Distinguido", las escuelas públicas tienen un porcentaje aún mayor de estudiantes en esta categoría en relación con su número total de estudiantes.

Se calculó un test de Chi-cuadrado obteniendo un p-value de $9.228e-14$, lo que indica que el tipo de gestión de la escuela y el desempeño en Lengua parecen estar relacionados de manera significativa. Por lo que se procedió a realizar el test de Cramer V, el coeficiente de asociación de Cramer es 0.27, lo que indica una asociación leve/moderada entre las variables.

Figura 14: Gráfico de barras apiladas por columna por desempeño y sector en Lengua de cuarto grado, departamento Uruguay Primaria.



Fuente: Elaboración propia (2024).

Tablas de contingencia según el tipo de Gestión en Secundaria

Se crearon tablas de contingencia para la variable desempeño según el tipo de gestión en la asignatura de Matemática para los alumnos de segundo año.

Porcentaje Total

	Distinguido	Muy Bueno	Bueno	Aprobado	Regular	Insuficiente	Total
Privada	2.9	2.8	2.9	2.6	2.6	0.4	14.2
Pública	9.4	10	13.2	21.1	27.4	4.8	85.8
Total	12.3	12.8	16.1	23.7	30.0	5.2	100

El porcentaje del total de los alumnos que es de escuela privada es el 14,2%. El porcentaje de la población total que ocupa cada desempeño es 12,3% "Distinguido", 12,8% "Muy Bueno", 16.1% "Bueno", 23.7% "Aprobado", 30% "Regular" y 5.2% "Insuficiente". El porcentaje de la población total tiene como desempeño "Distinguido" y es de escuela Privada es el 2.9%.

Porcentaje Fila

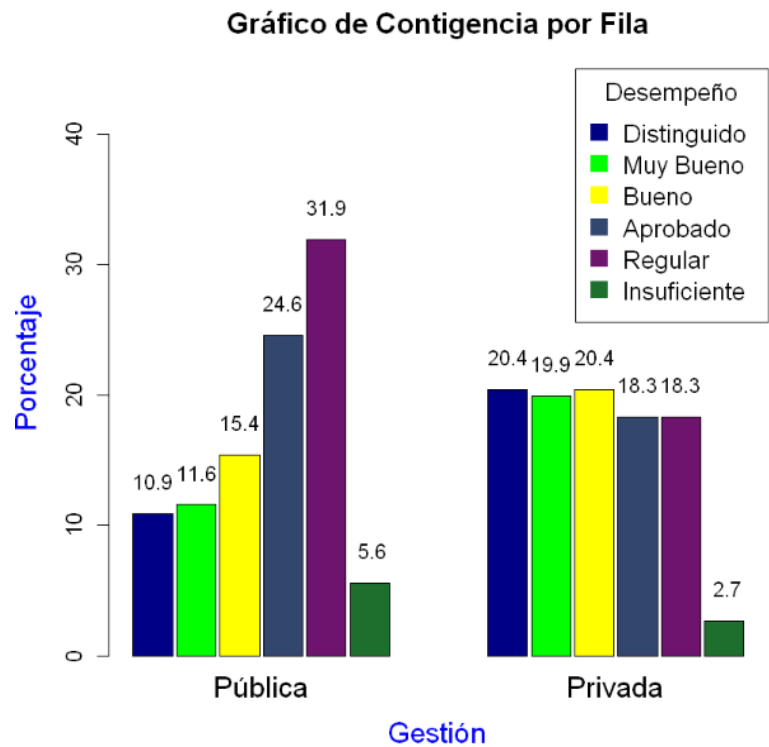
	Distinguido	Muy Bueno	Bueno	Aprobado	Regular	Insuficiente	Total
Privada	20.4	19.9	20.4	18.3	18.3	2.7	100
Pública	10.9	11.6	15.4	24.6	31.9	5.6	100
Total	12.3	12.8	16.1	23.7	30.0	5.2	100

El 20,4% de la escuela Privada son de desempeño "Distinguido", el 19.9 de "Muy Bueno", el 20.4 de Bueno, el 18.3 de "Aprobado", el 18.3 de Regular y el 2.7 de Insuficiente. Los de escuela pública el 10.9% son de Distinguido, el 11.6 de "Muy Bueno", el 15.4 de "Bueno", el 24.6 de "Aprobado", el 31.9 de "Regular" y el 5.6 de "Insuficiente".

Se realizó un gráfico de contingencia por filas, se puede observar que en el caso de las escuelas públicas la categoría de desempeño "Regular" tiene el mayor porcentaje del total, indicando que una mayor proporción de estudiantes en estas instituciones obtienen calificaciones en este rango.

Por otro lado, en las escuelas privadas, no hay una categoría de desempeño que predomine claramente sobre las demás. Los porcentajes de desempeño son similares en todas las categorías, lo que sugiere una distribución más equilibrada de los resultados académicos en comparación con las escuelas públicas.

Figura 15: Gráfico de barras adosadas por fila por desempeño y sector en Matemática de segundo año, departamento Uruguay Secundaria.



Fuente: Elaboración propia (2024).

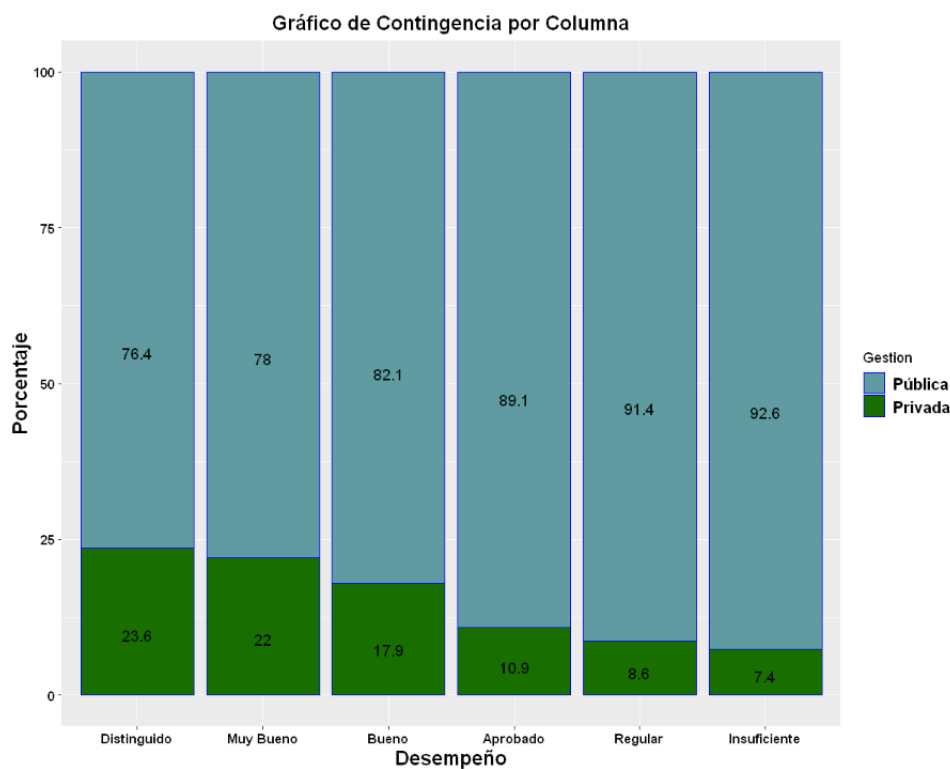
Porcentaje Columna

	Distinguido	Muy Bueno	Bueno	Aprobado	Regular	Insuficiente	Total
Privada	23.6	22	17.9	10.9	8.6	7.4	14.2
Pública	76.4	78	82.1	89.1	91.4	92.4	85.8
Total	100	100	100	100	100	100	100

Como se observa en la tabla, el 23.6% de los de desempeño “Distinguido” van a escuela Privada mientras que el 76.4% no. El 7.4% de los de desempeño “Distinguido” van a escuela Privada mientras que el 92.4% no. El desempeño que tiene mayor proporción en escuela pública es el “Insuficiente”.

Se realizó un gráfico de columnas y el mismo se ve escalonado, lo que indica que hay una relación de asociación entre las variables. En el mismo se puede apreciar que el desempeño. “Insuficiente” es el que tiene mayor proporción en escuela Pública y el desempeño “Distinguido” en escuela Privada.

Figura 16: Gráfico de barras apiladas por columna por desempeño y sector en Matemática de segundo año, departamento Uruguay Secundaria.



Fuente: Elaboración propia (2024).

Se realizó la prueba de Chi-cuadrado de independencia, el p-valor dio $2.2e-16$, que es un valor extremadamente pequeño, también se calcularon las frecuencias observadas y esperadas y dieron que las mismas tenían diferencias notables. Esto permitió establecer que hay una asociación significativa entre Desempeño y EsPrivada.

Se calcularon también los residuos estandarizados y los residuos ajustados que reforzaron la conclusión del test Chi-cuadrado: existe una asociación significativa entre el tipo de escuela (privada o no) y el desempeño de los estudiantes. Los residuos indican que las escuelas privadas tienen una mayor proporción de estudiantes con desempeño "Distinguido" y una menor proporción con desempeños más bajos, mientras que las escuelas no privadas tienen más estudiantes con desempeños "Aprobado" y "Regular".

Se calculó el coeficiente de Cramer que dio un valor de 0.2, la asociación entre el desempeño y el tipo de escuela se encuentra en el rango débil a moderada, más cercano a moderada. Esto significa que, aunque hay una relación entre el tipo de escuela y el

desempeño, esta relación no es muy fuerte y otros factores también pueden influir significativamente en el desempeño de los estudiantes. Por último, también se realizó un análisis de diferencias de las proporciones, el valor p fue 3.753e-07 esto indica que hay evidencia suficiente para afirmar que hay diferencias significativas en las proporciones de desempeños respecto al tipo de gestión.

Comparación del desempeño de escuelas públicas y privadas de estudiantes de primer año de Secundaria

Se quiso responder la siguiente pregunta: ¿Existe una diferencia entre los promedios de las escuelas públicas y privadas en las asignaturas troncales? Para responderla, se quería usar el test de manova multivariante, para ello se comprobó si se cumplían los supuestos. Primero se calculó la normalidad de las escuelas públicas y todas las asignaturas excepto Lengua y Matemática no dieron normales. En el caso de las escuelas privadas, si todas las materias cumplieron con el supuesto de normalidad. Luego se calculó la normalidad multivariante y el resultado del test en la prueba de Mardia-Asimetría dio un p-value= 7.364035e-06 y la Prueba de Mardia - Curtosis un p-value de 1.42327e-05. Por esto se concluye que se rechaza la hipótesis nula de normalidad multivariada en términos de asimetría y en términos de curtosis. Como no se cumplió con los supuestos de normalidad y la normalidad multivariante se concluye que no se pudo aplicar el test de Manova.

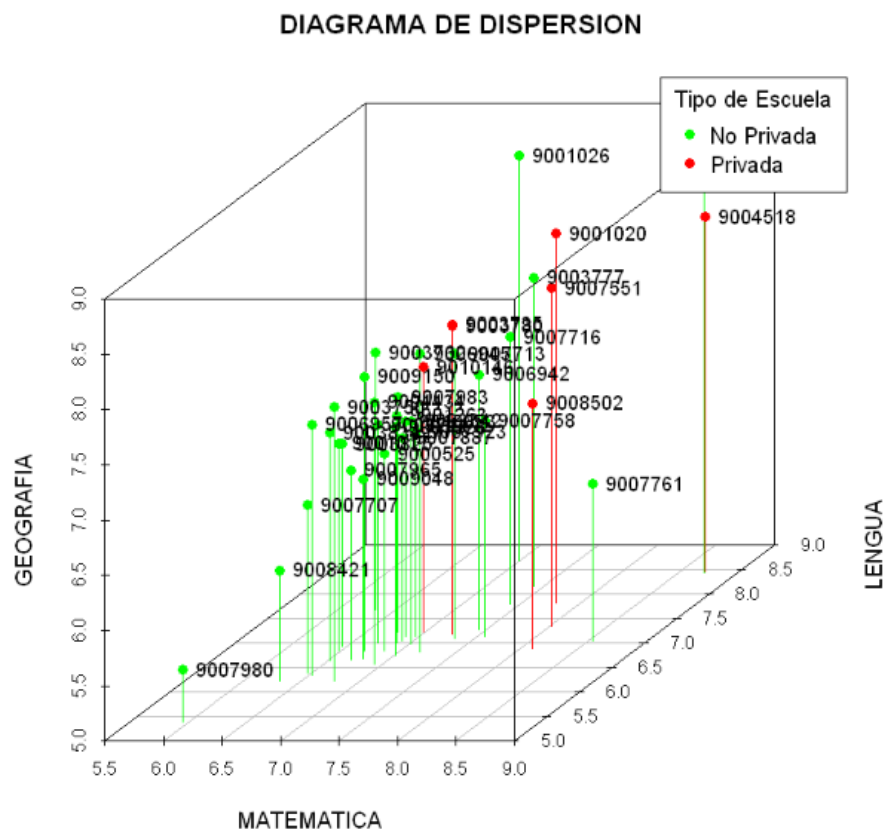
Para determinar si había outliers en los datos se procedió a calcularlos de dos formas diferentes. Por un lado, se calculó las distancias de Mahalanobis y se determinó si los datos son outliers en el conjunto de datos promedios_escuelas. El resultado fue que la mayoría de las observaciones no son outliers. Sólo la décima observación ha sido detectada como un outlier.

Luego se utilizaron tres métodos diferentes para estimar si hay outliers multivariados: el método Minimum Covariance Determinant (MCD), el método Minimum Volume Ellipsoid (MVE) y el método clásico. Los resultados obtenidos muestran que la mayoría de las observaciones tienen distancias de Mahalanobis relativamente bajas, lo que sugiere que están cerca de la media multivariada del conjunto de datos. Sin embargo, ciertas observaciones presentan distancias significativamente mayores, lo que indica que son potencialmente valores atípicos (outliers) notablemente altos en comparación con las demás observaciones. Son las observaciones en las filas 11, 31, 26, 33. Para obtener los CUE de las escuelas consideradas outliers: 9003762, 9007980, 9007761 y 9008421 se realizó el Local Outlier Factor que se muestran en la siguiente tabla.

CUE	EsPrivada	GEOGRAFÍA	HISTORIA	LENGUA	MATEMÁTICA
9003762	N	7.33	4.67	7.67	6.33
9007980	N	5.47	5.42	5.39	5.94
9007761	N	6.42	6.89	7.04	8.53
9008421	N	6.00	6.38	6.22	6.31

En el siguiente gráfico, se puede apreciar la distribución de las escuelas según sus promedios en las asignaturas troncales separadas por tipo de gestión.

Figura 17: gráfico 3D.



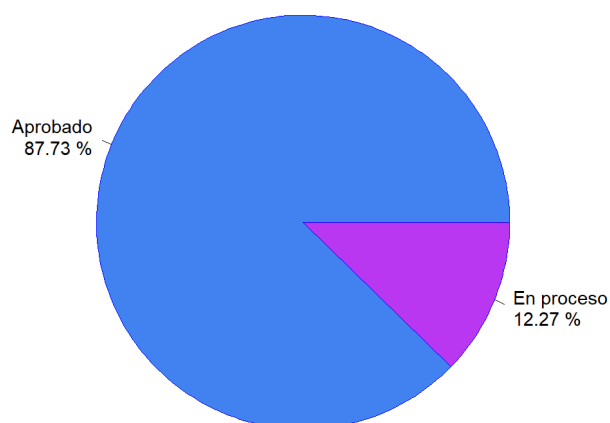
Fuente: Elaboración propia (2024).

Análisis del desempeño general en las asignaturas Lengua y Matemática

El objetivo fue obtener una idea general del desempeño de los alumnos en dos materias troncales importantes: Matemática y Lengua. Para esto consideramos realizar tablas de frecuencias y, en el caso de primaria, los resultados reflejaron la cantidad de estudiantes aprobados y de los que siguen en proceso de aprendizaje, siendo Matemática la que tenía mayor porcentaje de aprobados y Lengua con el mayor porcentaje de estudiantes en proceso de aprendizaje. Mientras que en el secundario, el resultado reflejó el total de alumnos aprobados y los que tuvieron que rendir, habiendo mayor porcentaje de aprobados en Lengua.

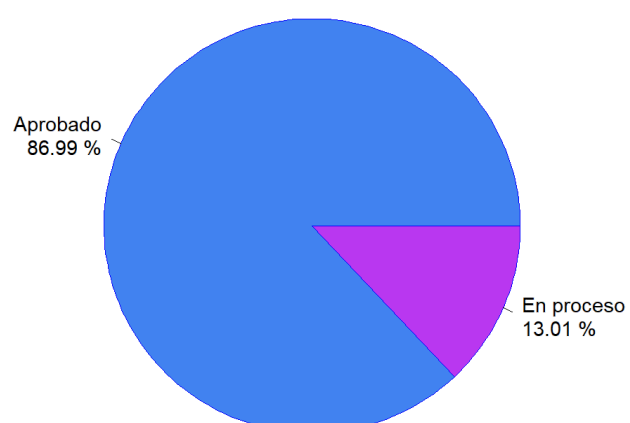
Figura 18 y 19: Gráficos de torta de las distribuciones de Resultados del nivel primario en Matemática y Lengua tomando en cuenta todos los grados del departamento Uruguay.

Distribución Resultado general en Matematica, Primaria



Fuente: Elaboración propia (2024).

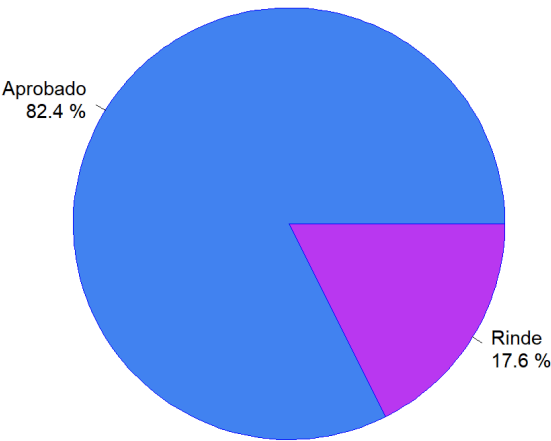
Distribución Resultado general en Lengua, Primaria



Fuente: Elaboración propia (2024).

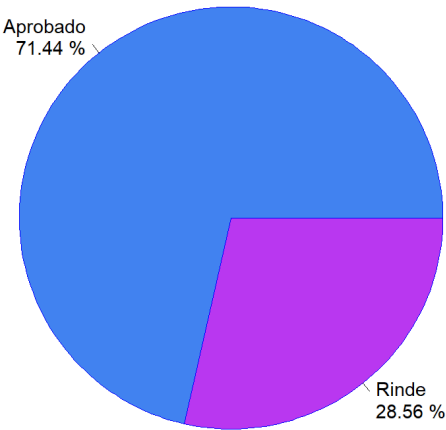
Figura 20 y 21: Gráficos de torta de las distribuciones de Resultados del nivel secundario en Matemática y Lengua del departamento Uruguay.

Resultado general en Lengua, Secundaria



Fuente: Elaboración propia (2024)

Resultado general en Matemática, Secundaria



Fuente: Elaboración propia (2024).

Después de realizar un análisis general, se avanzó hacia el cálculo de medidas estadísticas para evaluar el desempeño académico de los estudiantes. Inicialmente, el enfoque estuvo en las asignaturas fundamentales como lo son Matemática y Lengua, tomando los promedios de los estudiantes en ambas áreas. Sin embargo, también se consideraron incluir en el análisis y en cálculos de correlaciones otras asignaturas, como Educación Física, Geografía, Biología e Historia que no fueron estudiadas con tanta profundidad como Lengua y Matemática, pero que aun así aportaron información al análisis.

Se calculó el vector de medias para todos los años del nivel secundario de las materias troncales: Lengua, Matemática, Historia y Geografía.

Año Cursado	GEOGRAFÍA	HISTORIA	LENGUA	MATEMÁTICA
1	7.24	7.34	7.08	7.04
2	7.05	7.27	6.89	6.59
3	7.16	7.30	7.07	6.64
4	7.26	7.46	7.19	6.81
5	7.60	7.63	7.62	6.94
6	7.87	7.83	7.71	7.13

Se procedió a realizar la matriz de varianza y covarianza para estas asignaturas. En todos los años, Matemática tiene la mayor varianza. Las covarianzas varían entre los años, pero frecuentemente, hay una relación moderada entre Matemática y otras materias, especialmente Geografía e Historia. Aunque hay variabilidad en los puntajes de Matemática, hay también una tendencia a que los puntajes de Matemática estén relacionados con los puntajes de otras materias. A continuación se muestra la Matriz de covarianza de primer año, donde se puede observar que la asignatura con mayor varianza es Matemática. También se puede determinar que Geografía e Historia o Geografía y Lengua están muy correlacionadas.

	GEOGRAFÍA	HISTORIA	LENGUA	MATEMÁTICA
GEOGRAFÍA	2.26	1.86	1.79	1.89
HISTORIA	1.86	2.31	1.75	1.83
LENGUA	1.79	1.75	2.05	1.80
MATEMÁTICA	1.89	1.83	1.80	2.62

Se quiso responder la siguiente pregunta: ¿Existe una diferencia entre los promedios de las diferentes asignaturas de todos los años del secundario? Para responderla, se quería usar el test de anova multivariante, para ello se comprobó si se cumplían los supuestos. Para todas las combinaciones de Año_cursado, los valores p son significativamente menores a 0.05. Luego se calculó la normalidad multivariante y el resultado del test en la prueba de Mardia-Asimetría dio un p-value= 1.077147e-52 y la Prueba de Mardia - Curtosis un p-value de 3.500759e-10. Por esto se concluye que se rechaza la hipótesis nula de normalidad multivariada en términos de asimetría y en términos de curtosis. Como no se cumplieron los supuestos de normalidad y normalidad multivariada no se pudo aplicar el test.

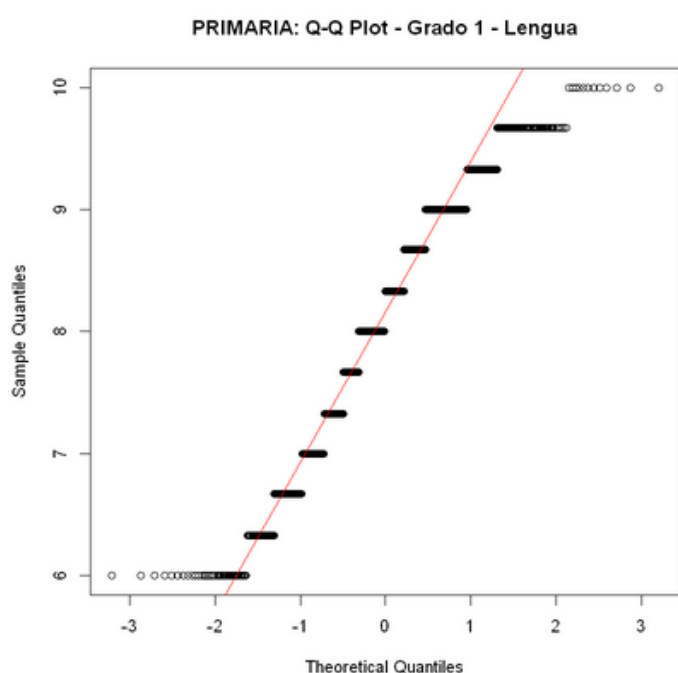
Análisis de la variable Promedio

Se procedió a estudiar en mayor profundidad la variable Promedio, ya que es una variable importante que nos ayuda a determinar en gran medida el desempeño en las asignaturas. Para ellos se estudió la normalidad, se procedió a estandarizar la variable y encontrar las puntuaciones Z. Para estudiar la normalidad se realizaron gráficos Q-Q plot de las asignaturas de Lengua y Matemática. Al observarlos, pareciera que la distribución de la variable Promedio podría ser normal. Sin embargo, para confirmar esta suposición, se

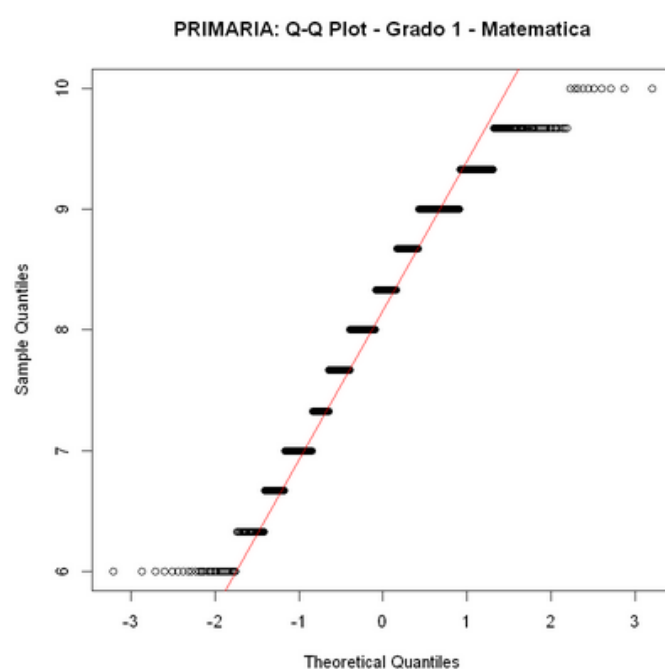
realizó el test de normalidad para estas variables, concretamente el de Kolmogorov-Smirnov. Los resultados del test indicaron que, en todos los años, la distribución de la variable no se ajusta a los parámetros considerados normales.

Procedimos a estandarizar el promedio en ambas asignaturas, al comparar los resultados de la estandarización de las puntuaciones en Matemática y Lengua en el nivel secundario, podemos ver que ambas asignaturas tienen una media estandarizada de cero. La dispersión de los datos, como se indica por los cuartiles y el rango intercuartil, es diferente entre Matemática y Lengua. Por ejemplo, la dispersión de las puntuaciones estandarizadas de Matemática parece ser un poco más amplia que la de Lengua, ya que los valores del primer y tercer cuartil son mayores en Matemática que en Lengua. Es decir, que un estudiante con una puntuación estandarizada de 1 en Matemática estaría un poco más por encima de la media que un estudiante con una puntuación estandarizada de 1 en Lengua, ya que la dispersión de las puntuaciones estandarizadas de Matemática es un poco mayor.

Figura 22 y 23: Gráficos Q-Q plot de los promedios de la asignatura Lengua y Matemática del Departamento Uruguay primaria.

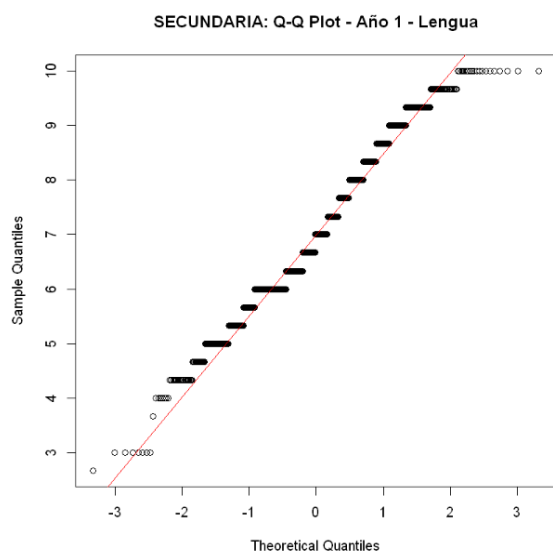


Fuente: Elaboración propia (2024).

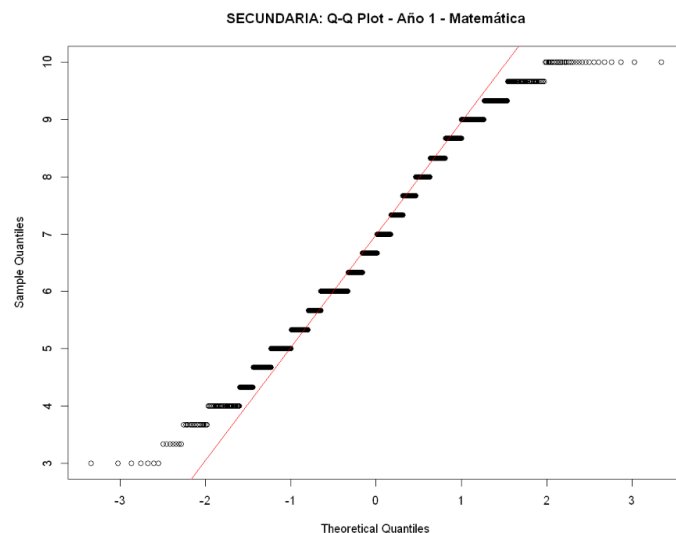


Fuente: Elaboración propia (2024).

Figura 24 y 25: Gráficos Q-Q plot de los promedios de la asignatura Lengua y Matemática del Departamento Uruguay secundaria.



Fuente: Elaboración propia (2024).



Fuente: Elaboración propia (2024).

Comparación del desempeño de un estudiante respecto a diferentes grupos de estudiantes

Para comparar el desempeño de un estudiante que se sacó un 4 en Matemática del nivel secundario con los resultados de estudiantes que asisten a una escuela secundaria pública común se comparó el promedio de estos alumnos con el del estudiante que sacó 4, como este es 6.992, el desempeño de este alumno particular se considera por debajo del promedio de estudiantes de tercero de escuelas públicas. También, como el promedio de 4 se encuentra por debajo del primer cuartil, significa que está en el 25% inferior de los promedios de notas. En el caso de la secundaria privada común, el promedio para la asignatura Matemática de tercer año es 6.992 y el valor del primer cuartil es 6. Esto quiere decir que su nota está por debajo del 25% de las notas y por lo mismo, es menor que el promedio. Por último, para la escuela secundaria técnica, como el promedio es 6.301, podemos decir que un promedio igual a 4 está por debajo del promedio y también debajo del 25% de las notas. Por otro lado, para la escuela agrotécnica, como el promedio de las notas es 6.670 y el valor del primer cuartil es 6.247 podemos decir que una nota igual a 4 está por debajo del promedio e inclusive del 25% de los datos.

Promedios de los alumnos de primer año de la escuela agrotécnica, en las materias troncales

Se realizó el test de Friedman para responder a la pregunta: ¿Hay diferencias en los promedios de los alumnos de primer año de la escuela agrotécnica, en Lengua, Matemática o Geografía? Para responder se realizó el test de Friedman el p-value obtenido fue 0.05772 lo que sugiere que no hay evidencias para afirmar que existen diferencias significativas en el desempeño de los alumnos en las asignaturas de Geografía, Lengua y Matemática en la escuela agrotécnica. La potencia del test calculado es 0.877.

Comparación del Desempeño de un Estudiante en Matemática en Escuelas con Distintas Poblaciones: Un Enfoque a través del Puntaje Z.

Se decidió además comparar el desempeño de este estudiante que obtuvo cuatro en Matemática con las notas de Matemática del primer año en dos escuelas con notorias diferencias en la cantidad de estudiantes. En la escuela con mayor cantidad de estudiantes, el puntaje Z de este alumno dió aproximadamente -1.35. Esto indica que una calificación de 4 está alrededor de 1.35 desviaciones estándar por debajo del promedio de las calificaciones de los estudiantes de esa escuela en Matemática. En la escuela con menor cantidad de estudiantes de Matemática, el puntaje Z fue de aproximadamente -2.29, lo que estaría indicando que una calificación de 4 está alrededor de 2.29 desviaciones estándar por debajo de la media de las calificaciones de los estudiantes de esa escuela en Matemática.

Promedios de los alumnos de primer grado/año de la escuela con menor cantidad de estudiantes

Se realizó el test de Cochran tanto en primaria como en secundaria para saber si las proporciones de las notas en Matemática de los alumnos de primer grado/año son iguales en todos los trimestres de la escuela con menor cantidad de estudiantes. En primaria el p-value dio 0.3679 y en secundaria dio 0.05393 lo que indica que no hay diferencias significativas entre las proporciones de los trimestres. Ambos con una potencia de 0.95.

Análisis de los promedios de las asignaturas Lengua y Matemática en Primaria

En el caso de la asignatura de Matemática, de primero a sexto año la media y la mediana ronda en el valor 8, lo que cambia es la moda, segundo, quinto y sexto tienen una

moda de 8, mientras que los otros años una moda de 9. Aunque la moda no coincide en todos, la mediana y media son valores cercanos esto quiero decir que en general los datos se distribuyen de forma simétrica. En todos los casos el valor de la media es representativa.

El porcentaje general de aprobados en esta asignatura es del 87,73% y 12.27% de los estudiantes se encuentran en proceso de aprendizaje. El grado con mayor porcentaje de estudiantes en proceso es quinto grado con un 14.46% mientras que el de menos es cuarto con un 10.91%. Se procedió a realizar un test de Wilcoxon para saber si la mediana poblacional de los promedios era mayor a 7. El resultado del test refleja que en todos los años el rendimiento es superior.

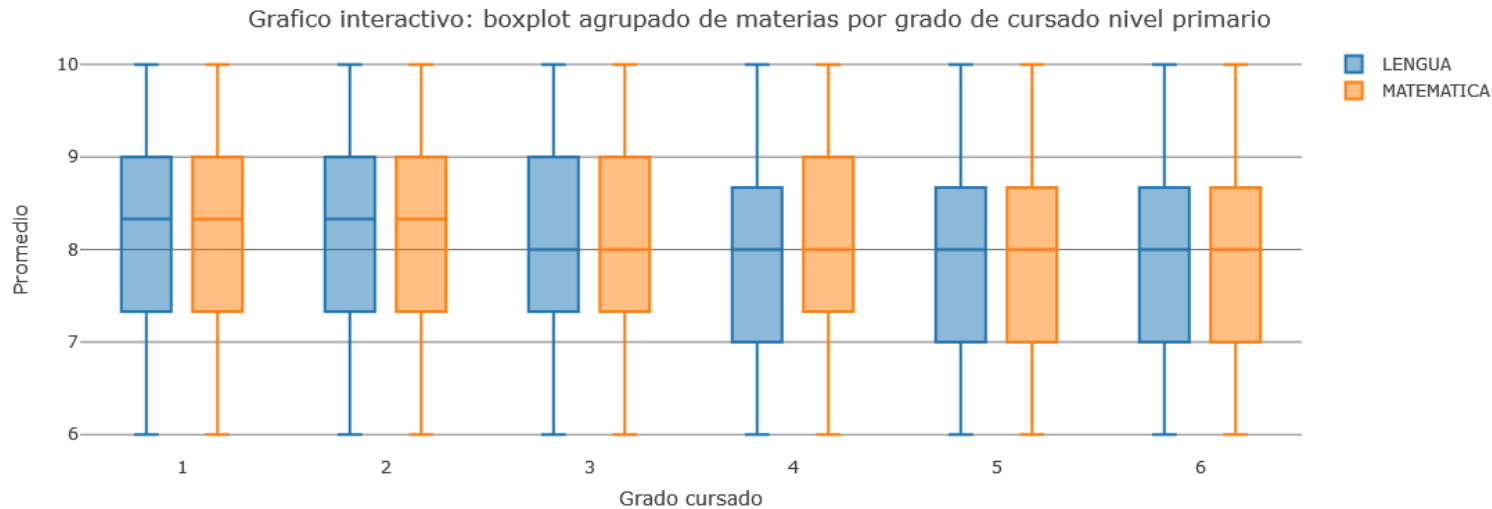
Si analizamos los datos de la asignatura Lengua, ocurre lo mismo que en Matemática, la media, mediana y rondan el valor 8, mientras que la moda varía de 9 a 8, aunque no son valores exactamente iguales son bastante cercanos, lo que refleja una distribución simétrica. Los datos están agrupados alrededor de la media en los seis años así que se la considera representativa.

El porcentaje general de aprobados en esta asignatura es del 86.99%. El grado con mayor porcentaje de estudiantes en proceso es segundo grado con un 15.55% mientras que el de menos es sexto con un 11.13%. Se procedió a realizar un test de Wilcoxon para saber si la mediana poblacional de los promedios era mayor a 7. El resultado del test refleja, igual que en Matemática, que en todos los años el rendimiento es superior.

Observando los valores del desvío estándar de las materias de Lengua y Matemática en primaria, se encontró que la variabilidad es bastante similar en ambas asignaturas en todos los grados. Sin embargo, en sexto grado, la asignatura de Matemática muestra una variabilidad ligeramente mayor (1.12). Por lo tanto, se podría decir que la asignatura de Matemática tiene una mayor variabilidad en las notas, pero la diferencia es muy pequeña. Este análisis es útil ya que nos indica que algunos estudiantes de sexto podrían estar teniendo dificultades en la asignatura de Matemática.

Mirando los diagramas de caja (boxplot) comparativos de las asignaturas de Lengua y Matemática, nos damos cuenta que todas tienen una distribución similar con cajas alargadas lo que nos indica bastante dispersión. En primero y segundo tienen la misma mediana ubicada en 8.33 aproximadamente, ambos con la misma dispersión, luego de tercero a sexto tienen la misma mediana en 8. El único grado en donde los gráficos no coinciden exactamente es en cuarto, donde Matemática parece tener un rendimiento un poco superior ya que tiene menos notas inferiores a 7.67, luego en quinto y sexto se mantienen los mismos valores.

Figura 26: Gráfico de caja agrupados por año y asignaturas en el departamento Uruguay Primaria.



Fuente: Elaboración propia (2024).

Para analizar si las distribuciones de los promedios de Lengua y Matemática son iguales, primero se comprobó que los datos no eran normales y se estudió con el test de Levene si las varianzas tenían una diferencia significativa. Esto dio que en todos los grados menos en primero no presentaron diferencias significativas. Se realizó el test de Wilcoxon-Mann-Whitney este dio como resultado que los promedios difieren en todos los grados.

Tablas de contingencia Desempeño de Lengua y Matemática Primaria

Se crearon tablas de contingencia para la variable desempeño según las asignaturas de Matemática y Lengua para los estudiantes de cuarto grado.

Porcentaje Total

	insuficiente	regular	aprobado	bueno	muy bueno	distinguido	Total
Lengua	0.9	5.5	7.5	11.7	13.6	10.7	49.9
Matemática	0.6	4.9	6.7	12.3	13.7	11.8	50.1
Total	1.5	10.4	14.2	24.0	27.3	22.5	100.0

En Lengua La mayor proporción de estudiantes se encuentra en las categorías de "Muy Bueno" (13.6%) y "Bueno" (11.7%), representando el 49.9% del desempeño total.

En Matemática es similar a Lengua, las categorías más representadas son "Muy Bueno" (13.7%) y "Bueno" (12.3%), representando el 50,1% del desempeño total.

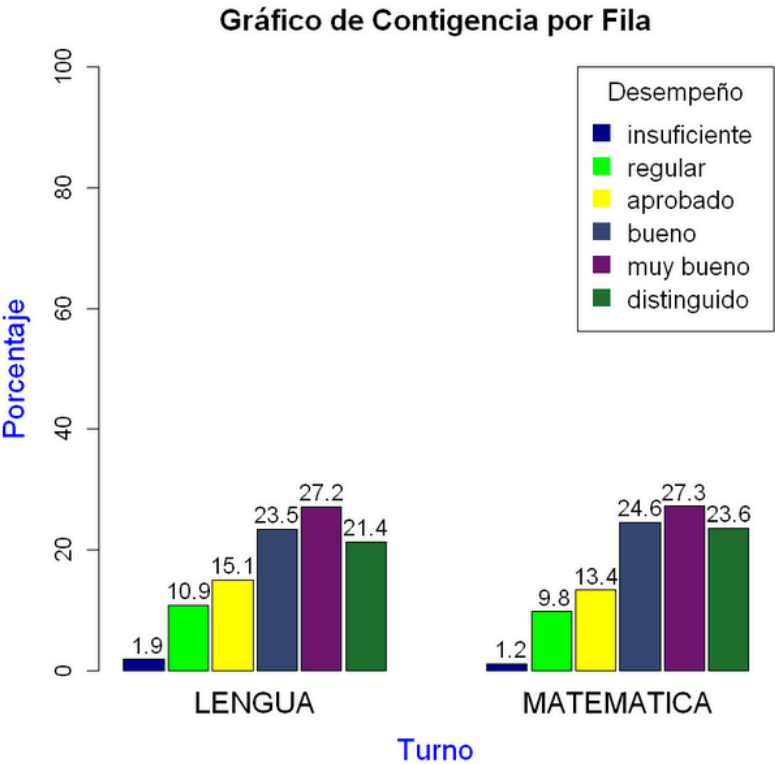
Las categorías de "Insuficiente" y "Regular" son las menos representadas en ambas asignaturas.

Porcentaje Fila

	insuficiente	regular	aprobado	bueno	muy bueno	distinguido	Total
Lengua	1.9	10.9	15.1	23.5	27.2	21.4	100.0
Matemática	1.2	9.8	13.4	24.6	27.3	23.6	100.0
Total	1.5	10.4	14.3	24.0	27.3	22.5	100.0

Tanto en Lengua como en Matemática, las categorías con mayor porcentaje son "Muy Bueno" y "Bueno".La distribución por fila muestra que, aunque hay ligeras diferencias, ambas asignaturas siguen un patrón de desempeño similar.

Figura 27 : Gráfico de barras adosadas por fila por desempeño y asignaturas de cuarto grado, departamento Uruguay Primaria.



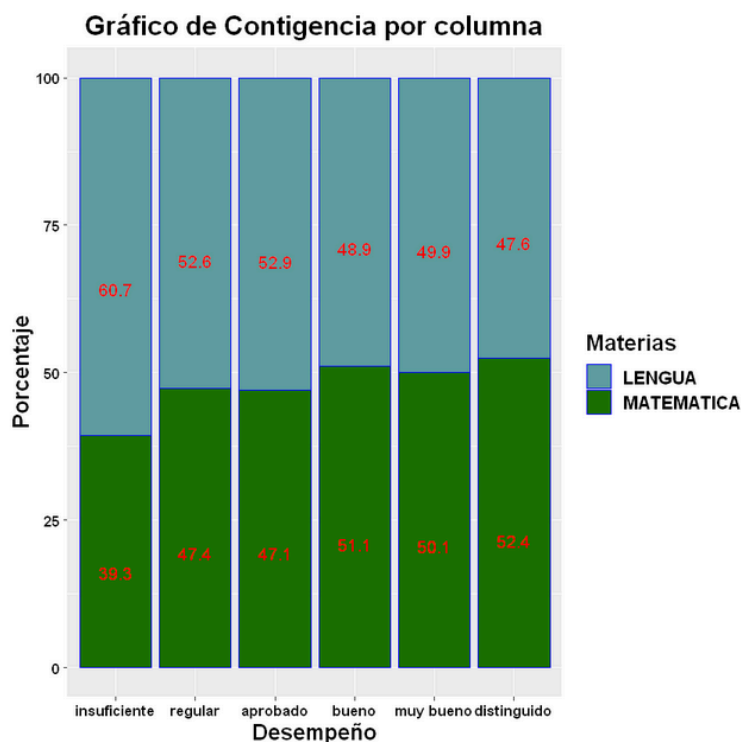
Fuente: Elaboración propia (2024).

Porcentaje Columna

	insuficiente	regular	aprobado	bueno	muy bueno	distinguido	Total
Lengua	60.7	52.6	52.9	48.9	49.9	47.6	50
Matemática	39.3	47.4	47.1	51.1	50.1	52.4	50
Total	100.0	100.0	100.0	100.0	100.0	100.0	100

La asignatura de Lengua tiene una mayor proporción de estudiantes en las categorías de "Insuficiente" (60.7%) y "Regular" (52.6%). Mientras que en Matemática la mayor proporción de estudiantes se encuentran en las categorías de "Bueno" (51.1%) y "Distinguido" (52.4%).

Figura 28: Gráfico de barras apiladas por columna por desempeño y sector en Lengua de cuarto grado, departamento Uruguay Primaria.



Fuente: Elaboración propia (2024).

Se realizó la prueba de Chi-cuadrado que dio como resultado un p-value de 0.5677. Esto significa que no se puede concluir que exista una relación significativa entre las dos variables.

Se calculó el coeficiente de Cramer V que dio como resultado 0.04601 lo que sugiere una asociación muy débil entre las variables. También se calcularon los valores esperados y residuos, los valores esperados para ambas asignaturas son iguales, reflejando un modelo de independencia, y los residuos son relativamente pequeños, lo que sugiere que las diferencias entre las frecuencias observadas y esperadas no son grandes. Esto apoya la conclusión de que las dos asignaturas tienen distribuciones de desempeño similares.

Análisis de los promedios de las asignaturas Lengua y Matemática en Secundaria

En el caso de la asignatura de Matemática del secundario, de primero a cuarto año la media muestral ronda en el promedio 6 y en quinto el promedio medio es un poco más alto con un valor de 6.98 y en sexto es igual a 7.13. La media muestral es un valor importante a considerar ya que en todos los años ocurre que no coincide con la moda y la mediana. Aunque son valores cercanos, esto quiere decir que en general los datos si se distribuyen de forma simétrica. Si sucede que en segundo, tercero y cuarto año los valores están centrados alrededor del valor de la media, lo que indica que la media es representativa de los datos de los promedios. En los otros tres años, ocurre que los datos están cercanos a la mediana. Al realizar tablas de frecuencia, pudimos obtener que el porcentaje de aprobados en la asignatura ronda desde el 50% al 65% en general, los porcentajes de estudiantes que sacaron notas menores a 4 va desde el 2,50% al 3,20% y los porcentajes de estudiantes que sacaron notas mayores a 8 son de alrededor del 20% en la mayoría de los años y solo en sexto supera el 30%. Luego de realizar un análisis de la mediana muestral se procedió a realizar un test de Wilcoxon para saber si la mediana poblacional era mayor a 7. El resultado del test reflejo que en primero, quinto y sexto esta era mayor a 7 pero en segundo, tercer y cuarto año no.

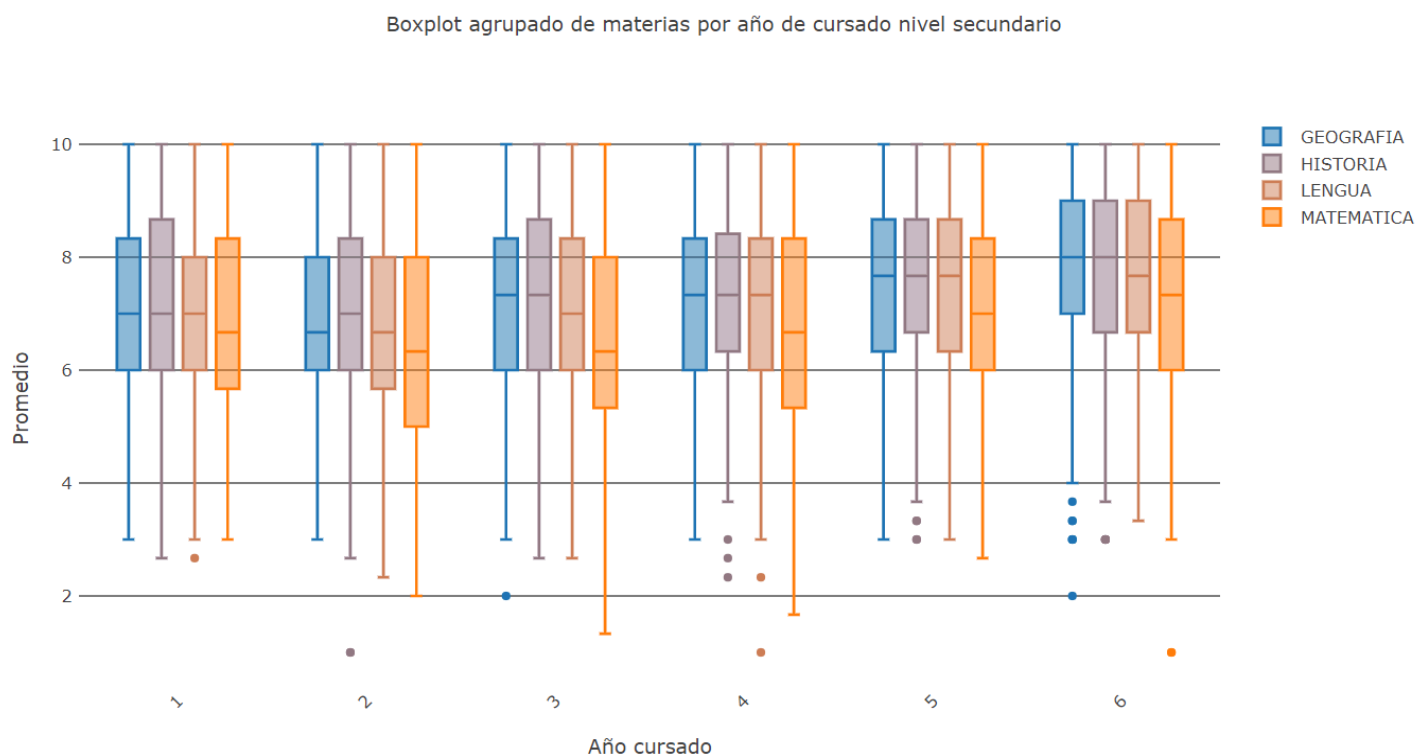
Si analizamos los datos de la asignatura Lengua, la media, mediana y moda tampoco tienen valores exactamente iguales pero sí son bastante cercanos, reflejando una distribución simétrica. Los datos están agrupados alrededor de la media en los seis años así que si se la considera representativa. El porcentaje de estudiantes aprobados es más alto, en todos los años es muy cercano al 80% y en primer año el 85% aprobó la materia. El porcentaje menor a 4 ronda en todos los casos en el 4% y el porcentaje de estudiantes con notas muy buenas, mayores a 8 también es mayor que en Matemática y en todos los años ocurre que el mismo está por encima del 30%. Particularmente quinto año tuvo un muy buen desempeño con un 40% de estudiantes con notas mayores a 8. En lengua también se procedió a realizar un test de Wilcoxon para saber si la mediana poblacional de los promedios

era mayor a 7. El resultado del test de Wilcoxon en Lengua mostró que en los tres primeros años la mediana si es mayor a 7 pero no ocurre esto en los 3 últimos años del secundario.

Al observar los diagramas de caja (boxplot) comparativos de las asignaturas de Lengua, Matemática, Geografía e Historia se observa que Matemática parece tener mayor variabilidad en la mayoría de los años. Las cajas que representan los promedios de Lengua, Geografía e Historia tienden a estar ligeramente por encima de las de Matemática, a excepción de primer año, lo que sugiere que los promedios de Matemática tienden a ser más bajos. Al observar la posición de las medianas (el valor en el centro de las cajas), vemos que en la mayoría de los años, esta está relativamente centrada en el centro de las cajas y que en todos los años la de Matemática es la que está más abajo.

Se identifican valores extremos en ambas asignaturas. Estos pueden observarse tanto cerca de los bigotes como a una distancia considerable de los mismos. Es importante señalar que muchos de estos valores atípicos se encuentran por debajo de 2.5, lo que indica la presencia de un número significativo de promedios muy bajos.

Figura 29: Gráfico de caja agrupados por año y asignaturas en el departamento Uruguay Secundaria.



Fuente: Elaboración propia (2024).

Para analizar si las distribuciones de los promedios de Lengua y Matemática son iguales, primero se comprobó que los datos no eran normales y se estudió si las varianzas eran iguales con el test de Levene. Esto dio que en los tres primeros años y en sexto las varianzas son iguales, y que en cuarto y quinto son diferentes. Como no se cumplía con el supuesto de normalidad de los datos se realizó el test de Wilcoxon-Mann-Whitney. El resultado de este, dio que solo en sexto año los promedios de los estudiantes en Matemática y Lengua son iguales.

Tabla de contingencia Desempeño de Lengua y Matemática Secundaria

Se crearon tablas de contingencia para la variable desempeño según la asignaturas asignaturas Lengua y Matemática para los alumnos de segundo año.

Porcentaje total

	Distinguido	Muy Bueno	Bueno	Aprobado	Regular	Insuficiente	Total
LENGUA	6.3	8.5	9.6	13.5	11.2	0.9	50
MATEMÁTICA	6.7	6.4	8.4	11.3	14.8	2.5	50
TOTAL	13.0	14.8	17.9	24.8	26.0	3.4	100

El porcentaje de la cantidad de alumnos que ocupa cada desempeño es 13.0% “Distinguido”, 14.8% “Muy Bueno”, 17.9% “Bueno”, 24.8% “Aprobado”, 26.0% “Regular” y 3.4% “Insuficiente”. El porcentaje de la población total que tiene como desempeño “Aprobado” y realiza Matemática es el 11,3%.

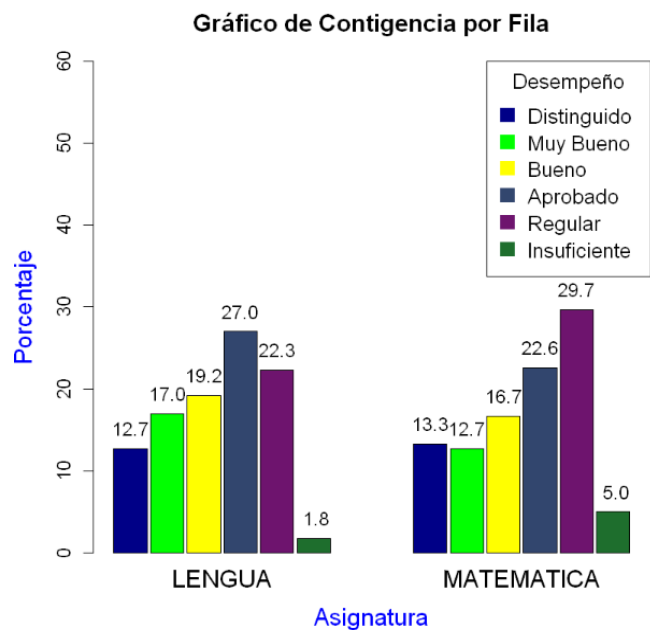
Porcentaje fila

	Distinguido	Muy Bueno	Bueno	Aprobado	Regular	Insuficiente	Total
LENGUA	12.7	17.0	19.2	27.0	22.3	1.8	100
MATEMÁTICA	13.3	12.7	16.7	22.6	29.7	5.0	100
TOTAL	13.0	14.8	17.9	24.8	26.0	3.4	100

En esta tabla de porcentaje fila se puede apreciar que el 12.7% de los que realizan Lengua son de desempeño “Distinguido”, el 17.0 de “Muy Bueno”, el 19.2 de “Bueno”, el 27.0 de “Aprobado”, el 22.3 de “Regular” y el 1.8 de “Insuficiente”. De los que realizan Matemática, el 12.7% son de desempeño “Distinguido”, el 17.0 de “Muy Bueno”, el 19.2 de “Bueno”, el 27.0 de “Aprobado”, el 22.3 de “Regular” y el 5.6 de “Insuficiente”.

En el gráfico de contingencia por filas se puede ver que en Lengua el 29.7% lo ocupa el desempeño “Regular” y es el turno que más porcentaje tiene del total. En Matemática el desempeño con mayor porcentaje es el “Aprobado” con un 27% y el de menor porcentaje el “Insuficiente” con un 1.8%.

Figura 30: Gráfico de barras adosadas por columna por desempeño y asignatura de Lengua y Matemática de segundo año, departamento Uruguay Secundaria.



Fuente: Elaboración propia (2024).

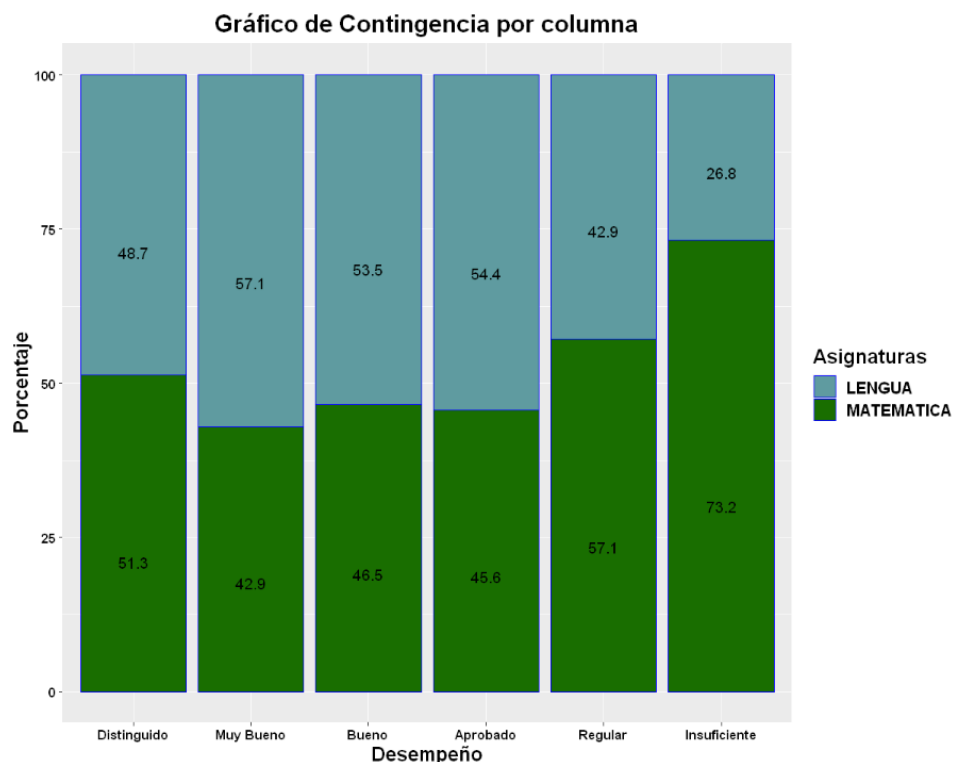
Porcentaje columna

	Distinguido	Muy Bueno	Bueno	Aprobado	Regular	Insuficiente	Total
LENGUA	48.7	57.1	53.5	54.4	42.9	26.8	50
MATEMÁTICA	51.3	42.9	46.5	45.6	57.1	73.2	50
TOTAL	100	1100	100	100	100	100	100

Como se observa en la tabla, el 48.7% de los de desempeño “Distinguido” realizan Lengua y el 51.3 realizan Matemática . El 26.8% de los de desempeño “Insuficiente” realizan Lengua y el 73.2% realizan Matemática. El desempeño que tiene mayor proporción en Lengua es el “Muy Bueno” y en Matemática el “Aprobado”.

En el siguiente gráfico de contingencia por columna se puede apreciar que el desempeño “Insuficiente” es el que tiene mayor proporción en Matemática y que el desempeño Muy Bueno es el que tiene menor porcentaje en Matemática. .

Figura 31: Gráfico de barras apiladas por columna por desempeño y asignatura de Lengua y Matemática de segundo año, departamento Uruguay Secundaria.



Fuente: Elaboración propia (2024).

Se realizó la prueba de Chi-cuadrado de independencia, el p-valor dio $1.922e-08$ y también se calcularon las frecuencias observadas y esperadas y dieron que las mismas tenían diferencias notables. Por lo tanto, puedes concluir que hay una asociación significativa entre el desempeño y las asignaturas de Matemática y Lengua. Esto significa que la variable desempeño no es independiente de las asignaturas de Matemática y Lengua, lo que sugiere que el rendimiento puede verse influido dependiendo la materia que se está considerando.

Se calculó el coeficiente de Cramer que dio un valor de 0.1354, esto quiere decir que la asociación entre el desempeño y el tipo de escuela se encuentra en el rango débil. Significa que, aunque hay una relación entre las asignaturas Lengua y Matemática y el desempeño, esta relación no es muy fuerte. Por último, también se realizó un análisis de diferencias de las proporciones, el valor p fue $3.753e-07$ esto indica que hay diferencias significativas entre las proporciones de los desempeños respecto a las asignaturas.

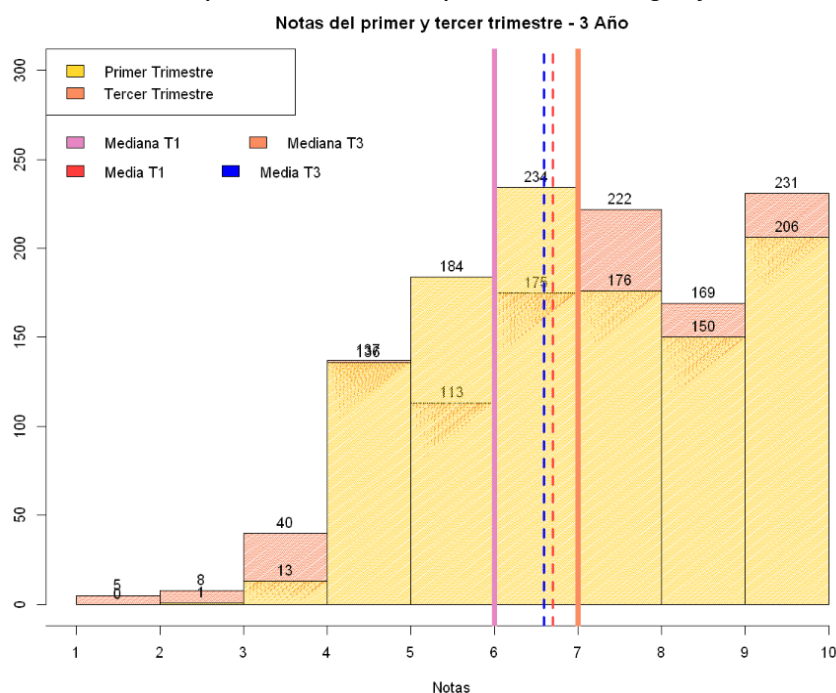
Análisis de las notas del Primer Trimestre y el Tercer Trimestre

Se quiso realizar una comparación entre las notas que obtienen los estudiantes en el Primer Trimestre y el Tercer Trimestre. Para eso se planteó la pregunta: ¿Hay una diferencia significativa entre las notas del tercer trimestre y las del primer trimestre? Para ello, se verificó si las diferencias entre las notas del tercer y el primer trimestre eran normales. El resultado de este análisis fue que tanto para el nivel primario como secundario esto no se cumplía. Por lo que se realizó el test Wilcoxon signed-rank.

En primaria, en general, el resultado del test da un p-value de $2.2e-16$, lo que nos dice que hay una diferencia significativa entre las medianas de las notas del primer y tercer trimestre.

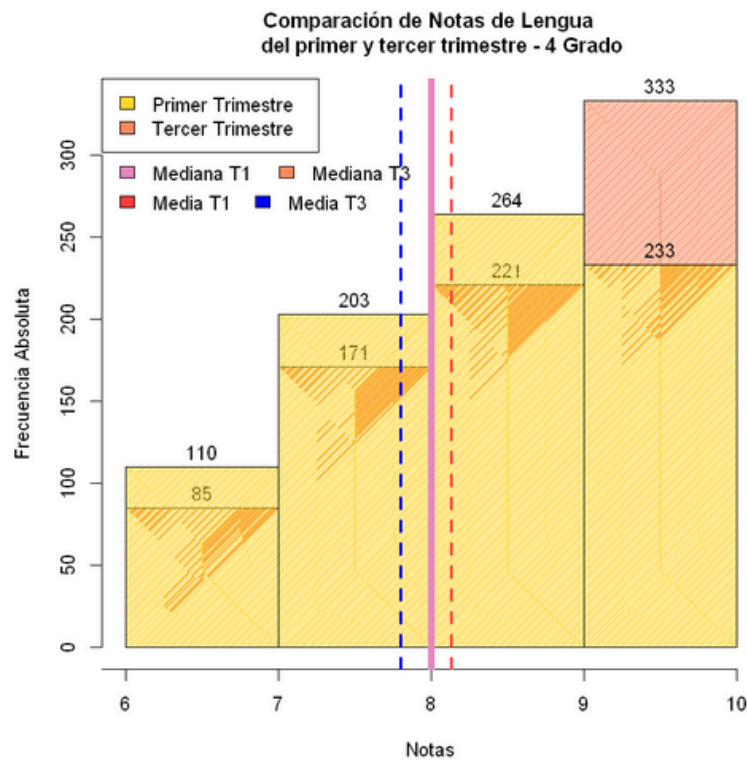
En el caso del nivel secundario, el resultado del test refleja que en primero p-value es 0.873 y en sexto año se obtuvo un p-value 0.1495 este test quiere decir que en estos años no hay una diferencia significativa entre las medianas de las notas del primer y tercer trimestre, pero en el resto de los años si hay diferencia ya que en general el p-value da un valor de $2.2e-16$. En el siguiente gráfico que representa las notas de matemática de los alumnos de tercer año, se puede notar que a los alumnos les fue mejor en el Tercer Trimestre, ya que sobresalen las barras del Tercer Trimestre en las notas de 7 hasta 10. Además, en este año ocurre que la mediana del tercer trimestre es 7 y la del Primer Trimestre 6.

Figura 32: Gráfico histograma del Análisis del Desempeño en Matemática del Primer y Tercer Trimestre para 3° Grado, Departamento Uruguay Secundaria



Fuente: Elaboración propia (2024).

Figura 33: Gráfico histograma del Análisis del Desempeño en Matemática del Primer y Tercer Trimestre para 4° Grado, Departamento Uruguay Primaria.



Fuente: Elaboración propia (2024).

Se decidió realizar el test de MANOVA unidireccional para comprobar si las Notas del Tercer Trimestre afectan tanto a los promedios como a la cantidad de materias que rinden los estudiantes de primer año que realizan la asignatura Matemática. Sin embargo, el mismo no se pudo realizar porque los datos no eran normales.

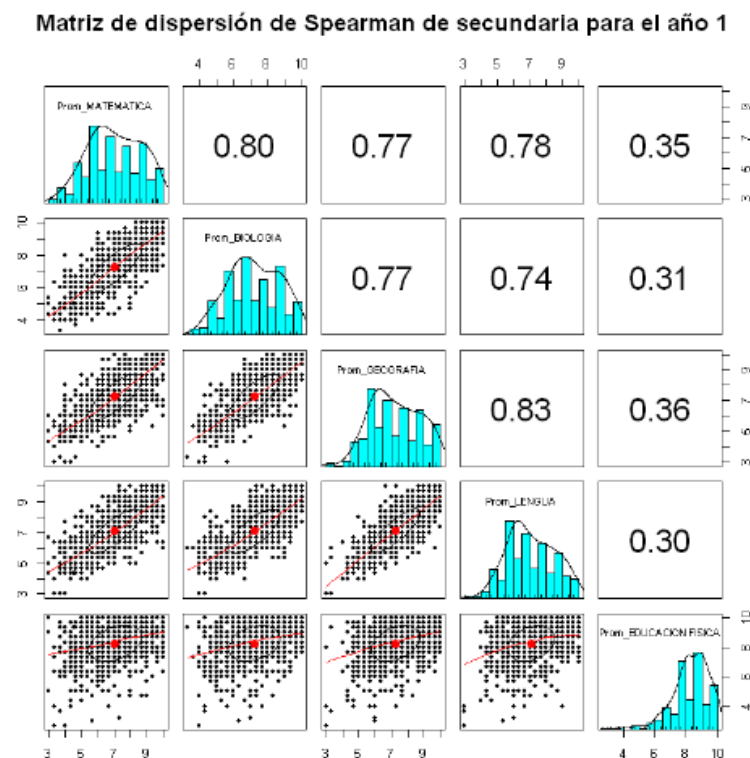
Correlación entre asignaturas

Se realizaron búsquedas de correlación en las diferentes materias que cursaron los estudiantes. Esto se llevó a cabo para averiguar si su rendimiento en algunas materias tenían relación con el rendimiento que obtuvieron en otras materias que cursaron.

Correlaciones en el nivel primario

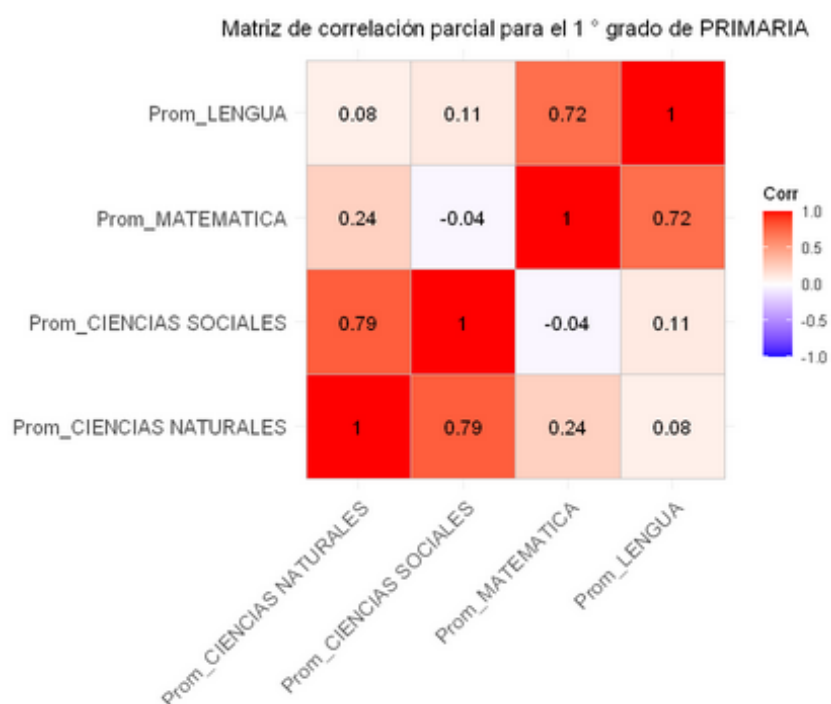
Se procedió a analizar las asignaturas ciencias sociales y ciencias naturales, entre estas dieron una correlación muy fuerte de primer grado hasta tercer grado, en cuarto tiene una correlación moderada y ya en los años siguientes carecen de correlación. Pero Matematica no tiene ninguna correlación con ciencias sociales o ciencias naturales, no obstante, si analizamos Lengua, podemos ver que de cuarto grado en adelante existe una correlación moderada con los promedios de ciencias sociales mientras que con ciencias naturales no tiene correlación en ningún grado.

Figura 34: Matriz de correlaciones de los promedios entre distintas materias, departamento Uruguay nivel primario.



Fuente: Elaboración propia (2024).

Figura 35: Matriz de correlaciones parciales de los promedios entre distintas materias, departamento Uruguay nivel primario.

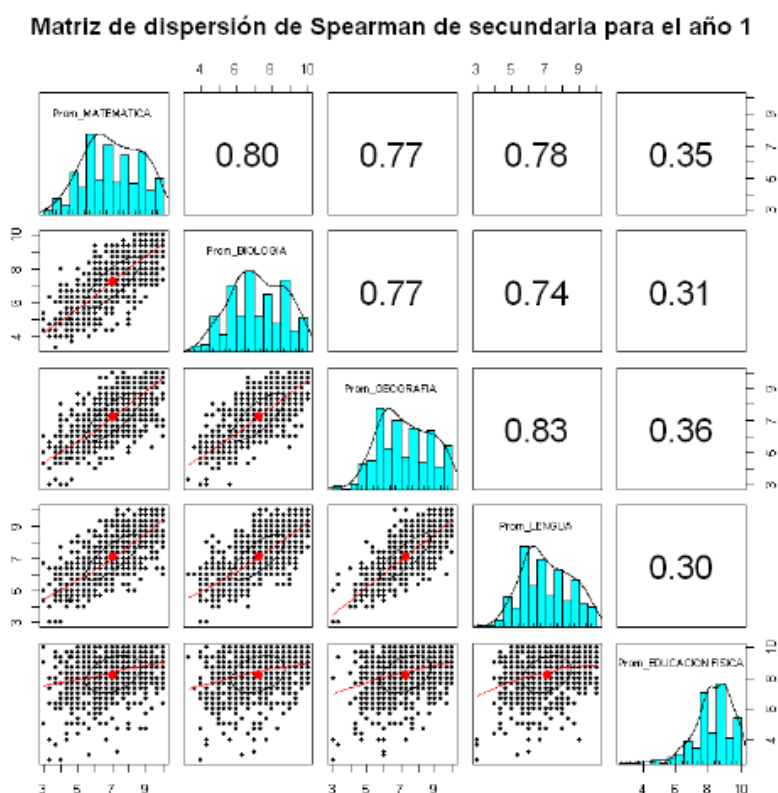


Fuente: Elaboración propia (2024)

Correlaciones en el nivel secundario

Se procedió a estudiar la correlación Matemática con Educación Física y se obtuvo que no había correlación. También se estudió la correlación entre las asignaturas: biología, Lengua y geografía. Al realizar la matriz de correlación el resultado obtenido era que en todos los años estas asignaturas no tenían una correlación mayor a 0.7 para ser considerada fuerte a excepción de primer año donde la correlación entre geografía y Matemática daba 0.77 y de Matemática y biología 0.80. También se notó otras correlaciones fuertes entre otras materias como biología y Lengua en los primeros años de secundaria.

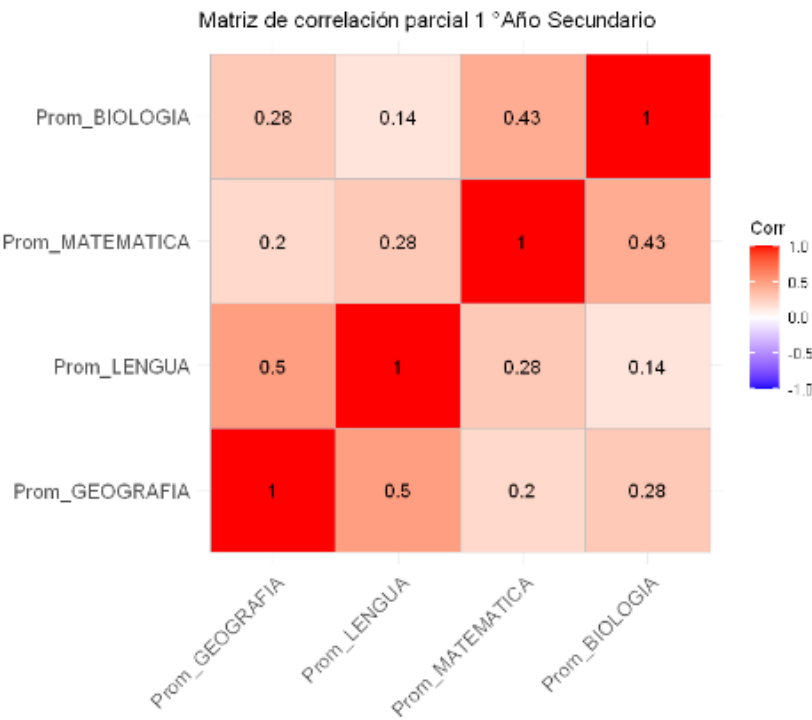
Figura 36: Matriz de correlaciones de los promedios entre distintas materias, departamento Uruguay nivel primario.



Fuente: Elaboración propia (2024)

Para realizar un análisis de estas correlaciones fuertes que se obtuvieron, se eligió realizar matrices de correlación parcial con el objetivo de eliminar el efecto de otras asignaturas que podrían estar influyendo en el resultado de las correlaciones que se obtuvieron. El resultado luego de aplicar la matriz de correlación parcial, fue que las correlaciones en todos los casos daban un valor de 0.5 en el mejor de los casos pero valores menores a 0.3 en la mayoría de los casos.

Figura 37: Matriz de correlaciones parciales de los promedios entre distintas materias del departamento de Uruguay secundaria.



Fuente: Elaboración propia (2024)

Se suiso realizar un test de ANCOVA para corroborar si existe una relación significativa entre los promedios de Matemática, Lengua e Historia controlando el efecto del promedio de Geografía. Aun así, los datos no cumplían los supuestos de normalidad, por lo que no se pudo realizar el mismo tanto en el nivel primario como secundario.

Conclusiones

Finalmente, se llegó a la conclusión de que el dataset presenta varias limitaciones. Entre ellas están: la presencia de valores faltantes, falta de carga de notas, materias no estandarizadas, numerosos registros duplicados y que los datos no seguían una distribución normal. De todas ellas, la que más afectó el tamaño del dataset fue la gran cantidad de datos duplicados. Aun así, una vez realizada la limpieza, se consideró que se cuenta con información suficiente para realizar un análisis inicial del sistema educativo en el departamento Uruguay. Ambos dataset que resultaron luego de la división del dataset de Uruguay, primaria y secundaria, se consideran de calidad para poder calcular el análisis que permitirá determinar qué asignaturas se les dificulta más a los estudiantes. Con respecto a la calidad de los datos en las notas, podemos decir que en todos los años la variable promedio no tiene una distribución normal para ninguno de los años de ambos niveles.

En el caso de primaria, los estudiantes tuvieron un mejor rendimiento en Matemática. En esta materia es menor el porcentaje de estudiantes en proceso, además de que tienen promedios muy buenos. En la materia de Lengua en cuarto grado se observa que les costó un poco más que en los grados anteriores, y en los grados siguientes sus promedios en Matemática y Lengua se mantienen en el mismo nivel. Algo a tener en cuenta es que estas materias se encuentran correlacionadas, por lo que si a un alumno le va muy mal en Lengua es muy probable que no le vaya bien en Matemática y viceversa.

Se encontró que al comparar el desempeño de los alumnos con la gestión de la escuela, esta puede influir significativamente en el aprendizaje. Los estudiantes que asisten a escuelas privadas son mayormente categorizados con un buen desempeño ("Bueno" o superior), a diferencia de los que asisten a escuelas públicas.

En el análisis del rendimiento académico en el nivel secundario, se puede decir que se observa un mejor desempeño en la asignatura de Lengua en comparación con Matemática. Esto se refleja en una mayor proporción de estudiantes con calificaciones sobresalientes (superiores a 8), aunque también se registra un mayor porcentaje de calificaciones por debajo de 4 en Lengua. Sin embargo, en todas las instancias, el número de estudiantes aprobados en Lengua supera al de Matemática.

Se amplió el análisis considerando otros factores que podrían influir en el promedio de los estudiantes. Una observación notable es que los alumnos de escuelas privadas tienden a tener promedios más altos en general. Además, al examinar específicamente la asignatura de Matemática en segundo año, se detecta una mayor proporción de alumnos

con desempeño "Distinguido" en instituciones privadas, mientras que en las públicas hay más estudiantes con desempeño "Insuficiente".

Si hablamos de los trimestres, en el caso de la asignatura Matemática, podemos decir que en general los alumnos tienen sus medianas parecidas en las notas, con lo que no hay diferencias significativas a excepción de tercer año, donde la mediana del tercer trimestre es más alta. Esto puede deberse a un esfuerzo adicional de los estudiantes en el último trimestre para mejorar sus calificaciones, especialmente si han tenido dificultades en los trimestres anteriores.

Al hablar de los turnos, como ya se mencionó anteriormente, el turno más concurrido es el de la mañana seguido del de tarde tanto en primaria como en secundaria. En segundo año de secundaria, en la asignatura de Matemática, los alumnos del turno mañana muestran un alto porcentaje de desempeño "Distinguido" y un bajo porcentaje de "Insuficiente", mientras que en el turno vespertino ocurre lo contrario.

En términos generales, el desempeño de los alumnos en primer año de secundaria es positivo tanto en Matemática como en Lengua, indicando un buen nivel previo adquirido en la primaria. Además, se observa un rendimiento satisfactorio en los últimos años en ambas asignaturas, lo cual es un indicador positivo, especialmente en Matemática, dada su importancia en los requisitos de ingreso universitario.

Recomendaciones

Se recomienda realizar un proceso de limpieza y validación exhaustivo en futuras recolecciones de datos, así como establecer protocolos claros para garantizar la calidad y la integridad de la información recopilada.

Recomendamos que para el caso de primaria, se manden acompañamientos a la asignatura de Lengua, con foco especial en las escuelas públicas y de cuarto grado en adelante. En el caso de secundaria, es necesario monitorear y mandar acompañamiento en segundo y tercer año que es donde son más los alumnos que no están aprobados. Además también se debe tener en cuenta a cuarto para matemática, ya que su rendimiento no fue tan bueno tampoco. Recomendamos que además de tener este enfoque en los años también se considere más enviarlos a escuelas que son públicas. Esto es así porque en todos los casos, según los resultados obtenidos, las escuelas privadas han demostrado tener mejores resultados. Esto puede deberse a que en las escuelas privadas se explican

mejor los contenidos o que los alumnos pueden acceder más fácil a clases particulares si no entendieron algún tema.