



**Milessi Ayrton y Torres Santiago**

**Universidad Nacional de Entre Ríos**

**Facultad de Ingeniería**

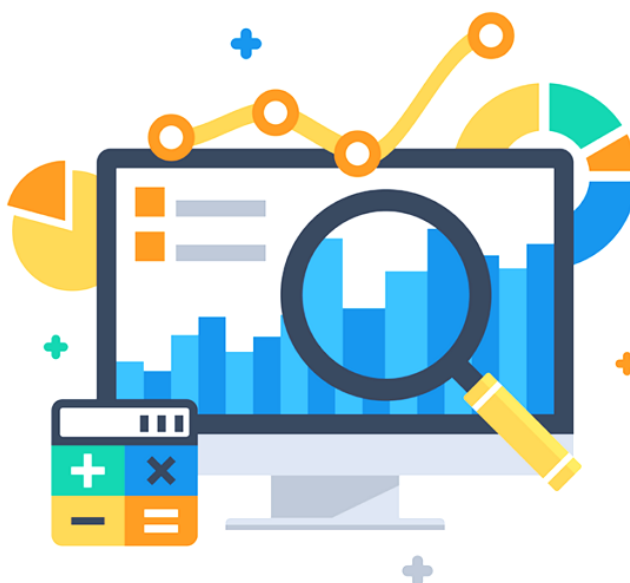
**Tecnicatura Universitaria en Procesamiento y Explotación de Datos**

**“Exploración de Datos Multivariados”**

**Prof. Melisa Fernández**

**“Análisis estadístico del nivel educativo del departamento Victoria en el año 2023”**

**Oro Verde - 05/06/2024**



## **Resumen**

A lo largo de todo el informe se explorará, investigará y analizará la educación en los niveles primaria y secundaria del departamento Victoria de la provincia de Entre Ríos. Los datos obtenidos serán del total del año 2023, donde contendrán algunos datos importantes como las notas de los estudiantes en diferentes cursos y distintas materias. También, a partir de los datos originales, se crearán nuevas variables para expandir el análisis.

El principal objetivo de este análisis es tener conocimientos de la situación educativa en el departamento, obteniendo resultados valiosos sobre el desempeño académico de los estudiantes, averiguando las áreas a mejorar y los puntos de éxito logrados.

Se utilizarán medidas estadísticas descriptivas y análisis de distribución de datos para examinar el desempeño en cada curso de cada nivel educativo. Los resultados serán visualizados con gráficos de barras, boxplot, histogramas y de sectores. Además, se analizarán las correlaciones entre los promedios de distintas materias, incluido el análisis de correlación parcial para controlar posibles efectos de materias que influyen entre las notas de las materias.

También se realizarán comparaciones entre los diferentes grupos de datos, para saber si estos test brindan resultados contundentes, se le realizarán sus correspondientes tests de potencia. Por último, a través de tablas de contingencia podremos hacer un análisis bivariado, logrando resultados muy abarcativos.

El informe busca proporcionar información valiosa que contribuya a la toma de decisiones en materia educativa, identificando áreas de oportunidad y fortalezas que puedan ser potenciadas para mejorar el desempeño académico de los estudiantes.

## **Introducción**

En el marco de la materia “Exploración de Datos Multivariados”, se desarrolló un trabajo de investigación enfocado en estudiar el nivel de educación en los niveles primaria y secundaria del departamento Victoria, ubicado en la provincia de Entre Ríos.

Nelson Mandela alguna vez afirmó que "la educación es el arma más poderosa que puedes usar para cambiar el mundo". Esta cita resalta la importancia de la educación tanto para las personas como para la transformación de las sociedades.

Por eso es muy importante examinar el desempeño académico en el departamento Victoria, con el objetivo de identificar áreas de éxito y oportunidades de mejora en el sistema educativo local. Como también, destacar a los estudiantes con resultados sobresalientes. Además, este análisis proporcionará información valiosa para la formulación de políticas y estrategias destinadas a fortalecer y optimizar el sistema educativo local.

A lo largo de todo el informe se mostrarán las etapas del análisis de datos, desde su limpieza hasta los resultados más destacados producto del riguroso análisis.

## **Metodología**

Para poder llevar a cabo el análisis, se usó el lenguaje de programación R dentro del entorno Jupyter, ya que es uno de los más robustos para el análisis de datos, a través de sus funciones, librerías y fácil lectura permite una variedad de análisis muy importante.

Antes de pasar a la etapa de procesamiento de los datos, nos dimos cuenta mientras se exploraban los datos que se necesitaban varias etapas de limpieza. Estas fueron:

- 1. Renombramiento de variables:** Se ajustaron los nombres de las variables para evitar confusiones futuras y asegurar la consistencia en la nomenclatura.
- 2. Eliminación de variables repetidas:** Al querer renombrar variables, nos encontrábamos con una variables que contiene los mismos datos que otra, se procedió a eliminar una de ellas.
- 3. Segmentación de la base de datos:** Para no mezclar niveles de educación, los datos se separaron en dos bases: una perteneciente a los estudiantes de primaria que pertenecían a escuelas comunes y ninas, evitando los grados multiaños. Por otro lado, la base de datos de secundaria contenía a los estudiantes que realizaban sus estudios en escuelas comunes, técnicas y agrotécnicas, no se tomaron estudiantes que cursaron multiaños.
- 4. Modificación de datos:** Luego de consultar con una docente que ejerció la profesión por 25 años, se modificaron las notas “EPAC”, “PA”, “UP” según la nota numérica correspondiente consultada ante la profesional en el área. También se cambiaron los registros de la variable ‘Gestión’ para mejorar su comprensión durante el análisis.
- 5. Creación de nuevas variables:** A partir de la variable 'periodoEvaluatorio' se tomó la decisión de crear las variables 'NotasPrimerT', 'NotasSegundoT', 'NotasTercerT' para tener ordenado las notas en cada trimestre para cada asignatura que cursó cada alumno.  
Además, generamos las variables 'Promedio' y 'Condición' donde se guarda el promedio logrado por cada estudiante en sus asignaturas y la condición final que consiguieron al finalizar el año lectivo.  
Se añadió la variable ‘Tendencia’ para indicar si las notas de los estudiantes a lo largo del año en las distintas materias eran 'Creciente', 'Estable' o 'Decreciente'.
- 6. Estimación de edades:** Después de una búsqueda exhaustiva para poder lograr este

complejo desafío, logramos calcular un rango de edad para cada estudiante según su DNI. Como este método puede tener errores, se utilizó un modelo de regresión lineal para predecir las edades usando datos de documentos y edades de alumnos de primaria, con autorización del director escolar.

7. **Renombramiento de asignaturas:** La mayoría de las materias ingresadas tenían los mismos contenidos pero tenían diferentes nombres asignados. Por lo cual, se procedió a reemplazar estos datos con las funciones 'gsub' y 'str\_replace\_all'.
8. **Creación de nuevas variables:** A través de las variables 'asignatura' y 'Promedio', generamos variables como el número total de materias por estudiante, el número de materias aprobadas y desaprobadas. Con estas, calculamos el porcentaje de materias aprobadas y desaprobadas por cada estudiante y su nivel de eximición.  
Por otro lado, en base al promedio generamos la variable 'Desempeño' donde se categoriza el promedio logrado por cada estudiante en las materias.
9. **Promedios completos:** Nos quedamos con aquellos estudiantes que tienen las notas en los tres trimestres, por lo cual los que tienen el promedio y no un registro incompleto.
10. **Modificación de los identificadores de los cursos:** para comprender mejor el curso al cual pertenece cada estudiante, se les cambió el formato que tenían a uno más simple. Un ejemplo es el del estudiante que cursa en "1 - A Prim" ahora su curso se lo identificará como "PRIMERO".
11. **Verificación de las etapas de limpieza:** Revisamos las variables modificadas para asegurar que los pasos de limpieza se realizaron correctamente, identificando y corrigiendo errores remanentes.
12. **Selección de variables a analizar:** para mejorar los costos computacionales a la hora de analizar y procesar datos, se seleccionaron las variables que sirven para el análisis. Estas fueron:  
'CodigoUnicoEscolar', 'departamento', 'Gestión', 'turno', 'Año\_curso', 'Documento', 'asignatura', 'idalumno', 'Modalidad', 'ModEnseñanza', 'NotasPrimerT', 'NotasSegundoT', 'NotasTercerT', 'Promedio', 'Condición', 'Desempeño', 'Tendencia', 'Eximido', 'Rango\_edad', "total\_asig", "porc\_aprob", "porc\_desaprob".

**13. Segmentación de las bases de datos por curso:** se separaron en varios datasets los datos, esto se hizo para poder analizar los datos por curso. Al final nos quedan 6 bases en primaria y 7 en secundaria.

**14. Creación de nuevas variables:** Con los promedios obtenidos para cada estudiante en diferentes materias, calculamos el promedio general. Para los cursos aptos para las banderas de ceremonia, se posicionó a los estudiantes según su promedio general. Además, se identificó la mejor y peor materia de cada estudiante según el promedio obtenido.

Por otro lado, calculamos el porcentaje de datos mal cargados para mostrar la robustez de la etapa de limpieza:

	Base sin etapa de limpieza	Base con etapa de limpieza	Porcentaje de datos eliminados
Primaria	119293-26	30890-17	83.0%
Secundaria	128743-26	51686-17	73.7%
Fuente de elaboración propia (Tabla N°1)			

Los métodos utilizados para el análisis de datos fueron:

- 1. Medidas estadísticas descriptivas:** A través de una tabla, se muestran resumidamente las medidas de centralidad, varianza y dispersión.
- 2. Representación de medidas estadísticas:** Uno de los gráficos que más representan las medidas estadísticas, son los boxplot. Por eso, además de mostrar en números las medidas, también se representan visualmente para una mejor comprensión.
- 3. Medidas de distribución:** Para saber si el comportamiento de los datos seguía una curva normal o no, se le calcularon a las variables correspondientes la curtosis, coeficiente de asimetría y pruebas de normalidad. En algunos casos donde sea necesario, se representarán gráficamente estos resultados con histogramas.
- 4. Puntuación Z:** Para comparar grupos de datos independientes con una característica común, se calcularon las puntuaciones Z. Este método permitió identificar qué conjunto de datos estaba más alejado de la media, facilitando la comparación entre los grupos.

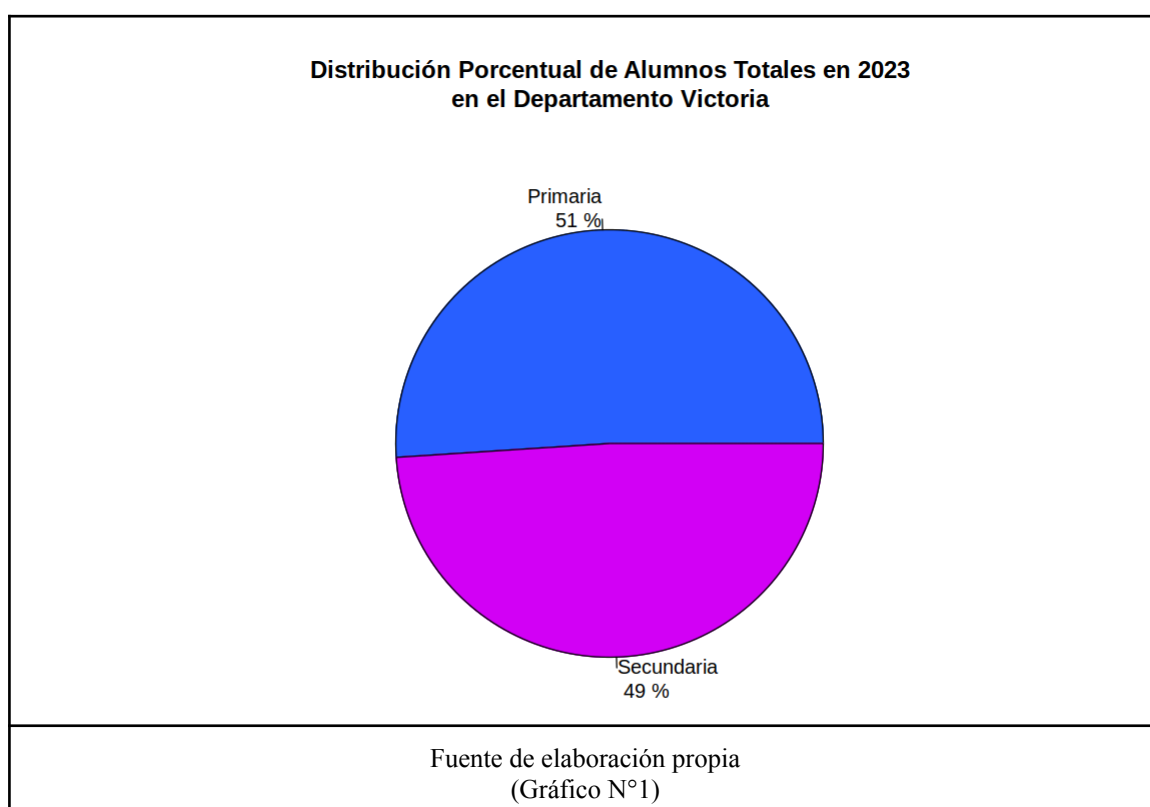
5. **Correlación:** Uno de los temas más importantes en la estadística es la correlación entre dos variables, por la importancia de este parámetro, queríamos calcular para cada grado que variable se correlaciona con otra. Para saber si esas variables están directamente correlacionadas y no influya otra variable se aplicará las correlaciones parciales.
6. **Test de una media:** Posibilitó la comparación de datos con un parámetro establecido, ayudando a identificar si un grupo de datos era mejor o peor.
7. **Test de medias:** Para poder comparar los rendimientos entre diferentes grupos, ya sean dos o más, dependientes o pareados, podemos hacerlos a través de los diferentes tests de medias, donde hay una gran variedad, lo que permitió hacer un análisis complejo y enriquecedor.
8. **Análisis Bi-variado:** Como se comentó anteriormente en el informe, a través de diferentes tablas de contingencia, podemos calcular porcentaje y facilitar la comparación entre las variables de un mismo grupo o diferentes.
9. **Análisis multivariado:** Este método permitió analizar múltiples variables simultáneamente, proporcionando resultados más completos y complejos al considerar varias variables a la vez.

## **Resultados**

A lo largo de las siguientes páginas, se expondrán los resultados logrados de diferentes análisis tanto en el nivel primario como en el nivel secundario.

Antes de comenzar el análisis por nivel de enseñanza, puede resultar interesante entender cómo es la distribución de estudiantes por nivel en el departamento Victoria.

En el gráfico N°1 podemos observar la distribución de la totalidad de alumnos en nuestro departamento. Vemos que un 49% de las matrículas se corresponden con alumnos del nivel secundario y el otro 51% se corresponde con alumnos del nivel primario.



### **Primaria**

Primero nos dedicaremos a los resultados obtenidos en el nivel de enseñanza primaria.

Antes de adentrarnos en lo que sucede en cada grado del departamento, haremos un análisis en general de la primaria.

A continuación se mostrará una tabla donde se encuentran la cantidad de escuelas primarias diferentes en Victoria.

Tipo de escuela	Cantidad
Escuelas primarias en Victoria	16
Escuelas primarias públicas en Victoria	13
Escuelas primarias privadas en Victoria	3

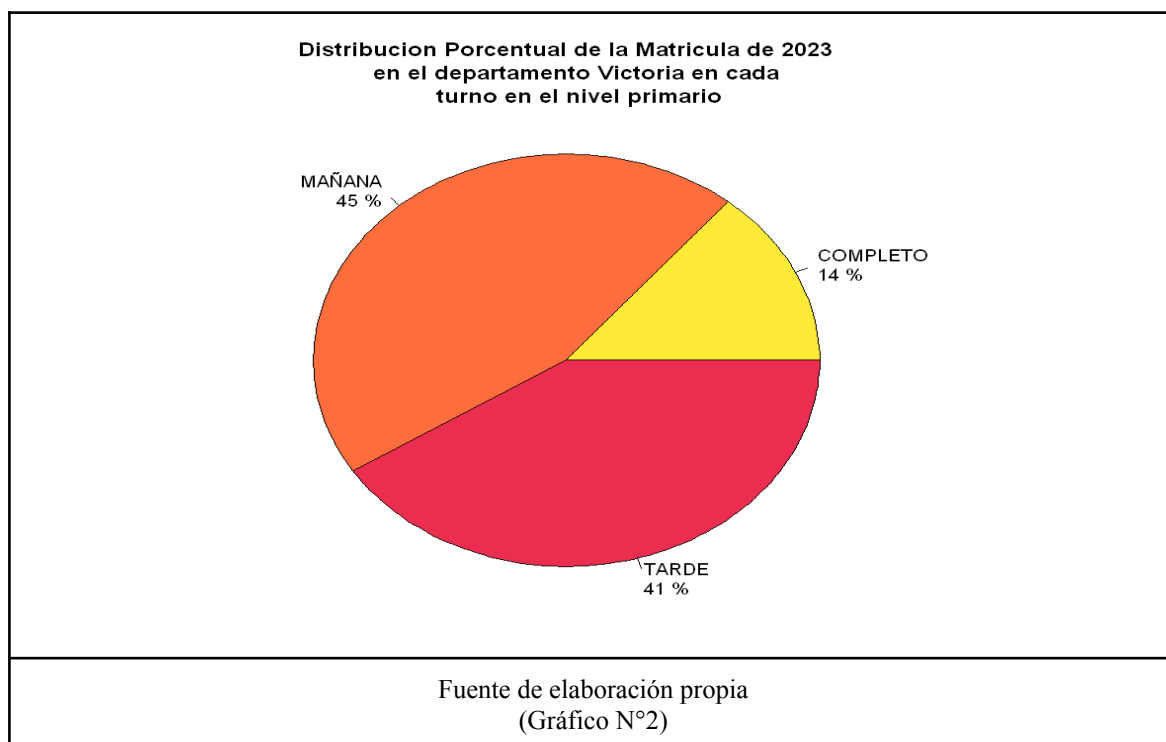


Escuelas primarias que tienen turno mañana en Victoria	12
Escuelas primarias que tienen turno tarde en Victoria	11
Escuelas primarias que tienen turno completo en Victoria	2
Escuelas primarias comunes en Victoria	16
Escuelas primarias NINA en Victoria	1
Fuente de elaboración propia (Tabla N°2)	

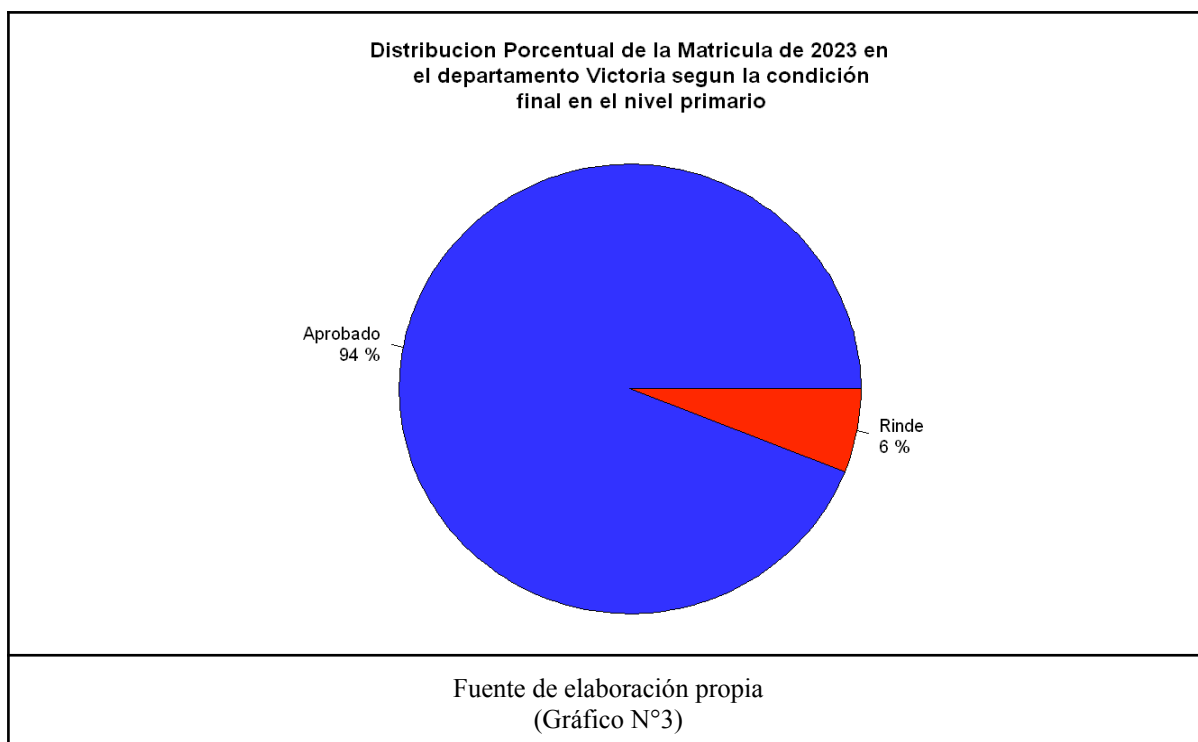
También es importante preguntarse ¿En qué turno cursan la mayoría de los estudiantes? ¿Asisten más a escuelas públicas que privadas? ¿La cantidad de aprobados finales es alto, bajo, está bien?

A continuación, abordaremos cada una de estas preguntas para proporcionar una visión integral del panorama educativo en el nivel primario.

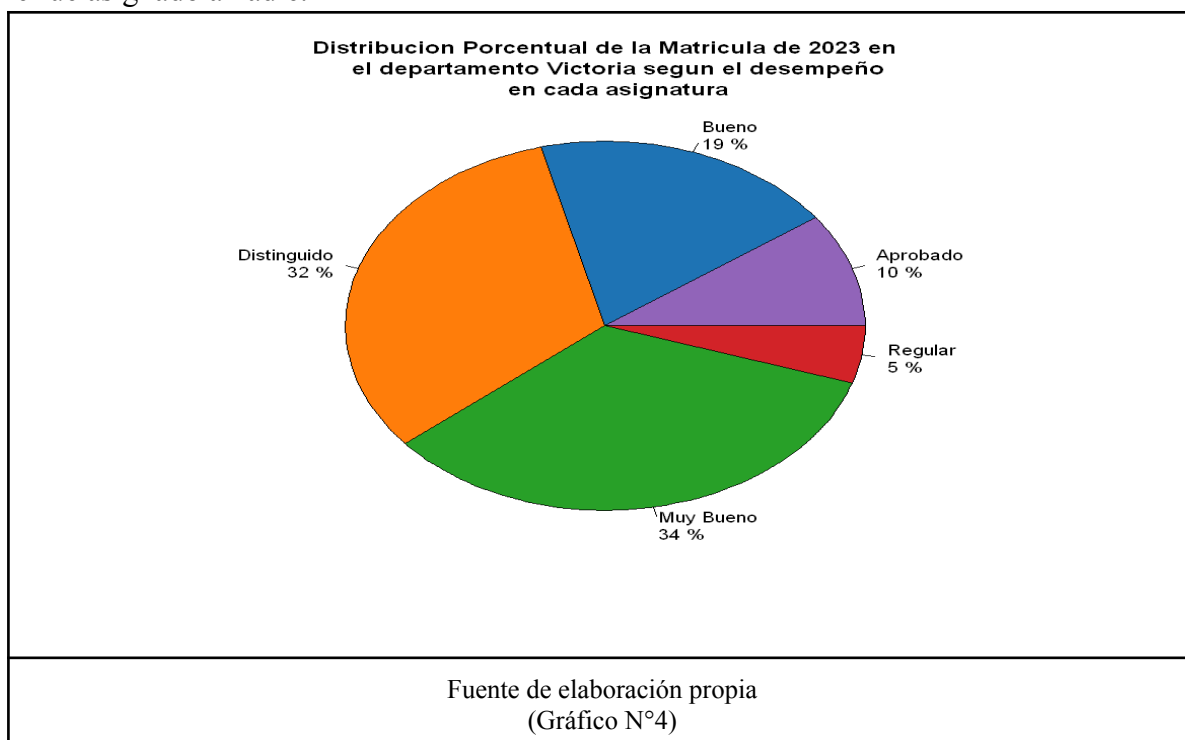
El mayor porcentaje de matriculados cursan la primaria en el turno mañana.



En cuanto a la condición final de cada estudiante, los resultados fueron positivos ya que el 94% de las materias fueron aprobadas, pero no hay que estar satisfechos, ya que un 6% tiene que rendir en las instancias de recuperación para poder aprobar la asignatura correspondiente.



Solo el 5% de las materias cursadas por los estudiantes de Victoria obtuvieron un desempeño regular. Mientras que un 32% lograron el mejor desempeño posible y el peor desempeño no le fue asignado a nadie.



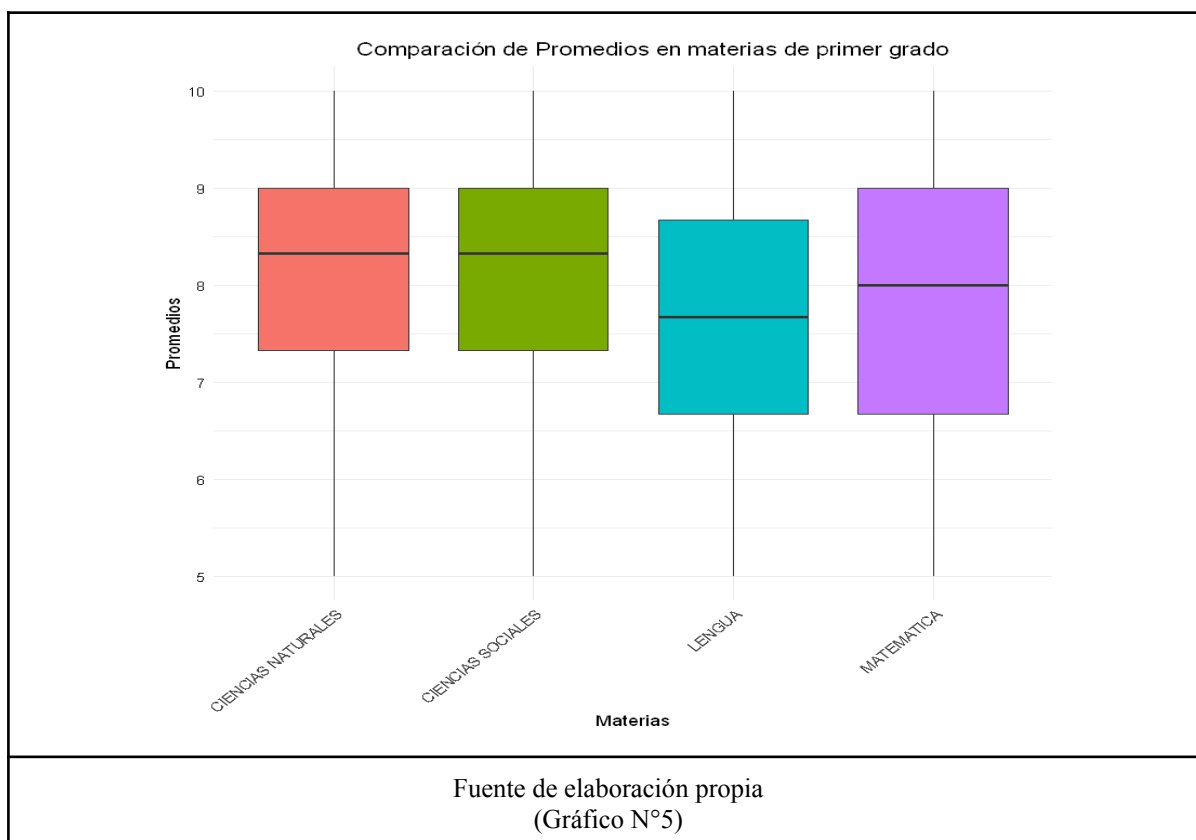
Antes de entrar en el análisis grado por grado, queríamos hacer una prueba bivariada entre los desempeños y los cursos. Aunque no se buscaba encontrar una asociación entre las variables, solo queríamos adelantar lo que podríamos encontrar en los grados.

	Regular	Aprobado	Bueno	Muy Bueno	Distinguido	Total
Primero	199	438	847	1547	1486	4517
Segundo	197	393	753	1666	1519	4528
Tercero	215	446	913	1731	1518	4823
Cuarto	205	411	801	1331	1117	3865
Quinto	311	441	755	1421	1363	4291
Sexto	255	520	839	1441	1499	4554
Total	1382	2649	4908	9137	8502	26578
Fuente de elaboración propia (Tabla N°3)						

### Primer grado

Constamos de 568 matriculados a lo largo de todo el departamento.

Los promedios obtenidos por los estudiantes en las asignaturas troncales como Cs. Naturales, Cs. Sociales, Matemática y Lengua fueron:



- ❖ Lengua tiene los peores promedios de las cuatro materias, ya que su mediana (7.67) y Q3 (8.67) son las más bajas.

- ❖ Por el contrario, las mejores materias son las Ciencias, tanto Naturales como Sociales. Tienen medianas y distribución muy similares. El 50% está ubicado en un promedio de 8.33.
- ❖ Además, Lengua es la materia con mayor variabilidad ( $\text{CoefVar} = 18.31$ ) respecto a la media de las comparadas.
- ❖ Ninguna materia presenta outliers. Todas tienen el mismo rango de datos, promedios que están dentro de  $[5 - 10]$ .

Para saber qué tan alejado está un estudiante con un promedio de 8 en Matemáticas según el turno, se calculó la puntuación Z donde los resultados fueron:

- ❖ **Turno mañana:** El promedio está a 0.11 desviaciones estándar de un promedio de 8, por lo cual el rendimiento es ligeramente inferior.
- ❖ **Turno tarde:** El promedio está a 0.14 desvíos estándar del promedio establecido.
- ❖ **Turno completo:** Es el grupo más alejado ya que su promedio está a 0.78 desviaciones estándar, indicando un rendimiento significativamente inferior a los otros turnos.

Por otro lado, se aplicó el mismo análisis pero esta vez teniendo en cuenta el tipo de gestión de la escuela. Los resultados fueron los siguientes:

- ❖ **Escuelas públicas:** El promedio está a 0.33 desviaciones estándar, ligeramente inferior a un promedio de 8.
- ❖ **Escuelas privadas:** El promedio está a -0.49 desvíos estándar del promedio. Esto quiere decir que el rendimiento es superior a un estudiante con promedio de 8.

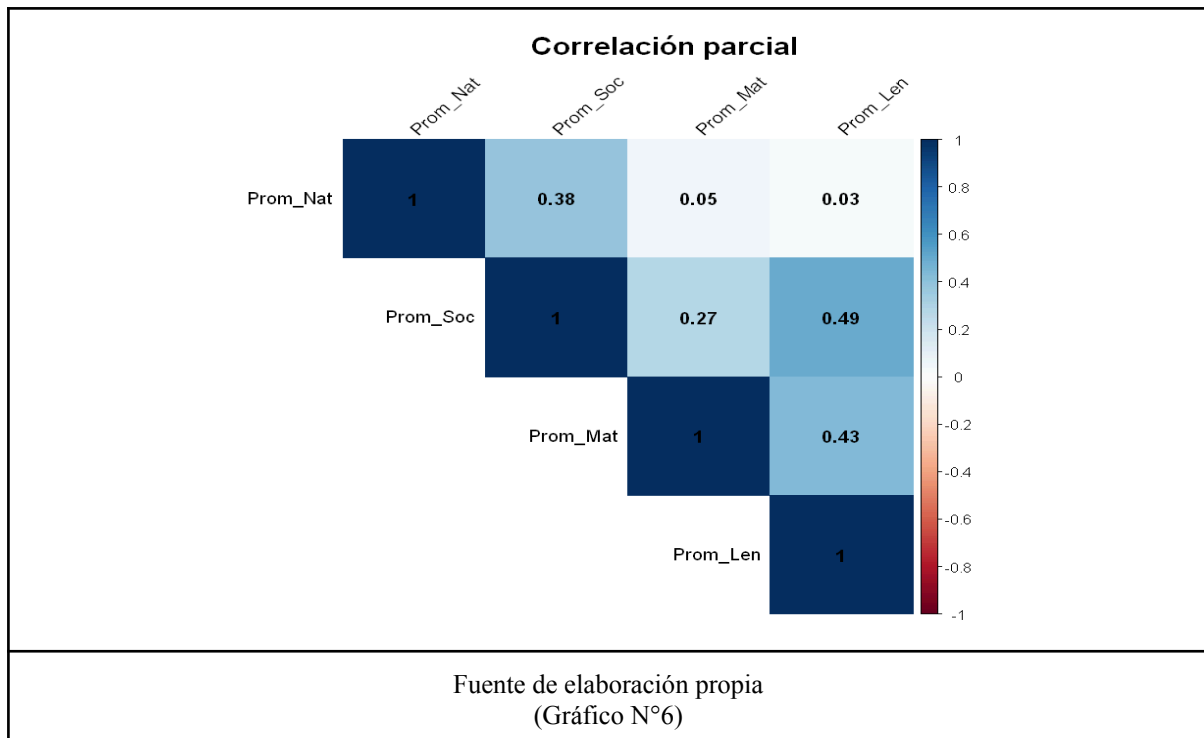
Los estudiantes de primer grado del Departamento Victoria que asisten a colegios privados tienen un rendimiento superior en Matemáticas que sus pares de escuelas públicas.

Los promedios de las materias comunes en primaria, al no seguir una distribución normal, cuando se calculen las correlaciones, se utilizará el coeficiente de Spearman.

Resumidamente, las correlaciones más fuertes fueron entre los promedios de:

- Ciencias Sociales con Matemática (0.8).
- Ciencias Sociales y Lengua (0.82).
- Lengua y Matemática (0.8).
- Ciencias Sociales y Ciencias Naturales (0.67)

Por eso se calcularon las correlaciones parciales para saber si realmente están correlacionadas entre sí o sufren el efecto de otras materias, los resultados encontrados son:



Como muestra el gráfico N°6, las correlaciones entre los promedios de las materias troncales no son muy fuertes. La más destacada es Cs. Sociales ~ Lengua con una correlación parcial de 0.49.

Para comparar los promedios de Matemática y Lengua de los alumnos de primer grado, se utilizó el test de Wilcoxon ya que los datos no tenían una distribución normal. Entonces el test está basado en evaluar las medianas.

Tanto en Matemática como Lengua nuestras hipótesis eran:

- H0: La mediana de los alumnos es 8.
- H1: La mediana de los alumnos es mayor a 8.

Las salidas de estos tests fueron concisos, el p-value en los dos casos nos dio 1, indicando que no hay evidencia suficiente para rechazar la hipótesis nula, dando entender que las medianas de los promedios son iguales a 8 o inferiores.

Ahora comparamos los promedios de Lengua entre escuelas de gestión públicas y privadas. Para esto se utilizó el test de Wilcoxon como en el caso anterior porque nuestros datos no siguen una distribución normal. Nuestras hipótesis fueron:

- H0: Los promedios de los alumnos de primer grado en Lengua son iguales en escuelas privadas y públicas.
- H1: Los promedios de los alumnos de primer grado en Lengua en escuelas privadas son mayores que en escuelas públicas.

El resultado del test de Wilcoxon nos arrojó que hay suficiente evidencia para rechazar  $H_0$  a favor de  $H_1$  ( $p\text{-value} = 1.401e-11$ ). Por lo que podemos concluir que los promedios en Lengua para los alumnos de primer grado son mejores en escuelas privadas que en públicas.

Uno de los análisis más desafiantes y entretenidos son los tests de más de dos medias independientes. En este caso, el primer planteo que tuvimos fue investigar si los tres turnos presentes en nuestra base de datos de primaria tenían la misma media o si alguna difiere significativamente en sus promedios de Matemática. Las hipótesis de este test son:

→  $H_0$ : Los tres turnos de primero tienen la misma media.

→  $H_1$ : Un grupo de datos que tiene una media distinta.

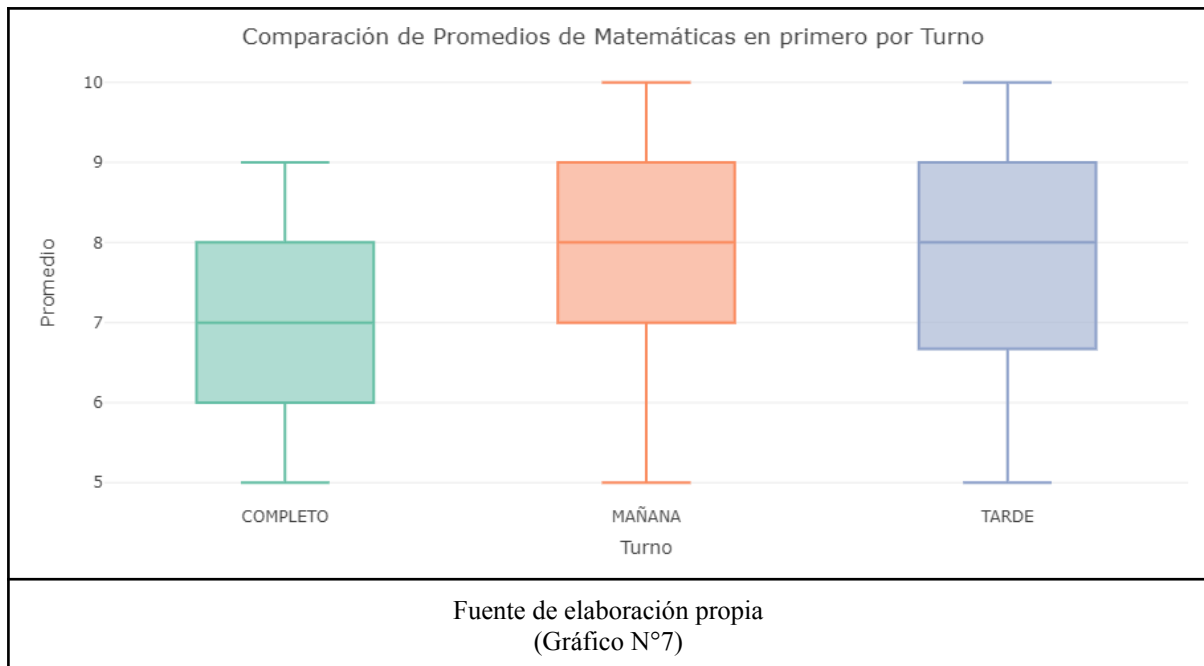
Con un nivel de significancia de 0.05, rechazamos la hipótesis nula a favor de  $H_1$  diciendo que hay un turno que su media es distinta a las demás. La potencia de este test es muy alta, con valores de 0.92, 0.99, 0.99.

Para saber exactamente qué grupo de datos difiere de los demás se le aplicó el test HSD de Tukey, este no dió que el turno que difiere con los demás es el completo como lo muestra la tabla N°4, mientras que entre el turno mañana y tarde las medias son muy similares.

Turno	Medias	p-adj
MAÑANA - COMPLETO	7.9 - 7.0	0.0000013
TARDE - COMPLETO	7.8 - 7.0	0.0000140
TARDE - MAÑANA	7.8 - 7.9	0.8557415
Fuente de elaboración propia (Tabla N°4)		

Pero el test de ANOVA tiene supuestos para saber si son representativos los resultados dados. Por eso a la hora de comprobar si los datos son de distribuciones normales, tenemos que descartar este análisis, ya que ningún conjunto de datos superó el test de normalidad. Así que aplicamos un test no paramétrico con los mismos conjuntos de datos.

En este caso también tomamos la hipótesis alternativa y decimos que hay un turno que difiere de los demás en cuanto a los promedios de Matemática. Las nuevas salidas dieron lo mismo, difiriendo levemente en el  $p\text{-value}$  entre la comparación de a pares de los turnos. El turno completo es el que difiere de las demás, lo cual tiene mucha lógica si vemos el siguiente gráfico:

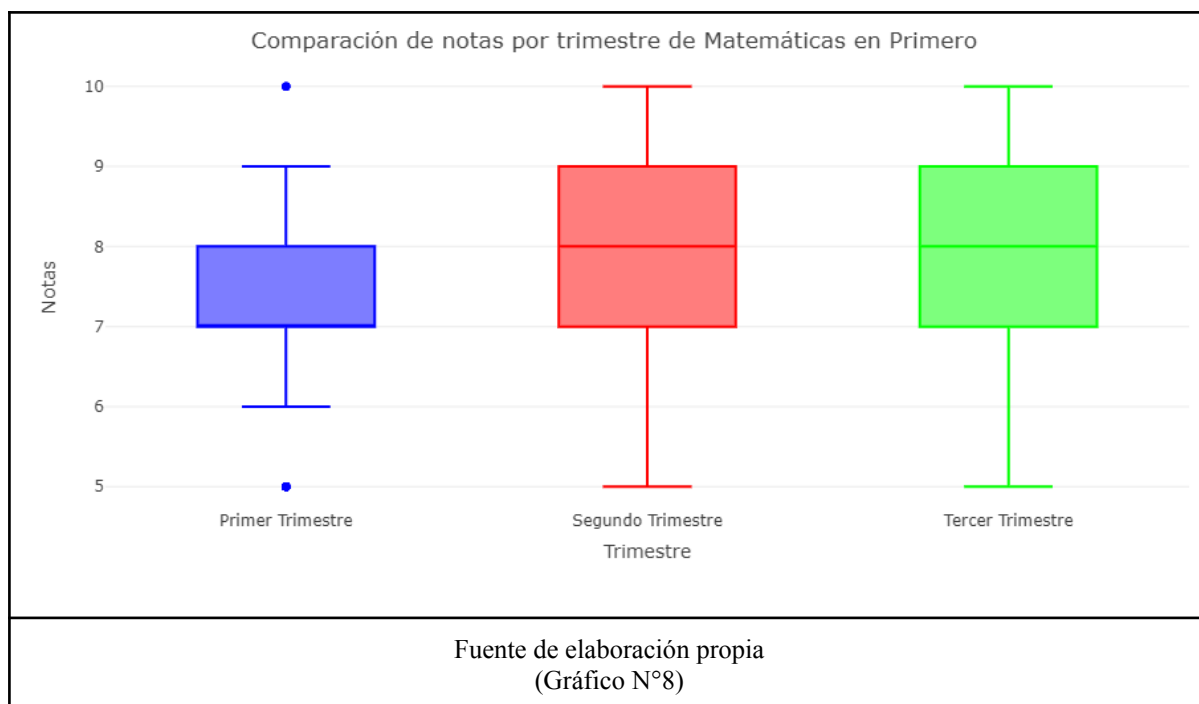


El mismo análisis se desarrolló con los promedios de Lengua. Los resultados fueron exactamente iguales. Inicialmente, con el test de ANOVA rechazamos  $H_0$ , pero al verificar la normalidad de los datos, descubrimos que estos no seguían una distribución normal. Por lo tanto, procedimos a utilizar el test de Kruskal-Wallis y pruebas pareadas (Pairwise). Los resultados fueron los siguientes:

Turno	Medianas	p-value
MAÑANA - COMPLETO	8 - 7	1.7e-06
TARDE - COMPLETO	8 - 7	8.9e-05
TARDE - MAÑANA	8 - 8	0.46
Fuente de elaboración propia (Tabla N°5)		

Como hemos visto, tanto en Matemática y Lengua, el turno completo es el que tiene una media diferente.

Pero además usamos el ANOVA de medias pareadas para saber si las notas de los trimestres de Matemática tenían la misma media o cual trimestre era diferente. Para ilustrar el planteo, decidimos graficarlo con un boxplot comparativo:



A simple vista parece ser que el primer trimestre tiene una distribución diferente a los otros dos trimestres. Para constatar esto aplicamos el ANOVA de medias pareadas y nos confirmó que sí hay diferencias entre los trimestres. Sin embargo, al comprobar los supuestos, encontramos que los datos no siguen una distribución normal y sus varianzas no son constantes. Por lo tanto, descartamos el análisis inicial y procedimos con un test no paramétrico, el test de Friedman.

El test de Friedman nos dice que si hay diferencias entre las medianas de los grupos ya que devuelve un p-value menor a  $2.2e-16$ . Este resultado es confiable ya que su test de potencia es de 0.99, siendo muy fuerte. Sorprendentemente, cuando aplicamos el test de Pairwise para identificar entre qué trimestres están las diferencias, encontramos que todos los trimestres difieren entre sí, no hay medianas iguales o cercanas entre los trimestres.

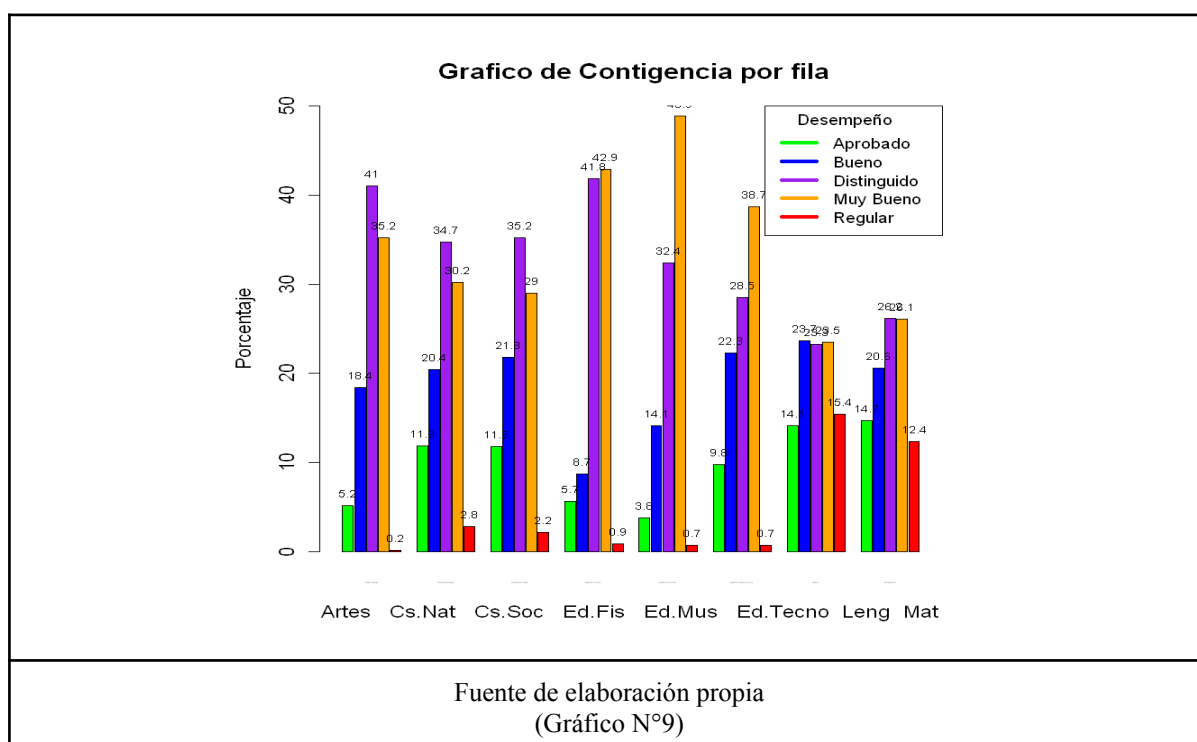
Para el análisis bivariado cualitativo decidimos tomar las variables de las asignaturas (filtrando las materias que consideramos comunes en primaria) y el desempeño logrado por los estudiantes en cada una de las materias. La tabla ordenada y con sus totales marginales tanto en filas como en columnas quedó conformada de la siguiente manera:

	Regular	Aprobado	Bueno	Muy Bueno	Distinguido	Total
Artes Visuales	1	30	107	204	238	580
Cs. Naturales	16	68	117	173	199	573
Cs. Sociales	12	64	118	157	191	542

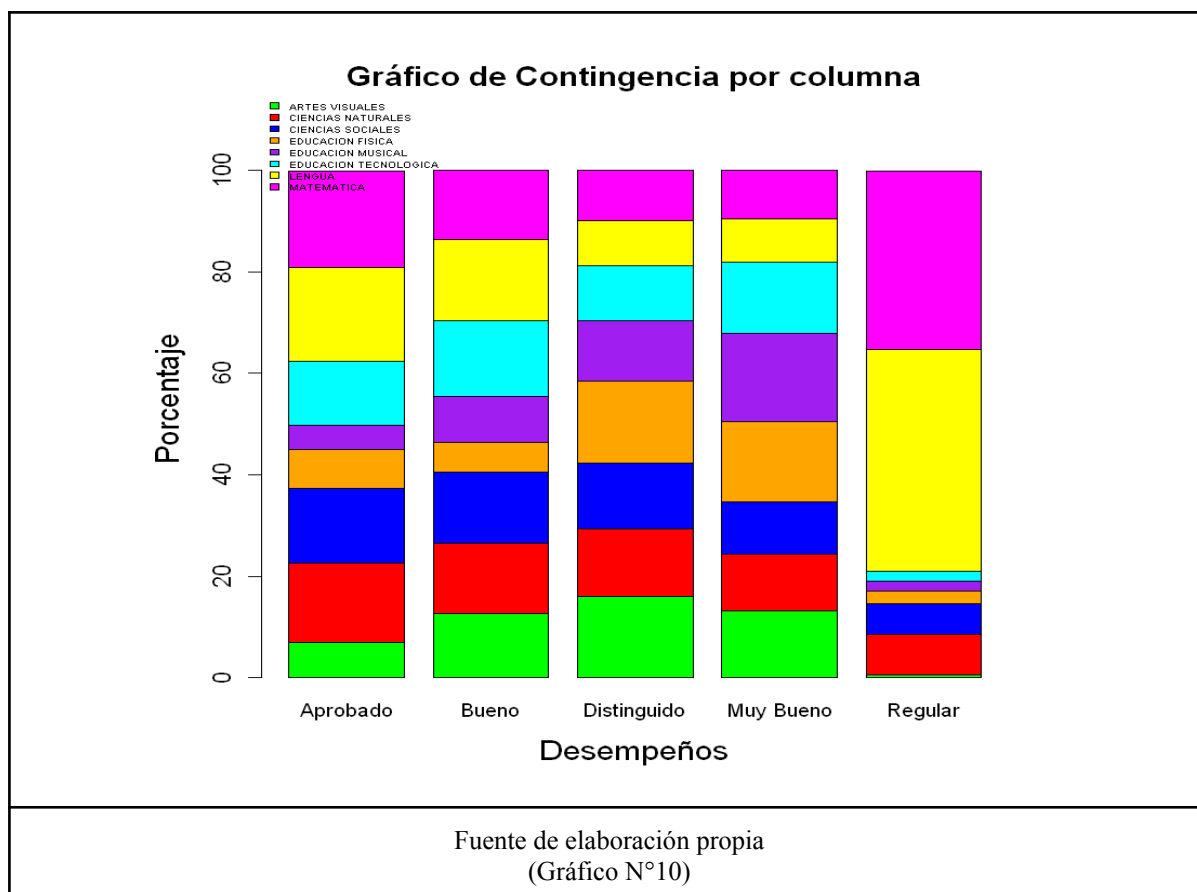


Ed. Física	5	33	50	246	240	574
Ed. Musical	4	21	77	267	177	546
Ed. Tecnológica	4	55	123	218	161	564
Lengua	87	80	134	133	132	566
Matemática	70	83	116	147	147	564
Total	199	434	845	1545	1486	4509
Fuente de elaboración propia (Tabla N°6)						

La tabla de contingencia con los porcentajes por fila fue representada con un gráfico de barras adosadas:



Mientras que la tabla de contingencia con los porcentajes por columna fue representada con un gráfico de barras apiladas:



Como muestran la tabla N°6 y los gráficos, podemos determinar que según el tipo de materia, “causa” el desempeño final de cada alumno. Vemos como en materias más prácticas hay menos desempeños regulares, mientras que en materias más teóricas como lo son Matemática y Lengua, tienen porcentajes elevados.

Gracias al test de chi-cuadrado, podemos afirmar nuestra hipótesis de que hay relación entre las variables. La fuerza de esta relación es débil, ya que el coeficiente de Cramer nos dio 0.18.

En cuanto a las observaciones esperadas si fueran variables independientes se muestran una mayor frecuencia de la esperada en:

- Lengua ~ Regular: El residuo ajustado es de 13.6.
- Educación Musical ~ Muy Bueno: El residuo ajustado es de 7.7.

Mientras que las materias que se esperaban más frecuencias de las observadas son:

- Educación Física ~ Bueno: El residuo ajustado es de -6.6.
- Artes Visuales ~ Regular: El residuo ajustado es de -5.3.

Estos resultados demuestran que nuestra hipótesis es correcta. El desempeño de los estudiantes está condicionado según la materia

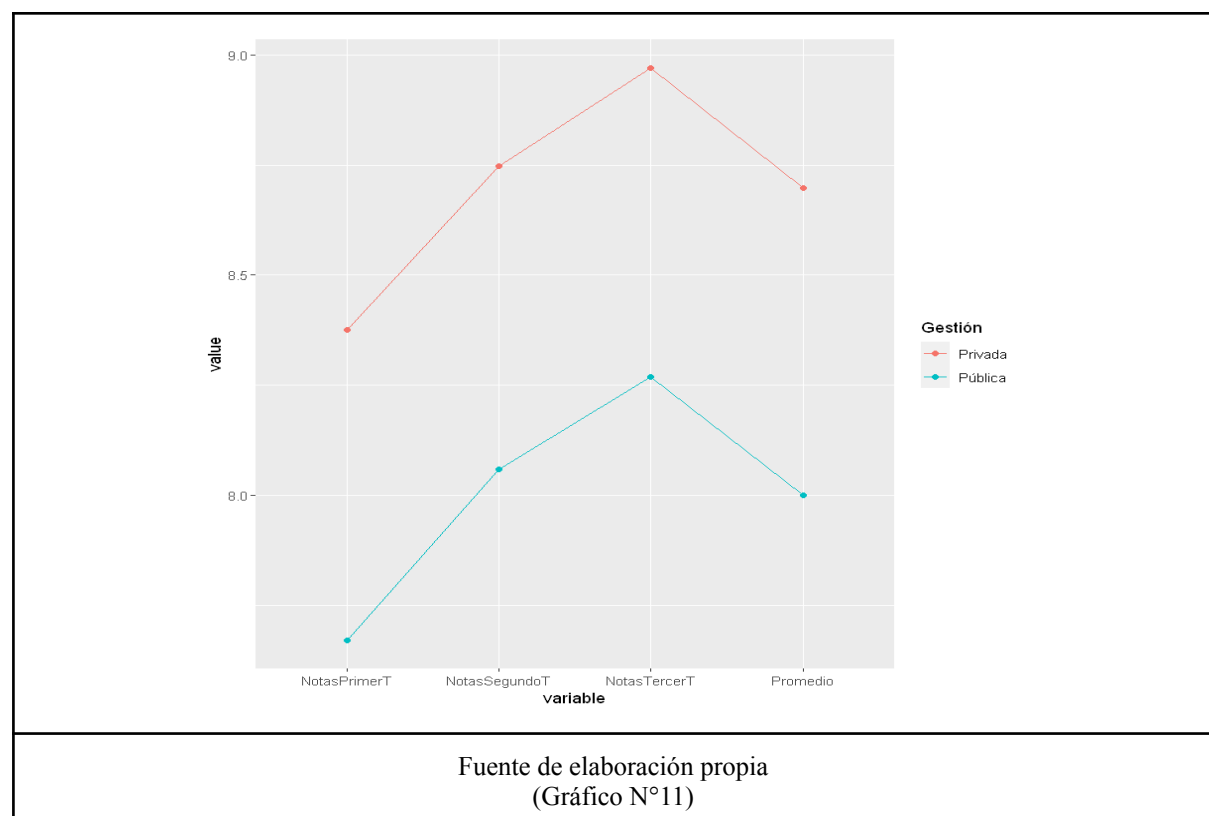
El análisis multivariado realizado abarcó una amplia gama de investigaciones, generando resultados diversos y significativos. Los hallazgos más relevantes encontrados fueron:

La búsqueda de valores outliers a través de la distancia de Mahalanobis, utilizando las variables numéricas como las notas de los trimestres y el promedio. Esta búsqueda finalizó con un total de 41 valores atípicos.

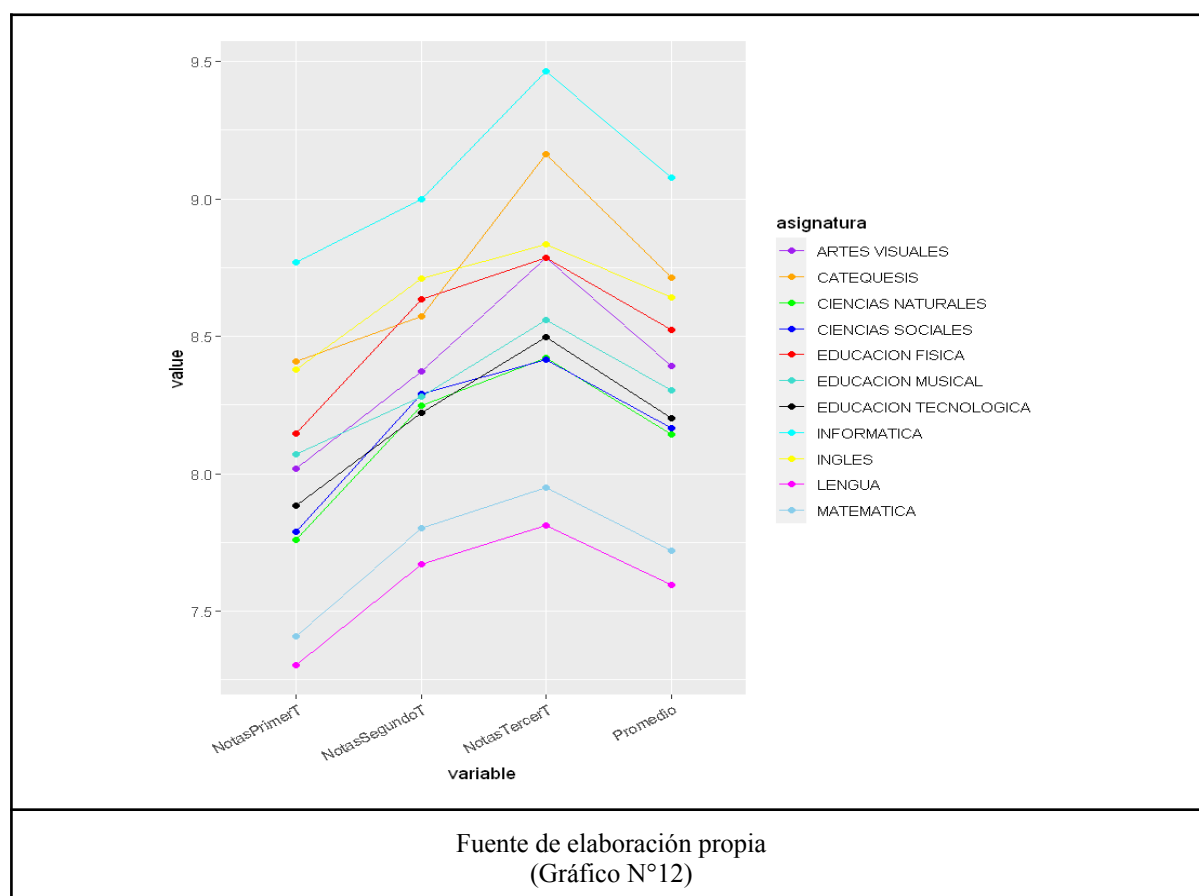
Se llevó a cabo un análisis de la distribución normal multivariada utilizando la función de Mardia. Sin embargo, los resultados indicaron que esta distribución no se cumple en el conjunto de datos.

Se compararon dos poblaciones multivariadas independientes como lo son el tipo de gestión que tienen las escuelas en el departamento. Para hacer posible esto, se asumió que las medias de las variables para cada grupo se aproximan a una distribución normal multivariada, debido al Teorema del Límite Central Multivariado (TLCM).

Los resultados del test de Hotelling indicaron que se rechaza la hipótesis nula a favor de la hipótesis alternativa, lo que sugiere que los vectores medios difieren entre los tipos de gestión. Este hallazgo se respalda visualmente en el Gráfico N°11, donde se observa claramente la diferencia significativa entre los grupos.



Para concluir el análisis del primer grado, nuestro objetivo fue comparar las medias de las materias para determinar cuál es la mejor. Esta comparación nos permitió obtener información relevante sobre el desempeño en cada asignatura.



La representación gráfica muestra como las materias Lengua y Matemática están muy por debajo de las demás. Hay que destacar la performance en Informática y Catequesis siendo las mejores.

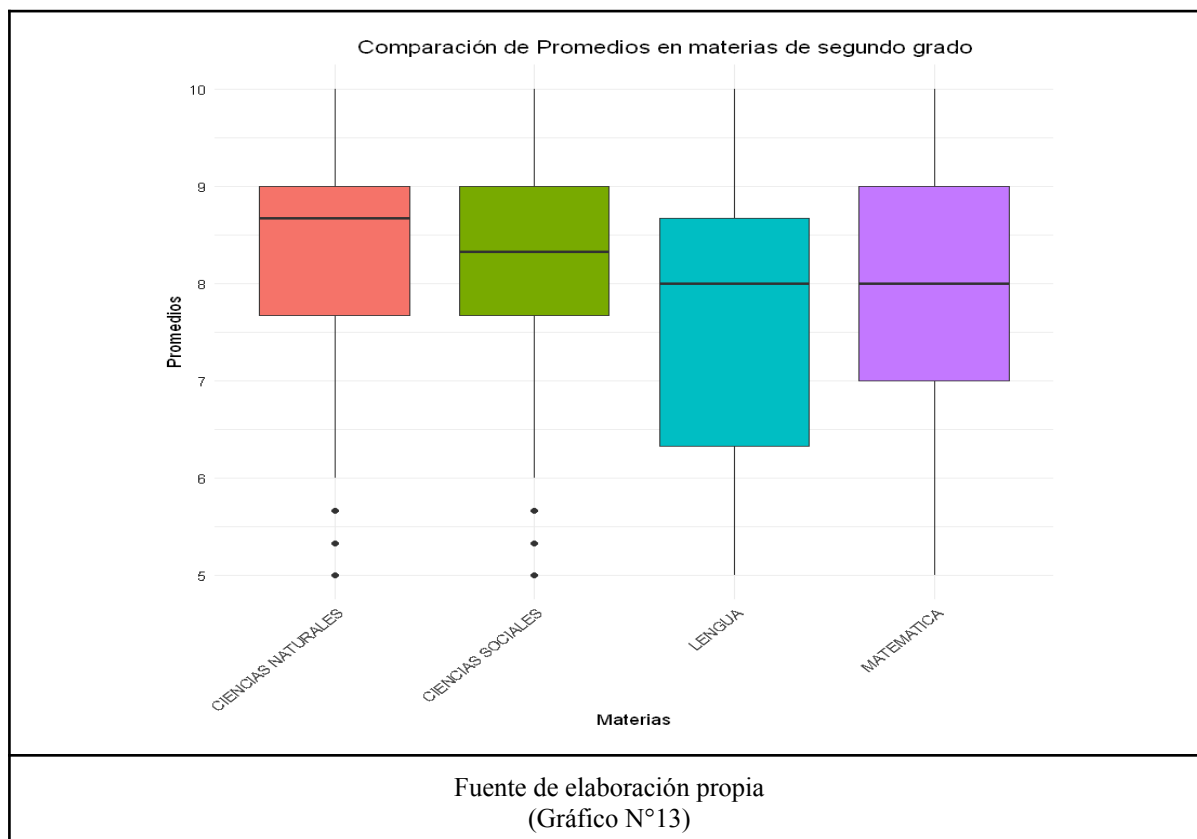
## Segundo grado

Nuestra base de datos de segundo grado quedó conformada por 563 matriculados, un 14% menos que los datos oficiales relevados por la C.G.E.

El análisis de las distintas medidas estadísticas descriptivas dieron distintos y sorprendentes resultados, como:

- ❖ Lengua siendo la que peor promedios tuvo y mayor variabilidad respecto a la media. El 75% está por debajo de un promedio de 8.67.
- ❖ Las ciencias, tanto Naturales como Sociales presentaron outliers a los promedios por debajo de 5.68.
- ❖ Ciencias Naturales posicionándose como la que mejor rendimiento ha logrado. Un 50% de los matriculados consiguió un promedio mayor a 8.67.

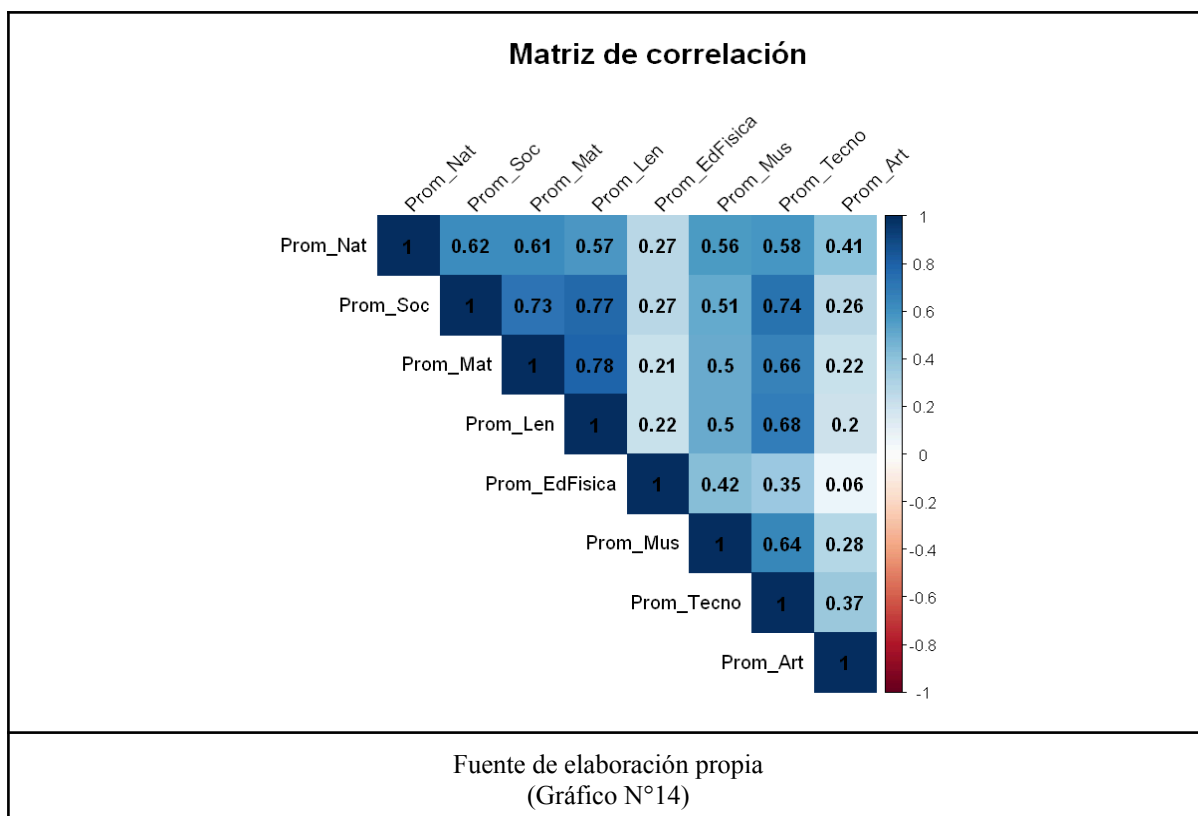
- ❖ Ciencias Sociales le sigue a Naturales como la mejor materia en segundo grado.
- ❖ El rango de promedios en las 4 materias es el mismo [5 - 10].
- ❖ En el siguiente gráfico de boxplot por grupo, se aprecian los resultados comentados y muchos más.



En cuanto a un promedio igual a 8 en Matemática en escuelas privadas en este grado, es inferior a la puntuación Z obtenida, ya que el resultado es -0.66. Mientras que en escuelas públicas, un promedio de 8 es mejor que el punto Z de este grupo.

Se puede ver una diferencia notable entre los promedios de Matemática según el tipo de gestión de las escuelas.

Ahora nos enfocaremos en las correlaciones de las materias más importantes a nivel primario:

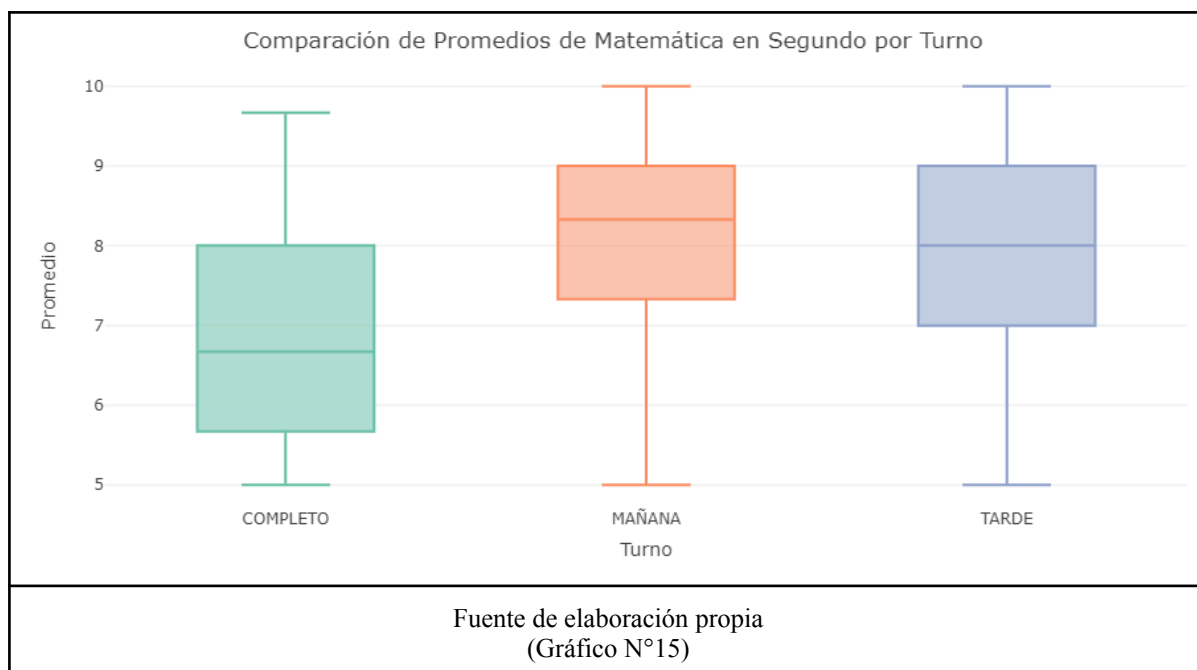


Como muestra el gráfico N°14, hay muchas correlaciones altas y moderadas, pero si le calculamos la correlación parcial ¿Se mantendrán la dirección y fuerza de las correlaciones? Previo al cálculo de las correlaciones parciales, se eliminarán las variables de los promedios de Educación Física y Artes Visuales ya que presentan correlaciones directas bajas.

Cuando se calcularon las correlaciones parciales, solo dos correlaciones son moderadas, estas son entre los promedios de las materias de Lengua ~ Matemática y Educación Tecnológica ~ Cs. Sociales.

Como ya sabemos que un promedio en Matemática es más alto en escuelas privadas que en escuelas públicas. Ahora lo constatamos en Lengua con un test de dos medias independientes. Como la distribución de los datos no es normal, se aplicará con un test no paramétrico. Esta prueba refleja que se rechaza la hipótesis nula (promedios iguales en escuelas privadas y públicas) a favor de la hipótesis alternativa (los promedios en Lengua son mayores en escuelas privadas que en escuelas públicas).

En cuanto al test de ANOVA para más de dos medias independencia lo utilizamos para saber si los promedios de Lengua eran iguales en los diferentes turnos. Con el siguiente gráfico podemos ver un indicio de lo que nos puede arrojar la prueba.



Como se puede aparentar, el test nos dió que hay diferencias entre las medias. Especialmente las que se comparan con el turno completo, ya que el par de turno mañana y tarde no tienen una diferencia significativa. Pero este test no es representativo porque los datos no vienen de una distribución normal. Por lo cual, se aplicará un test donde se tenga en cuenta la mediana, como lo es la prueba de Wilcoxon. Este nos dió resultados similares, donde hay diferencias entre las medianas.

La asignatura Matemática se empleó el ANOVA para más de dos medias pareadas, donde se usaron las notas de los trimestres como variables.

A pesar de que el resultado diera que si hay diferencias entre las notas, los supuestos de las varianzas constantes y distribución normal, no se cumplen en este caso. Por lo cual, derivamos la investigación al test de Friedman donde se encontró diferencias significativas entre las notas de los trimestres. Este resultado es muy de confiar, ya que la potencia conseguida es de un 0.99.

El análisis bivariado cualitativo fue el mismo que se realizó en primer grado, ya que nos resulta interesante investigar la performance lograda por los estudiantes en los diferentes grados.

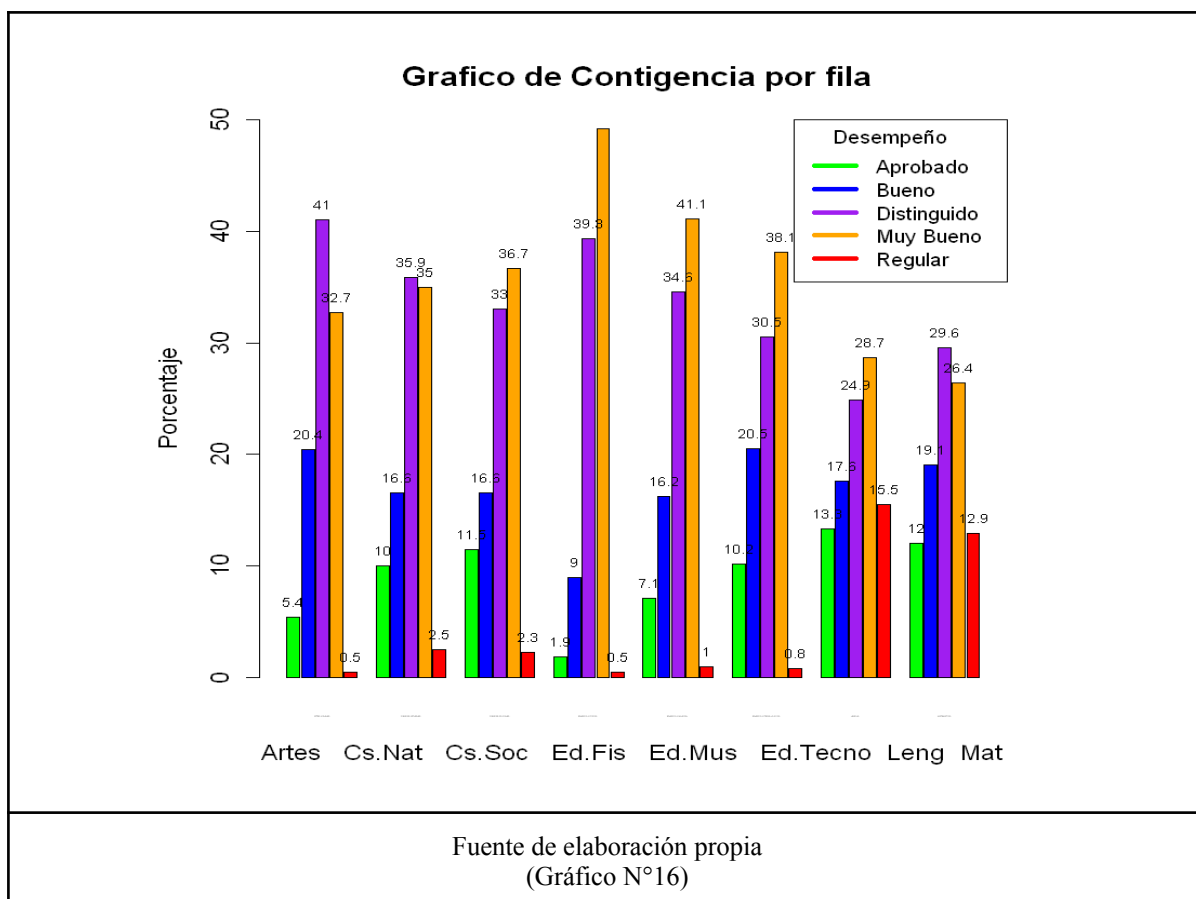
En este grado, los resultados serán representados con una tabla donde la primera fila de los datos lo compondrán las observaciones y la segunda fila el porcentaje total, la tabla quedó conformada de la proxima manera:

	Regular	Aprobado	Bueno	Muy Bueno	Distinguido	Total
Artes Visuales	3 (0.1%)	32 (0.7%)	121 (2.8%)	194 (4.4%)	243 (5.5%)	593 (13.5%)
Cs. Naturales	14 (0.3%)	56 (1.3%)	93 (2.1%)	96 (4.5%)	201 (4.6%)	560 (12.7%)
Cs. Sociales	12 (0.3%)	61 (1.4%)	88 (2.0%)	195 (4.4%)	175 (4.0%)	531 (12.1%)
Ed. Física	3 (0.1%)	11 (0.3%)	51 (1.2%)	279 (6.3%)	223 (5.1%)	567 (12.9%)
Ed. Musical	5 (0.1%)	37 (0.8%)	84 (1.9%)	213 (4.8%)	179 (4.1%)	518 (11.8%)
Ed. Tecnológica	4 (0.1%)	54 (1.2%)	108 (2.5%)	201 (4.6%)	161 (3.7%)	528 (12.0%)
Lengua	85 (1.9%)	73 (1.7%)	97 (2.2%)	158 (3.6%)	137 (3.1%)	550 (12.5%)
Matemática	71 (1.6%)	66 (1.5%)	105 (2.4%)	145 (3.3%)	163 (3.7%)	550 (12.5%)
Total	197 (4.5%)	390 (8.9%)	747 (17%)	1581 (36.0%)	1482 (33.7%)	4397 (100%)
Fuente de elaboración propia (Tabla N°7)						

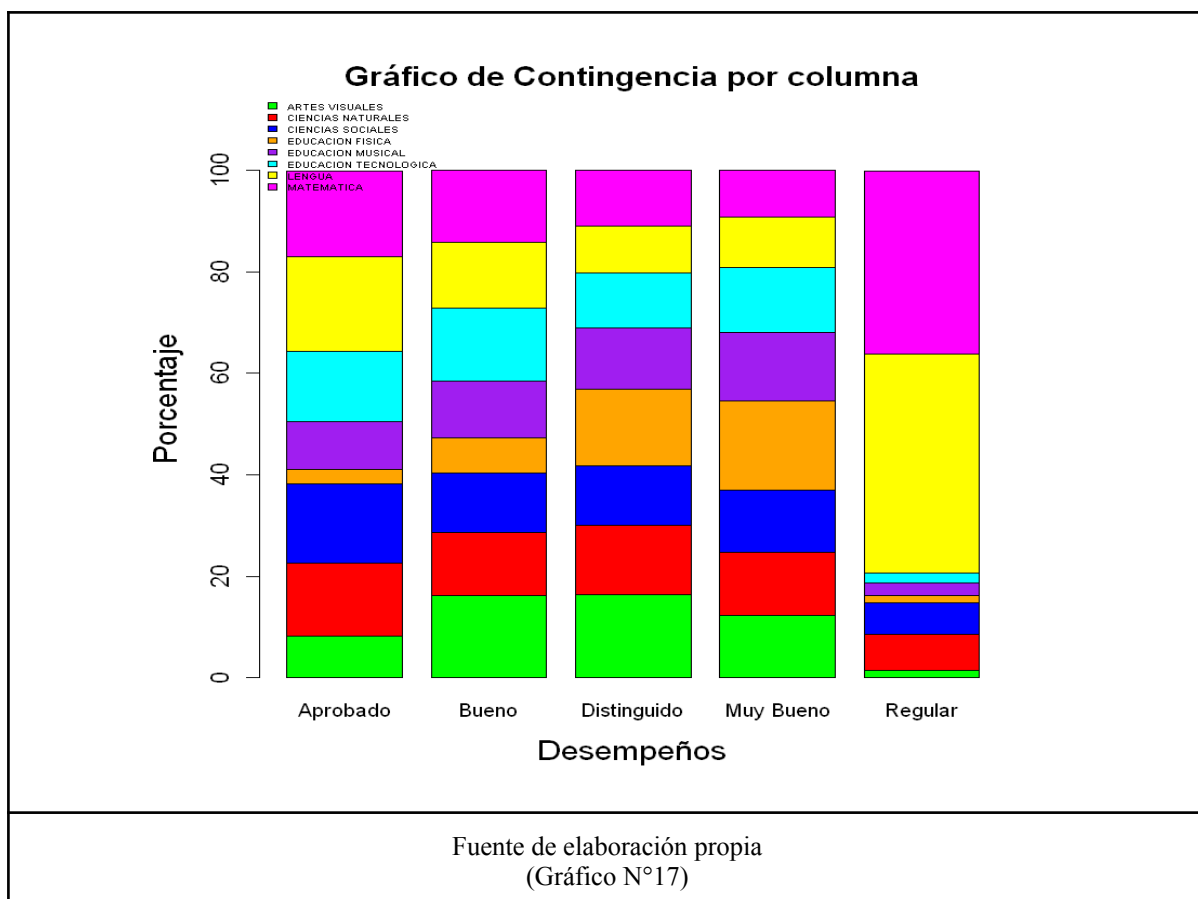
Como podemos leer de la tabla N°7 , un 33.7% consiguió el desempeño más alto, mientras que un 4.5% se desempeñó en las materias de forma regular.

A continuación, con gráficos de barras adosadas primero y posteriormente con un gráfico de barras apiladas, representaremos las tablas de contingencia de los porcentajes por fila y por columna. Los gráficos en cuestión son:





La mayoría de asignaturas tienen un porcentaje en los desempeños de alto nivel que supera el 30%. Mientras que Matemática y Lengua, en esos niveles no supera el 30%. Además, se observa que la barra roja, que representa los desempeños regulares, generalmente se mantiene por debajo del 3%. No obstante, en Lengua y Matemática, hay un aumento significativo, alcanzando el 15% y 12% respectivamente.



Como es presentado en el gráfico de barras apiladas, Matemática y Lengua concentran el 79.1% de desempeños regulares.

Toda la información recopilada nos permite plantear que según la materia, el estudiante logra un desempeño más alto o más bajo. Por lo cual nuestra hipótesis alternativa es que hay relación entre las variables.

Con la prueba de chi cuadrado tenemos la evidencia suficiente para rechazar  $H_0$  y afirmar nuestra hipótesis. La fuerza de esta asociación de las variables es débil.

Nuestra investigación para segundo grado en análisis multivariado fue la comparación de tres o más poblaciones multivariadas independientes, tomando las variables de las notas trimestrales y el promedio, en los diferentes turnos.

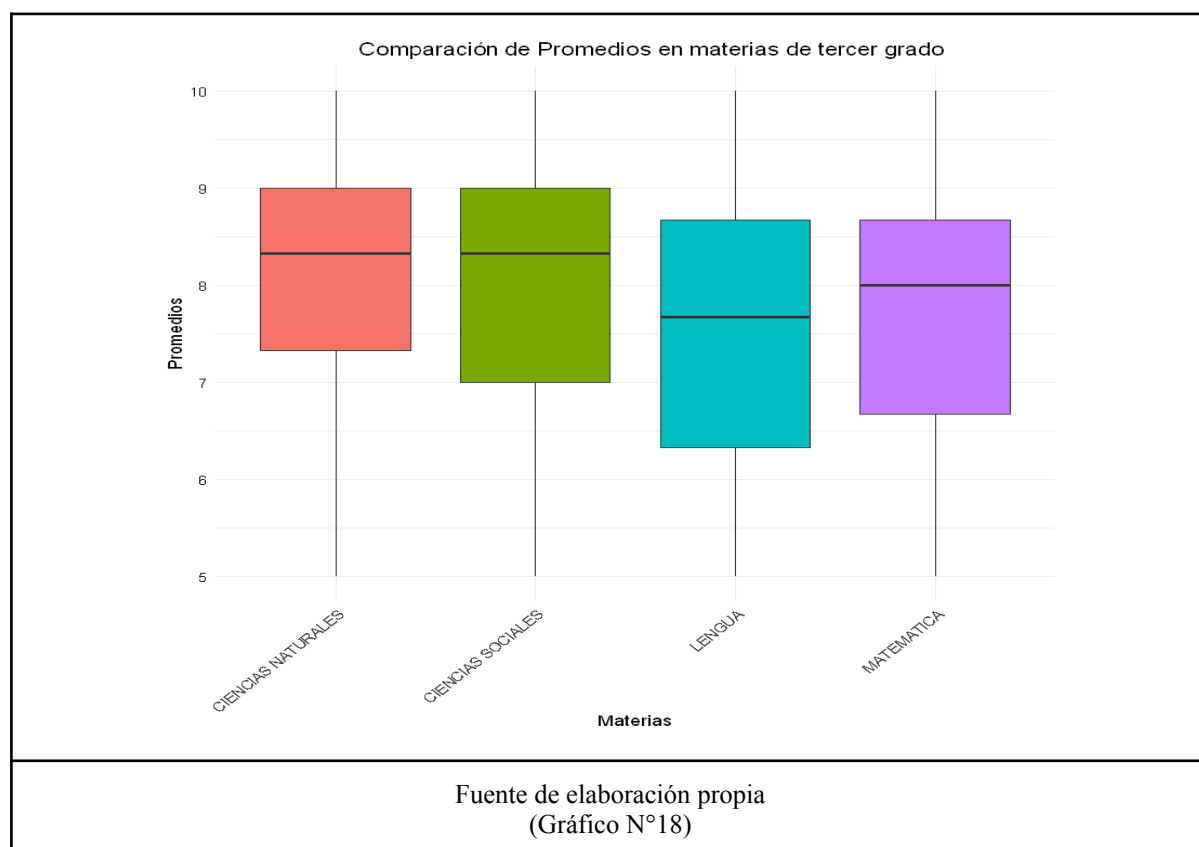
Pero antes de hacer el MANOVA, calculamos la normalidad multivariada con la función de Mardia, la cual nos devolvió que los datos no siguen una distribución normal multivariada. Pero teniendo en cuenta el teorema del límite central multivariado, asumimos que nuestros datos tienen una media que se aproxima a una distribución normal. Por lo cual, planteamos la prueba.

Los resultados de la prueba fueron que hay diferencias significativas entre los vectores medios.

### Tercer grado

Contamos con 567 estudiantes matriculados en nuestra base de datos para analizar. Después de aplicar medidas estadísticas descriptivas en los promedios en las materias consideradas troncales, algunos de los resultados fueron:

- ❖ Lengua es la que más bajos promedios tiene de las 4 materias. El 50% de los promedios no supera el 7.67.
- ❖ Matemática tiene un 75% de los promedios por encima de 6.67.
- ❖ Ciencias Naturales es la que mejores promedios ha logrado, solo un 25% está por debajo del promedio 7.33.
- ❖ Ninguna materia presenta outliers porque el rango de datos es de [5 - 10] y los límites tanto inferiores y superiores están muy lejos de esos valores.



Un promedio de 8 en Lengua lo podemos comparar con dos grupos de datos independientes a través de la puntuación Z. En una escuela pública, en tercer grado, el promedio es mayor ya que está a 0.53 desvíos de la media. En cambio, en escuelas privadas, el promedio está a -0.51 desviaciones de la media.

Se puede ver una gran diferencia entre un alumno que tiene un promedio de 8 en Lengua y el tipo de gestión. Para una escuela pública sería uno de los mejores promedios, pero para una escuela privada sería del grupo del medio.

Al calcular las correlaciones entre las materias comunes en primaria, no se hallaron grandes resultados. La relación más destacada es entre los promedios de Cs. Sociales y Lengua (correlación parcial de 0.31).

Para saber si lo mismo que en la puntuación Z sucede en Matemática, se le aplicará el test de dos medias independientes. Será con un test no paramétrico porque los datos no vienen de una distribución normal. El p-value de la prueba arrojó 2.235e-15, por lo que tenemos suficiente evidencia para rechazar  $H_0$  a favor de  $H_1$  y afirmar que los promedios de Matemática son mayores en las escuelas privadas que en las escuelas de gestiones públicas.

Ahora el análisis es para los promedios de Matemática, pero en esta ocasión buscaremos diferencias entre los distintos turnos con el test de ANOVA. La prueba nos mostró que si hay diferencias entre los diferentes turnos.

Como viene sucediendo, las diferencias están entre el turno completo y tarde/mañana. Mientras que entre el turno mañana y tarde son muy similares. Esto fue descubierto por el test HSD de Tukey.

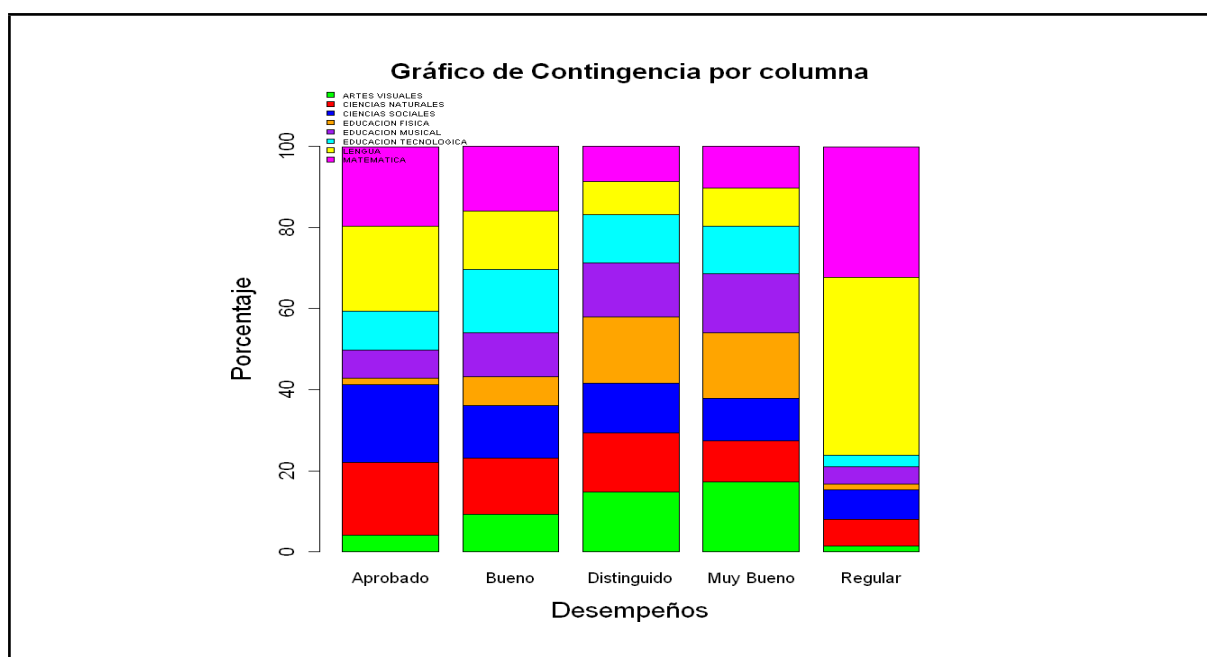
Pero para saber si los resultados son representativos al conjunto de datos, las distribuciones tienen que ser normales. En este caso, no se cumplen los supuestos.

Entonces basaremos nuestro análisis en un test no paramétrico que los compare según la mediana de cada conjunto. El test de Kruskal-Wallis también nos refleja que existen diferencias, estas diferencias están en:

<b>Turno</b>	<b>Medianas</b>	<b>p-value</b>
MAÑANA - COMPLETO	8 - 6.67	7.0e-09
TARDE - COMPLETO	8 - 6.67	1.4e-07
TARDE - MAÑANA	8 - 8	0.93
Fuente de elaboración propia (Tabla N°8)		

Los desempeños en las materias en tercer grado se distribuyen de una forma peculiar, como lo muestra la tabla N°9 , la mayoría de materias tienen un alto porcentaje en Muy Bueno y bajos porcentajes en Regular. Esto es realmente bueno, significa que el progreso educativo es positivo. A continuación, la tabla con las frecuencias y porcentaje totales:

	Regular	Aprobado	Bueno	Muy Bueno	Distinguido	Total
Artes Visuales	3 (0.1%)	18 (0.4%)	79 (1.7%)	285 (6.2%)	209 (4.6%)	594 (13%)
Cs. Naturales	14 (0.3%)	79 (1.7%)	117 (2.6%)	167 (3.7%)	207 (4.5%)	584 (12.8%)
Cs. Sociales	15 (0.3%)	84 (1.8%)	110 (2.4%)	171 (3.7%)	175 (3.8%)	555 (12.2%)
Ed. Física	3 (0.1%)	7 (0.2%)	59 (1.3%)	268 (5.9%)	231 (5.1%)	568 (12.5%)
Ed. Musical	9 (0.2%)	31 (0.7%)	93 (2.0%)	239 (5.2%)	190 (4.2%)	562 (12.3%)
Ed. Tecnológica	6 (0.1%)	42 (0.9%)	132 (2.9%)	194 (4.3%)	167 (3.7%)	541 (11.9%)
Lengua	91 (2.0%)	93 (2.0%)	121 (2.7%)	155 (3.4%)	118 (2.6%)	578 (12.7%)
Matemática	67 (1.5%)	86 (1.9%)	135 (3.0%)	168 (3.7%)	124 (2.7%)	580 (12.7%)
Total	208 (4.6%)	440 (9.6%)	846 (18.5%)	1647 (36.0 1%)	1421 (31.1%)	4562 (100%)
Fuente de elaboración propia (Tabla N°9)						



Como podemos ver en los porcentajes totales, específicamente en el desempeño Regular hay más frecuencias en áreas como Matemática y Lengua, en el gráfico de barras apiladas se puede apreciar mejor ya que representa a una tabla de contingencia con los porcentajes por columna.

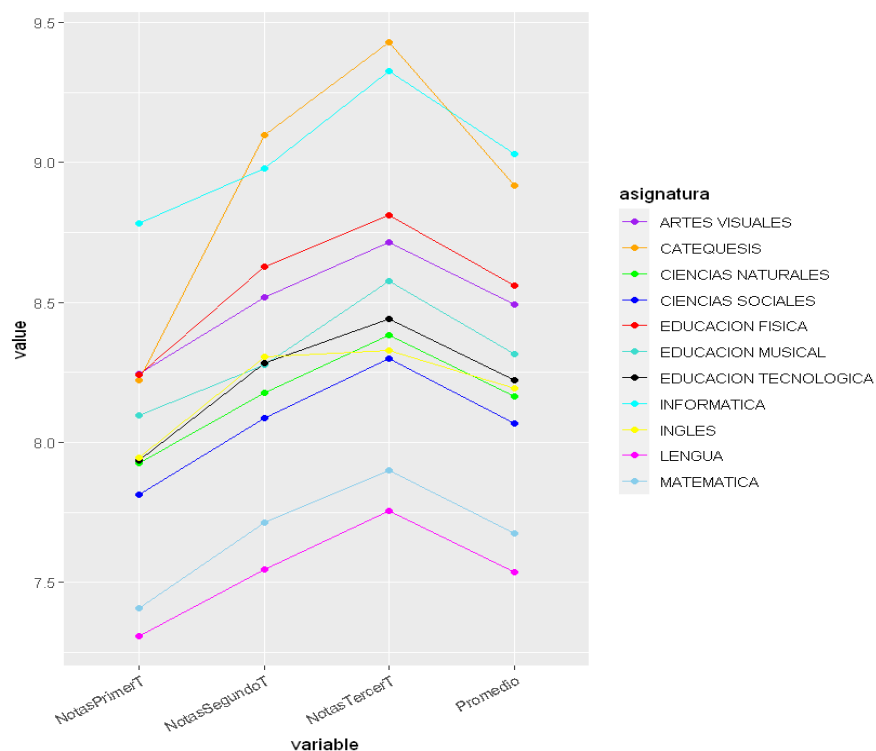
Si vemos de manera creciente del peor desempeño al mejor, podemos ver que cada vez que es mejor el nivel, menos porcentaje ocupan dos materias troncales como lo son Lengua y Matemática. Destacamos el área de trabajo de Educación Física, ya que es la que mayor porcentaje tiene en el mejor desempeño.

Toda esta información, nos permite seguir planteando lo de los grados anteriores. El desempeño de cada estudiante, está ligado al área de aprendizaje. Hay una asociación entre las variables, pero no es tan fuerte como se estimaba. El coeficiente de Cramer apenas dió 0.18, indicando que la fuerza de esta asociación no es tan fuerte.

Para el análisis multivariado nos resultó interesante saber cuántos outliers había en tercer grado si tomamos las variables cuantitativas como las notas en cada trimestre y el promedio. A través de la distancia de Mahalanobis, logramos identificar 36 valores atípicos. Pero esto no es un dato tan representativo, ya que la función de Mardia nos dió que los valores no tienen una distribución normal multivariada.

Aún así, se decidió seguir investigando sobre tercer grado y logramos graficar el rendimiento de las materias en este grado. Los resultados son sorprendentes, ya que se pueden apreciar grupos de rendimiento, donde materias como Catequesis e Informática se destacan muy por encima del resto. Mientras que existe un grupo del medio en cual, en diferentes niveles se están ubicando. Por la parte inferior, encontramos las asignaturas claves, como son Lengua y Matemática. Este gráfico nos demuestra una vez más las diferencias entre las áreas de enseñanza.

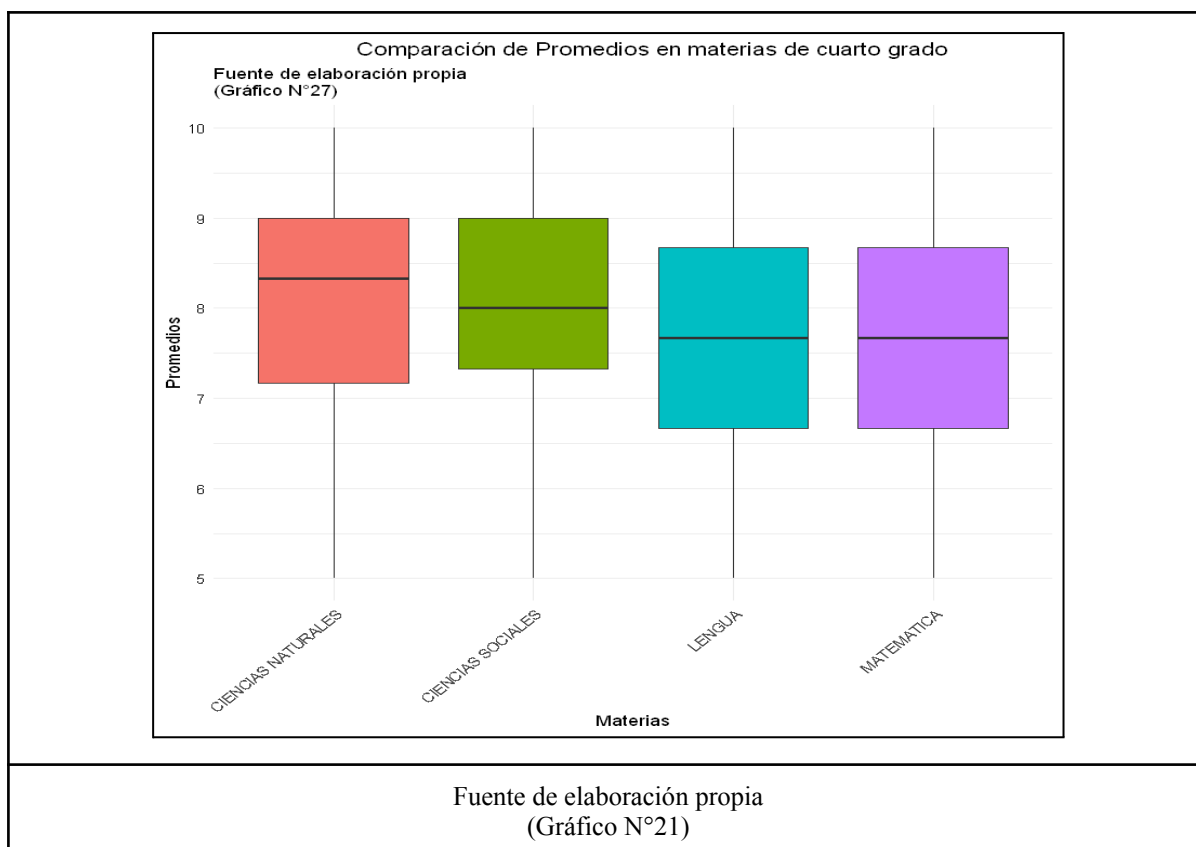
El gráfico en cuestión quedó:



Fuente de elaboración propia  
(Gráfico N°20)

## Cuarto grado

En este grado es donde se encuentra la mayor diferencia entre nuestra base de datos y el relevamiento de datos brindado por la C.G.E. La diferencia asciende a un 23%, a pesar de eso, nuestras muestras se pueden utilizar para el análisis.



Los resultados más destacables de las medidas estadísticas son:

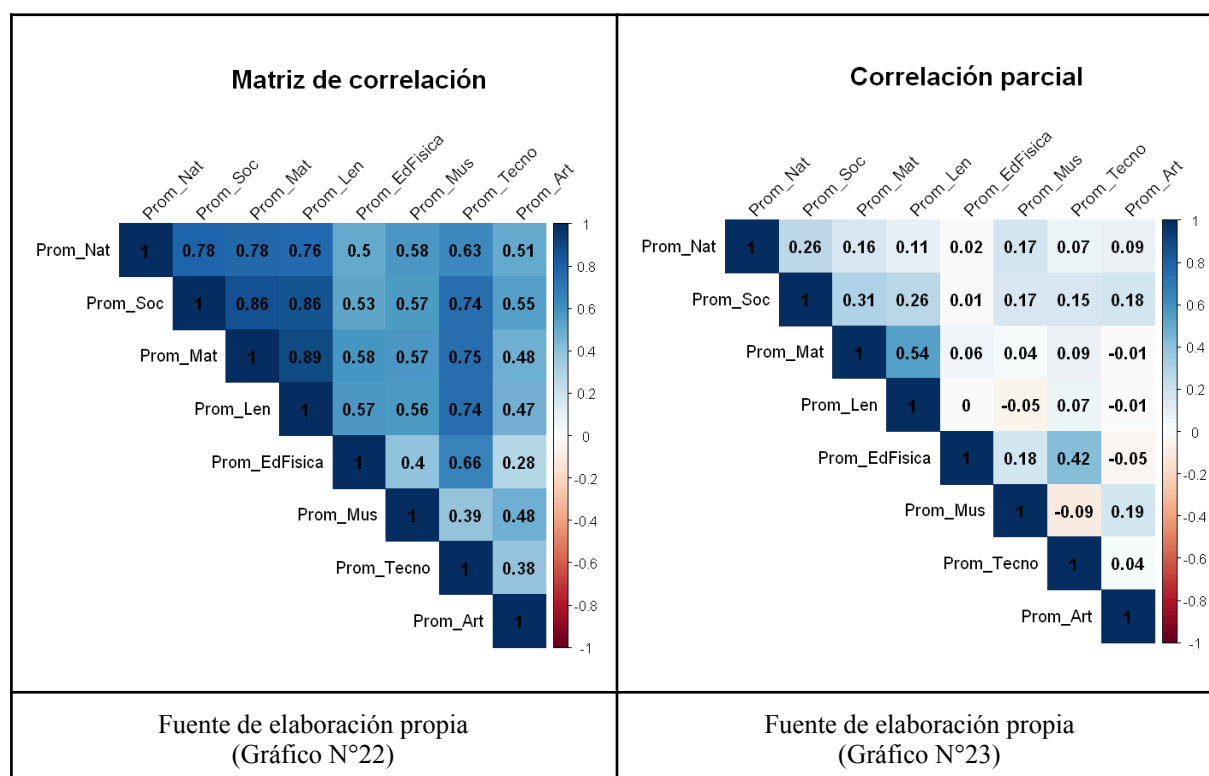
- ❖ Matemática y Lengua son las que peores promedios tienen. Tienen una distribución similar. Un 25% tienen los promedios por encima de 8.67.
- ❖ Un 15% tiene desaprobadas la materia Lengua. El mismo porcentaje sucede en Matemática.
- ❖ La mejor materia sigue siendo Ciencias Naturales como sucede en los anteriores grados, ya que el 50% de los promedios está por encima de 8.33.
- ❖ Ciencias Sociales tiene la caja más achatada. El 25% está situado en un promedio 7.33, el 50% en un promedio de 8 y un 75% de los promedios no supera el 9.
- ❖ No hay outliers, el rango de datos es de [5 - 10].

En este grado, un estudiante que tiene un promedio de 8 en Matemática, en el turno mañana está a 0.33 desvíos de la media. En el turno tarde, está más cerca, a un 0.21. Por último, en el turno completo se posiciona a 0.62 desvíos. Es decir que un promedio de 8 en Matemática para este grado, se ubica entre los mejores alumnos.

Nos resultó interesante fijarnos las correlaciones entre las materias más frecuentes en las primarias, pero a la hora de aplicar su correlación parcial, no dio lo mismo que en segundo y



tercer grado. Muchas correlaciones directas son positivas y altas, pero cuando nos fijamos si hay una tercera variable que interviene, nos indica que si, incluso que pueden ser más de una.



Los más destacable que podemos extraer son las correlaciones como la es entre Matemática ~ Lengua, siendo una correlación moderada, y como lo es Educación Física ~ Educación Tecnológica, siendo también una correlación moderada.

Si comparamos los promedios de Lengua, lo hacemos entre escuelas públicas y escuelas privadas. A través de un test no paramétrico podemos aceptar o rechazar la hipótesis nula. La salida de la prueba, nos da como resultado el rechazo de la hipótesis nula, o sea que la evidencia es suficiente para decir que los promedios de Lengua en cuarto grado son mayores en escuelas privadas. Esto tiene sentido si vemos las medianas para cada grupo. En escuelas de gestión privada, la mediana encontrada es igual a 8. Mientras que en escuelas públicas, la mediana es inferior, siendo un 7.67.

Ahora si comparamos los promedios para Matemáticas, pero según el turno al que asisten, con el test de ANOVA de más de dos medias independientes podemos encontrar diferencias entre los turnos. Pero esa diferencia no es tan significativa como en grados anteriores. Para este grado, el p-value devuelto por el ANOVA es de 0.00792.

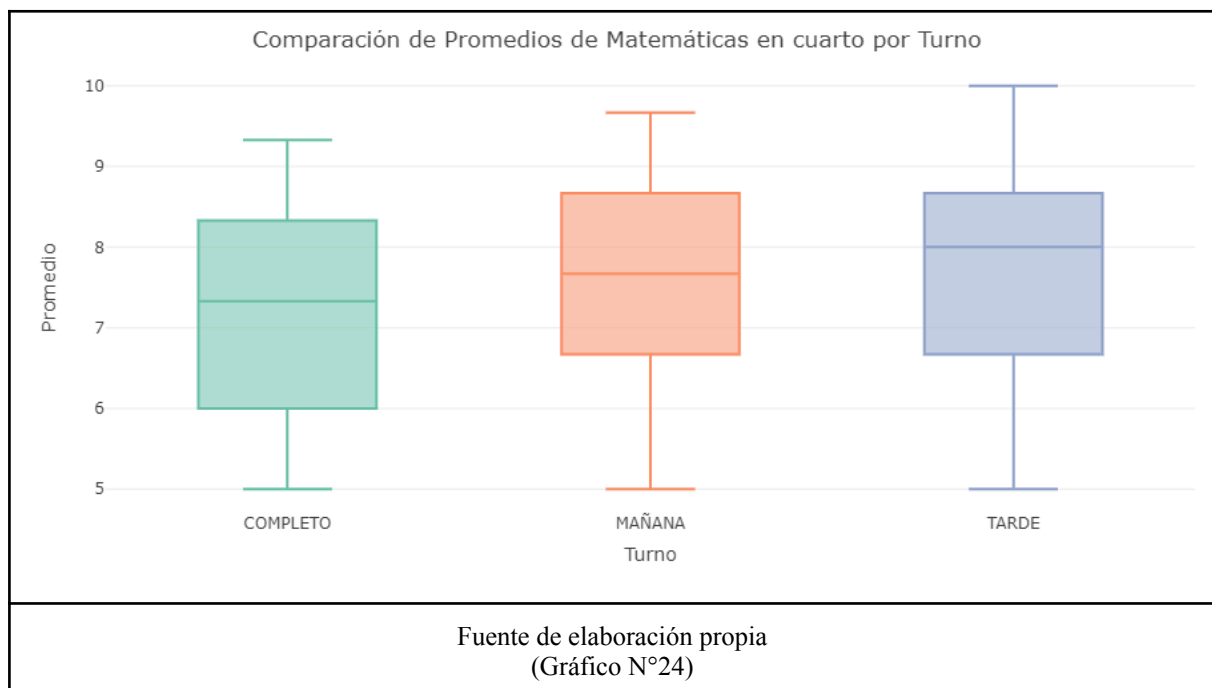
Las diferencias radican entre las comparaciones del turno completo con el tarde y mañana. Pero este análisis no es tan representativo, ya que las distribuciones de los conjuntos no son

normales, por lo cual derivamos el mismo análisis pero con un test no paramétrico, donde tendremos en cuenta el valor que está en el medio del conjunto ordenado, o sea la mediana.

En este caso, el Kruskal test nos dió un valor de p menor al 0.05. Entonces decimos que las medianas de los grupos difieren. El caso es que con el test de Pairwise, encontramos diferencias entre:

Turno	Medianas	p-value
MAÑANA - COMPLETO	7.67 - 7.33	0.0442
TARDE - COMPLETO	8 - 7.33	0.0093
TARDE - MAÑANA	8 - 7.67	0.2058
Fuente de elaboración propia (Tabla N°10)		

Si vemos el siguiente gráfico, podemos notar estas diferencias entre el turno el completo y sus pares:



En los promedios de Lengua sucede exactamente lo mismo, excepto en las diferencias entre el turno completo y tarde (p-value = 0.019), y mañana (p-value = 0.037).

Como se viene haciendo en los cursos anteriores, nos interesa saber cómo están distribuidos los desempeños en las materias, para saber si nuestra hipótesis sigue siendo válida en este grado. En este caso nos interesa saber los porcentajes conseguidos por columna.

	Regular	Aprobado	Bueno	Muy Bueno	Distinguido	Total
Artes Visuales	1.5	3.0	8.2	15.6	15.4	12.2
Cs. Naturales	12.6	16.9	1.2	12.2	13.5	13.2
Cs. Sociales	8.6	16.6	13.2	10.8	11.9	12.0
Ed. Física	1.0	2.7	8.2	16.2	16.0	12.5
Ed. Musical	4.5	9.9	11.1	9.8	17.3	12.1
Ed. Tecnológica	2.5	8.1	15.5	14.1	9.0	11.7
Lengua	34.3	19.9	15.8	10.6	7.9	13.0
Matemática	34.8	22.9	14.8	10.6	8.9	13.4
Total	100	100	100	100	100	100
Fuente de elaboración propia (Tabla N°11)						

Como vemos, entre Lengua y Matemática se llevan el 69.1% de los desempeños regulares. En los desempeños distinguidos, las mismas materias ocupan solo el 16.8%. Por lo que podemos seguir planteando nuestra hipótesis de que hay una asociación entre las variables analizadas. El test de chi cuadrado nos arroja un  $p\text{-value} < 2.2e-16$ , esto quiere decir que tenemos la evidencia suficiente para rechazar la hipótesis nula (no hay asociación) a favor de  $H_1$  (nuestra hipótesis). Esta asociación tiene una fuerza débil.

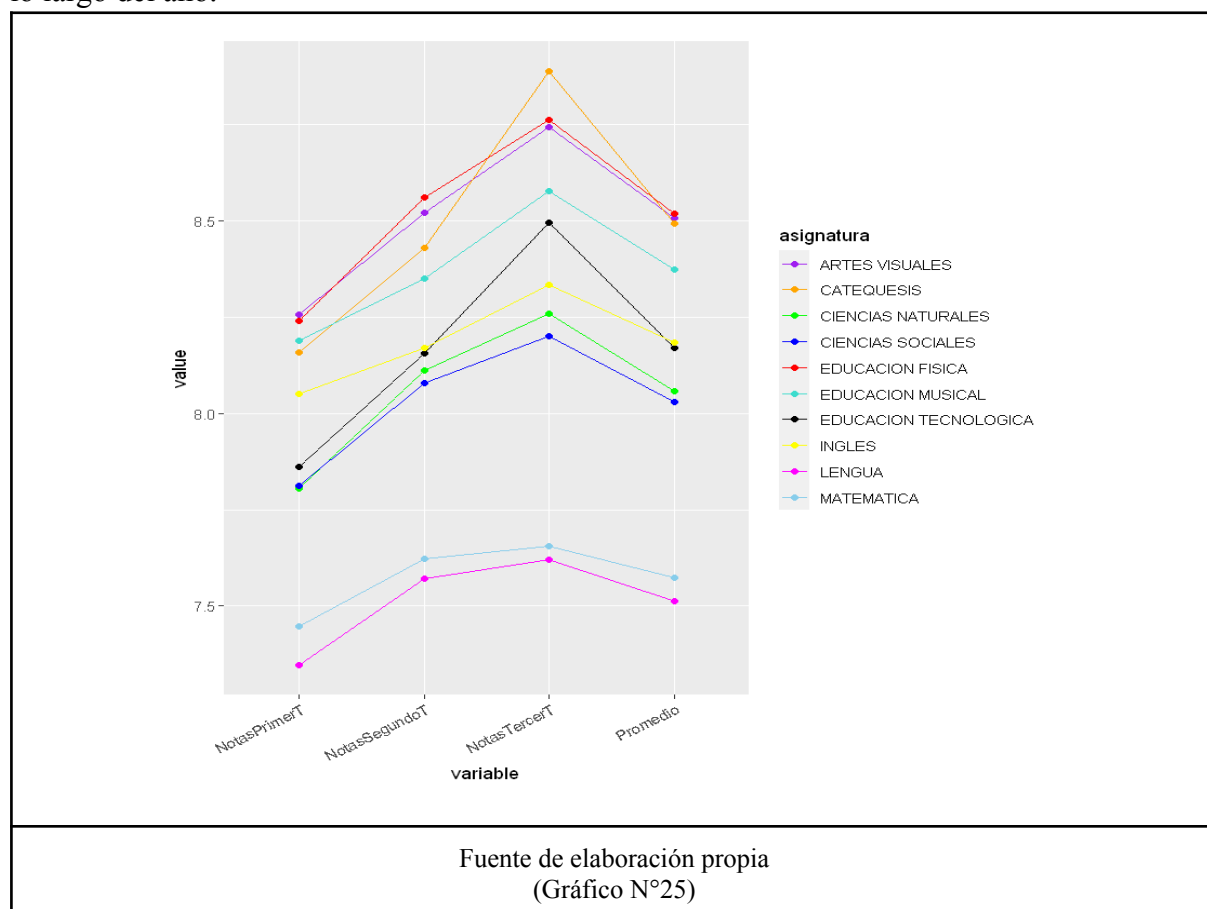
Si las variables fueran independientes, se esperaban más observaciones en la asociación entre:

- Lengua ~ Distinguido: Los residuos ajustados indican que se esperaban -6.0.

En cambio, hay una sobrefrecuencia en la asociación entre:

- Lengua ~ Regular: El residuo ajustado es de 9.2.

Al realizar un vector de medias, podemos calcular las medias de cada variable cualitativa. Gracias a esto, podemos representar el rendimiento obtenido en las materias de cuarto grado a lo largo del año.



Las áreas destacadas en este nivel de enseñanza primaria son Artes Visuales, Catequesis y Educación Física. Tienen la media de los promedios muy similares.

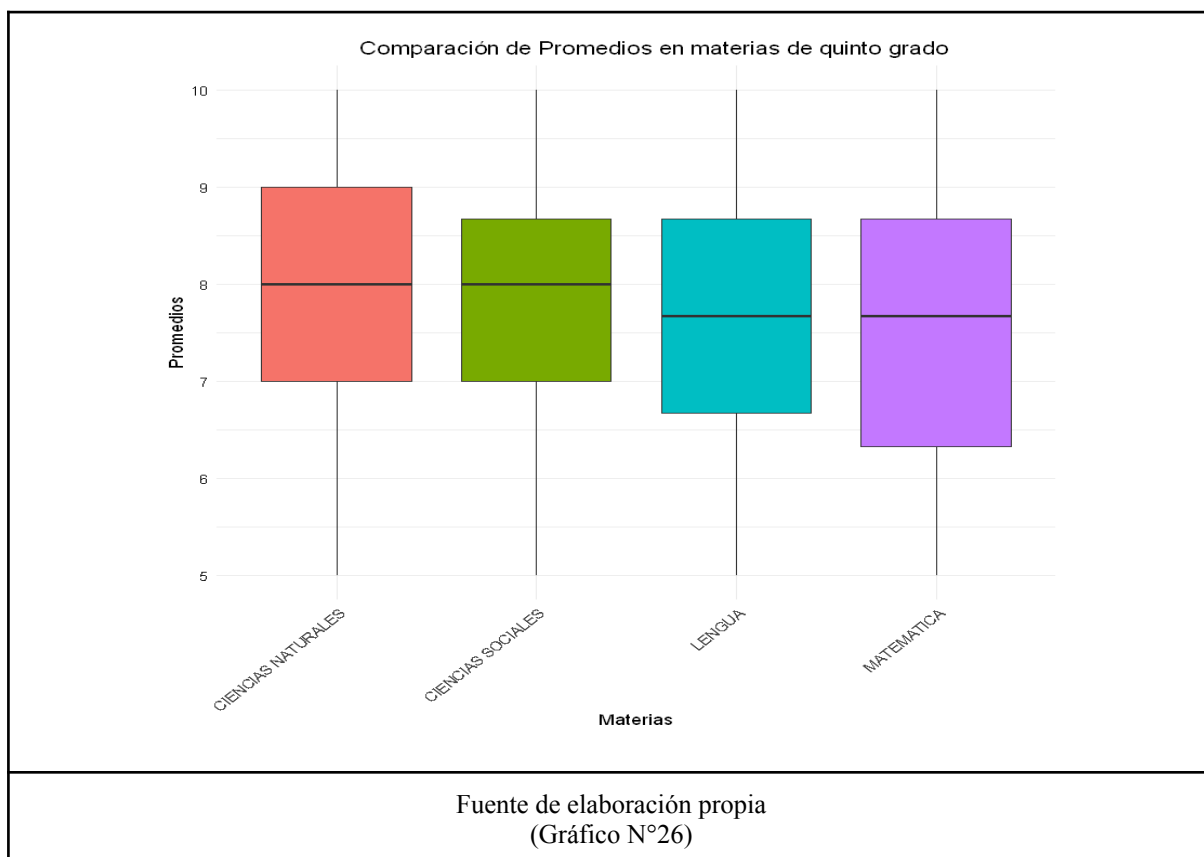
Las líneas que están en la parte inferior del gráfico pertenecen a Matemática y Lengua. Podemos notar cómo están alejosamente separadas de las medias de otras materias.

Por otro lado, queríamos ver si en este grado hay diferencias o no entre los vectores medios. Para hacer posible esto, asumimos que las medias de las variables se aproximan a una distribución normal multivariada, debido al T.L.C.M.

El MANOVA nos devuelve un p-value menor a  $2.2e-16$  por lo cual rechazamos nuestra hipótesis nula a favor de la alternativa y decimos que los vectores medios tienen una diferencia significativa entre los turnos de cuarto grado, tomando las variables cuantitativas como lo son las notas en cada trimestre y el promedio.

## Quinto grado

Con 503 matriculados, podemos describir la variable 'Promedio'. Se eligieron como se viene haciendo en los grados anteriores, las materias troncales.



Del gráfico N°17 podemos decir que:

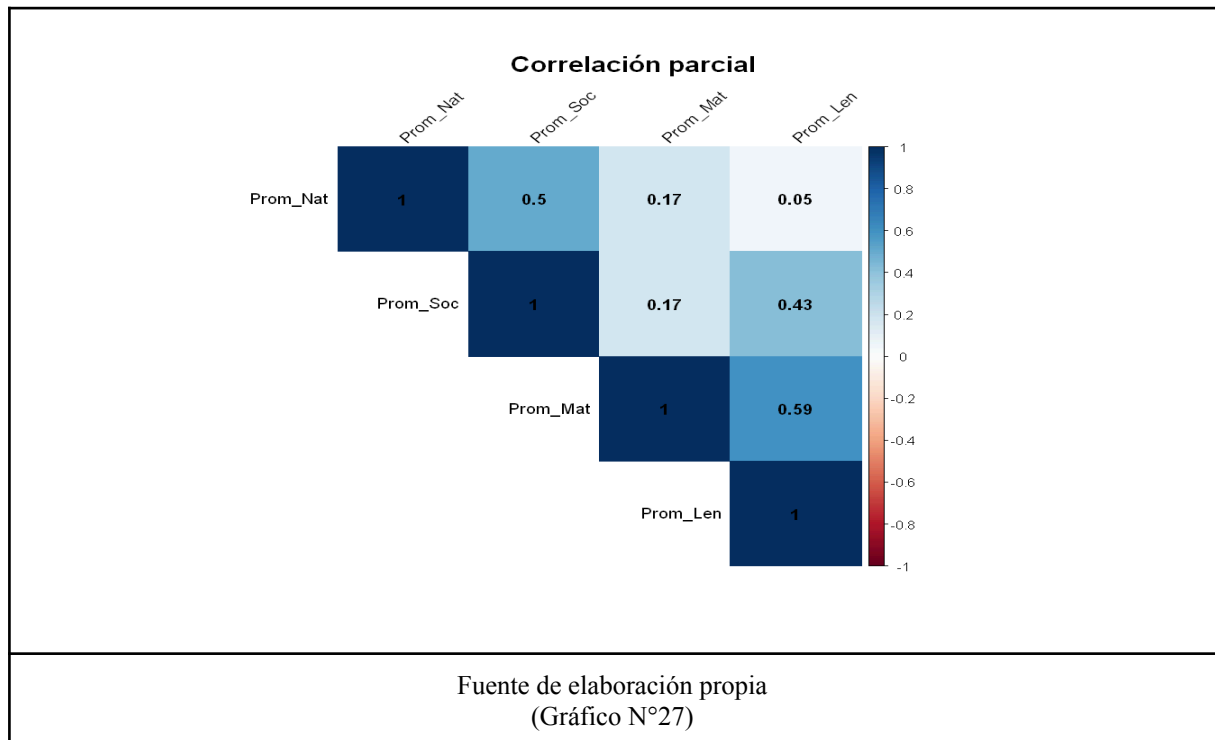
- ❖ La que consiguió peores promedios es Matemática, ya que el 25% de los promedios son menores a 6.33.
- ❖ Lo contrario a Ciencias Naturales que sigue siendo la que mejores promedios consigue, tienen un 75% de los promedios por encima de 7.
- ❖ Lengua tiene un 14% de los promedios desaprobados.
- ❖ Además, se demuestra que ninguna materia presenta outliers, su rango es el mismo [5 - 10] en todas las materias seleccionadas.

En quinto grado, el estudiante que tenga un promedio de 8 en Lengua, estará de la media a:

- 0.23 desviaciones de la media del turno mañana.
- 0.23 desvíos de la media del turno tarde.
- 0.77 desviaciones de la media del turno completo.

Por lo cual el estudiante, con ese promedio se posiciona como uno de los mejores en el turno completo, pero en el turno mañana como el tarde, estaría ubicado muy levemente por encima de la media.

Buscando si una materia de quinto grado se correlaciona con otra, aplicamos la correlación de pearson, ya que no siguen una distribución normal. Los resultados obtenidos reflejan altas correlaciones entre los promedios de las materias troncales. Sin embargo, las correlaciones parciales no son tan altas como se mostraban en primera instancia.



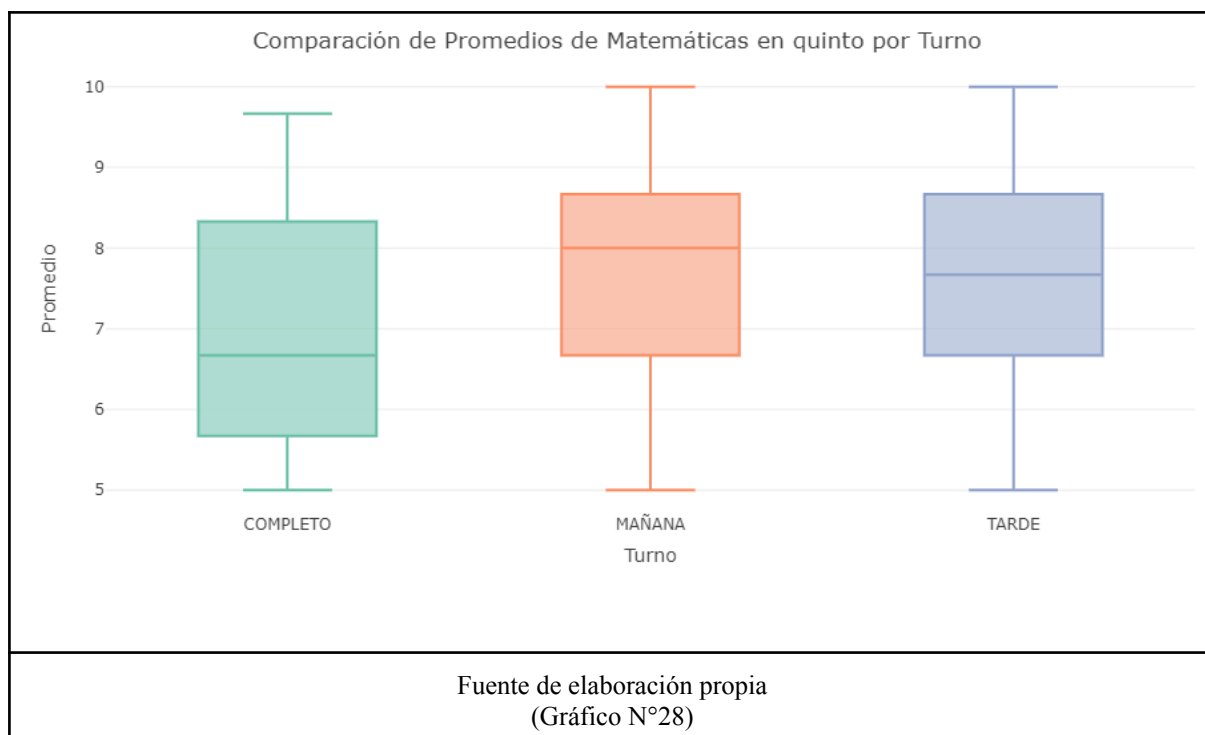
Para los promedios de Lengua de quinto grado, se le aplicó el test no paramétrico de Wilcoxon, ya que se compararon los promedios en escuelas públicas contras los promedios de escuelas privadas, y como no tenían una distribución normal, la comparación se realizó a través de la mediana. Nuestras hipótesis eran:

- H0: Los promedios de los alumnos de quinto grado en Lengua son iguales en escuelas privadas y públicas.
- H1: Los promedios de los alumnos de quinto grado en Lengua en escuelas privadas son mayores que en escuelas públicas.

Con un nivel de significancia de 0.05, la evidencia es suficiente para rechazar H0 en favor de H1 y decir que los promedios en Lengua en escuelas privadas son mejores que los promedios de escuelas públicas.

Con los promedios de Matemática hicimos un análisis con el test de más de dos medias independientes, donde los conjuntos de datos estaban integrados por los diferentes turnos.

La siguiente representación gráfica demuestra la distribución de los promedios en los diferentes turnos:



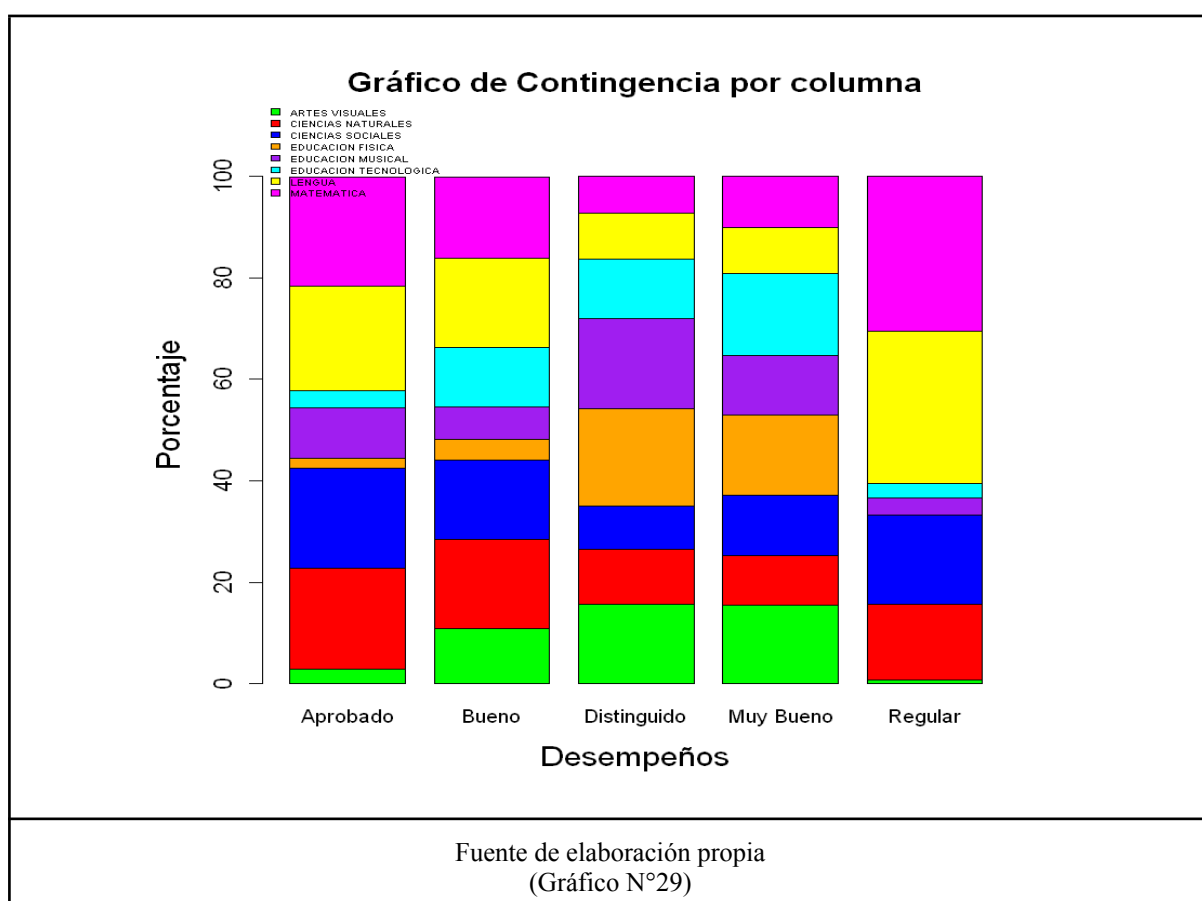
Este gráfico es importante de interpretar ya que al no tener distribuciones normales, la prueba estadística se realiza en torno a la mediana. Ya que los resultados del ANOVA no serían representativos. Por eso, el test de Kruskal devuelve que si hay diferencias en los promedios de Matemática entre los diferentes turnos. Específicamente entre el turno completo con el de la mañana y turno completo con el de la tarde.

En el análisis bivariado cualitativo, nosotros creemos que los desempeños están determinados según la materia. ¿Por qué? Por los resultados conseguidos en las tablas de contingencia y las distintas representaciones gráficas que se mostraran a continuación.

	Regular	Aprobado	Bueno	Muy Bueno	Distinguido	Total
Artes Visuales	0.4	2.2	15.2	42.2	40.0	100
Cs. Naturales	7.3	14.9	24.3	26.3	27.2	100
Cs. Sociales	8.8	14.9	21.8	32.4	22.2	100
Ed. Física	0.0	1.6	5.8	43.3	49.4	100
Ed. Musical	1.9	7.8	9.3	33.3	47.7	100

Ed. Tecnológica	1.5	2.7	17.3	46.6	31.9	100
Lengua	14.6	15.2	23.8	24.0	22.4	100
Matemática	15.3	16.3	22.4	27.3	18.8	100
Total	6.3	9.6	17.6	34.2	32.3	100
Fuente de elaboración propia (Tabla N°12)						

Por ejemplo en esta tabla de contingencia de los porcentajes por fila, podemos ver como en todos las asignaturas existe una diferencia considerable entre los desempeños bajos y altos, siendo los mejores desempeños los que ocupan mayor porcentaje. Excepto en materias como Lengua y Matemática que tienen una distribución casi equitativa entre los desempeños bajos y altos.



En este gráfico vemos como el 60.6% de los desempeños regulares lo concentran las materias más importantes.

Con lo presentado podemos concluir que existe una asociación entre las variables. Para verificar esto, lo haremos con el test de chi cuadrado, con una significancia del 5%.



El p-value devuelto es menor a  $2.2e-16$ . Por lo cual nuestra hipótesis es correcta, existe una asociación entre las variables de una fuerza de 0.209 según el coeficiente de Cramer.

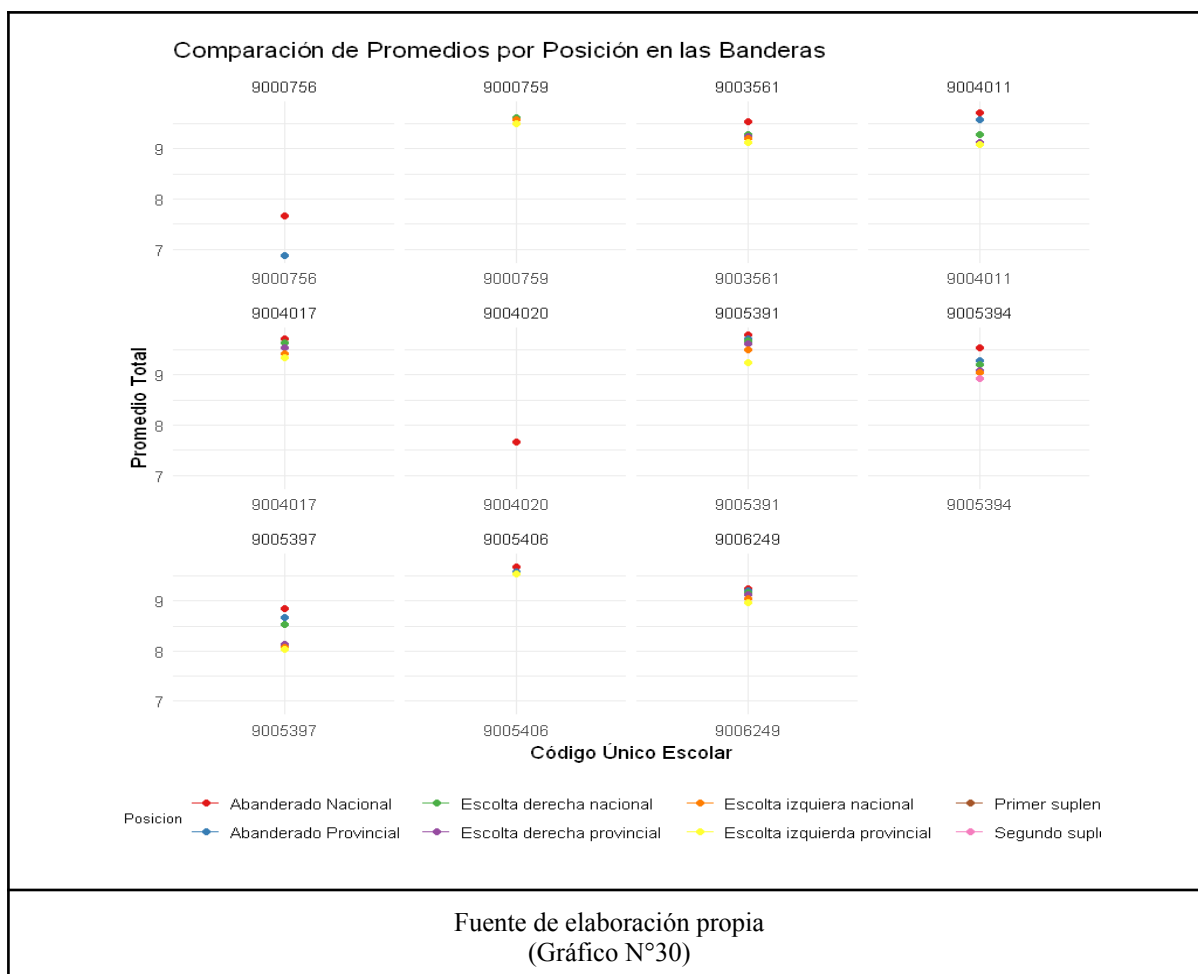
Para finalizar con quinto grado, como es un curso que al finalizar el año lectivo, los mejores alumnos recibirán las banderas de ceremonia. Nos propusimos buscar los abanderados entrantes para cada escuela de nuestro dataset.

Para hacer posible esto, tomamos a los estudiantes que tengan aprobadas las 8 materias que denominados comunes en primaria. Entonces, calcularemos el promedio total de cada estudiante según los promedios obtenidos en cada materia. Estos filtros se hicieron para que la competencia sea pareja entre los estudiantes, ya que si tienen menos o más materias, afecta directamente en el promedio total, dando posiciones en las banderas incorrectas.

Después de todos los pasos de filtrado, cálculo, selección, agrupación y asignación, se armó el data frame con los resultados, algunos de ellos fueron:

CodigoUnicoEscolar	idalumno	Documento	Gestión	turno	Promedio _Total	Posición
9003561	636680	52491513	Pública	MAÑANA	9.54	Abanderado Nacional
9003561	547731	52946674	Pública	MAÑANA	9.29	Abanderado Provincial
9003561	592438	53005877	Pública	MAÑANA	9.29	Escolta derecha nacional
9003561	668028	53250955	Pública	TARDE	9.25	Escolta derecha provincial
9003561	551293	52820596	Pública	TARDE	9.21	Escolta izquierda nacional
9003561	570943	52491522	Pública	TARDE	9.12	Escolta izquierda provincial

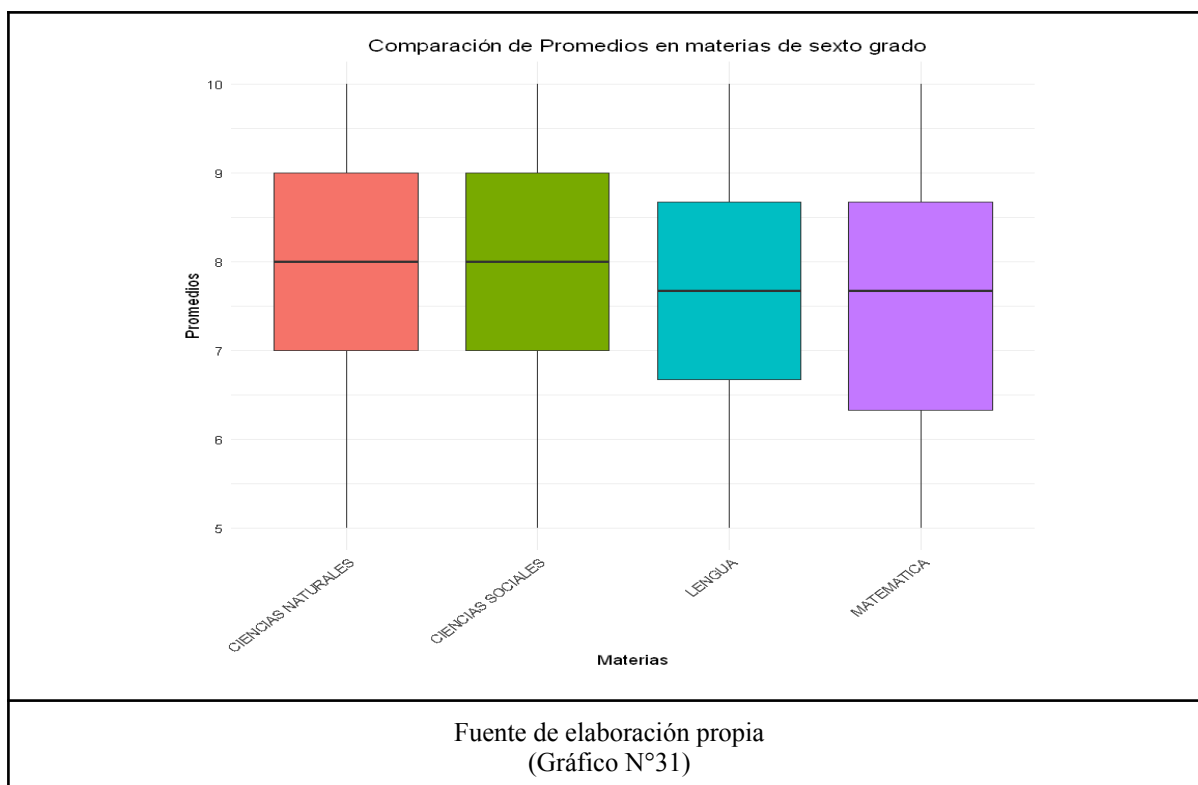
En el siguiente gráfico se muestran los abanderados para cada escuela de una manera más práctica:



Si analizamos el gráfico podemos ver como la escuela N°9005406 de gestión pública tiene los mejores promedios totales. El Abanderado Nacional entrante consiguió un promedio de 9.67, mientras que el sexto mejor promedio de esa escuela es de 9.54, muy buenos logros para la escuela y los estudiantes.

### Sexto grado

Teniendo 506 matriculados en nuestra base de datos, alguna de las medidas estadísticas que podemos obtener de la variable 'Promedio' en algunas materias comunes de la primaria las podemos ver resumidamente en el siguiente boxplot:



A simple vista podemos decir:

- ❖ Los peores promedios obtenidos de las cuatro materias comparadas es Matemática, ya que el 75% de sus promedios están entre 6.33 y 8.67.
- ❖ Lengua tiene el 88% de los promedios por encima de 6, pero solo el 41% supera un promedio de 8.
- ❖ Mientras que las mejores son las ciencias, tanto Naturales como Sociales. Pero si los analizamos con los percentiles, el 93% de los promedios en Sociales fueron mayores a 6. En cambio, en Naturales el 94% tienen un promedio mayor a 6.
- ❖ También es importante destacar que no presentan outliers, el rango de los promedios en todas las asignaturas es de [5 - 10].

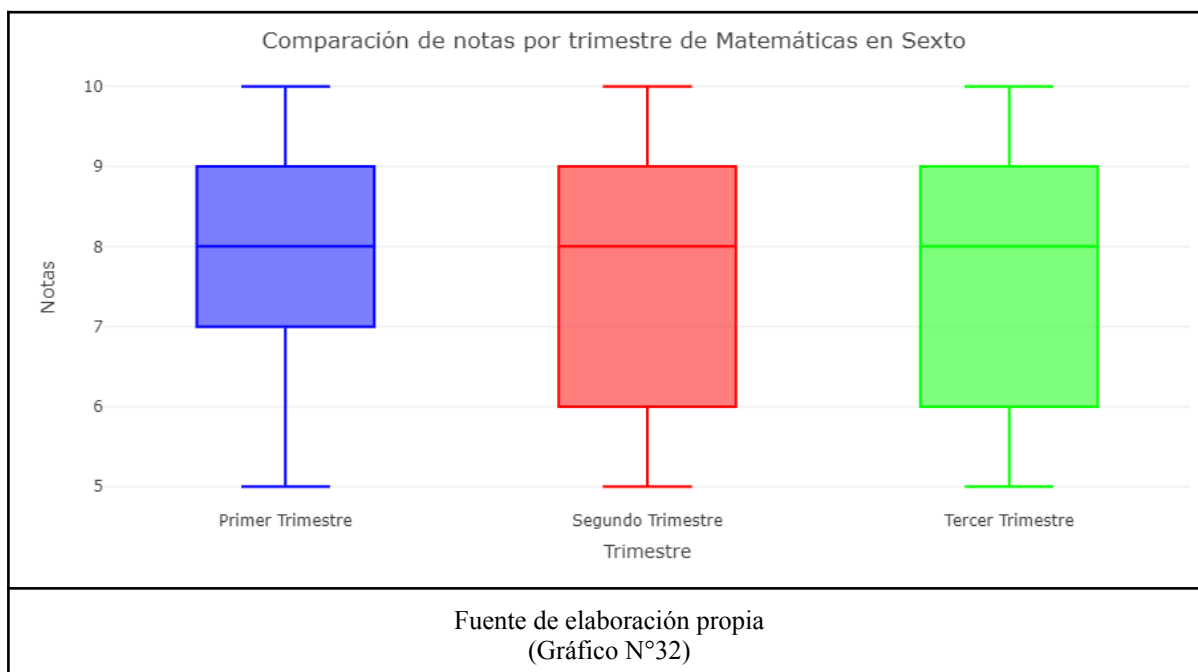
Por otro lado, para confirmar el tipo de distribución que siguen, se les calcularon las medidas de distribución y test de normalidad, los resultados son interesantes porque no todos arrojaron de igual manera que los grados ya analizados.

Materia	Coficiente de simetría	Curtosis	Media	Mediana	Desvío	¿Es normal la distribución?
Matemática	-0.03	1.99	7.55	7.67	1.40	No
Lengua	0.02	2.11	7.52	7.67	1.33	No
Cs. Naturales	-0.24	2.13	7.96	8.00	1.31	No

Cs. Sociales	-0.18	2.12	7.93	8.00	1.34	No
--------------	-------	------	------	------	------	----

En este caso, Matemática, Cs. Sociales y Cs. Naturales están levemente sesgadas hacia la izquierda, mientras que Lengua tiene una leve asimetría hacia la derecha, a pesar de ser un número tan bajo y no tan significativo, es importante destacar por lo que significa que la mayoría de datos esté por debajo de la media.

Nos resultó interesante investigar las notas de cada trimestre en Matemática. Para eso empleamos el uso del test para más de dos medias pareadas. El gráfico de las distribuciones quedó muy interesante:



Parece ser que no hay tantas diferencias entre las notas de los trimestres. Las salidas del ANOVA efectivamente confirmaron nuestra visión, el p-value devuelto es 0.9494. Por lo cual no podemos rechazar  $H_0$ , lo que significa que las medias de nuestros conjuntos de datos son iguales. Pero no es resultado de fiar ya que no cumplen con los supuestos.

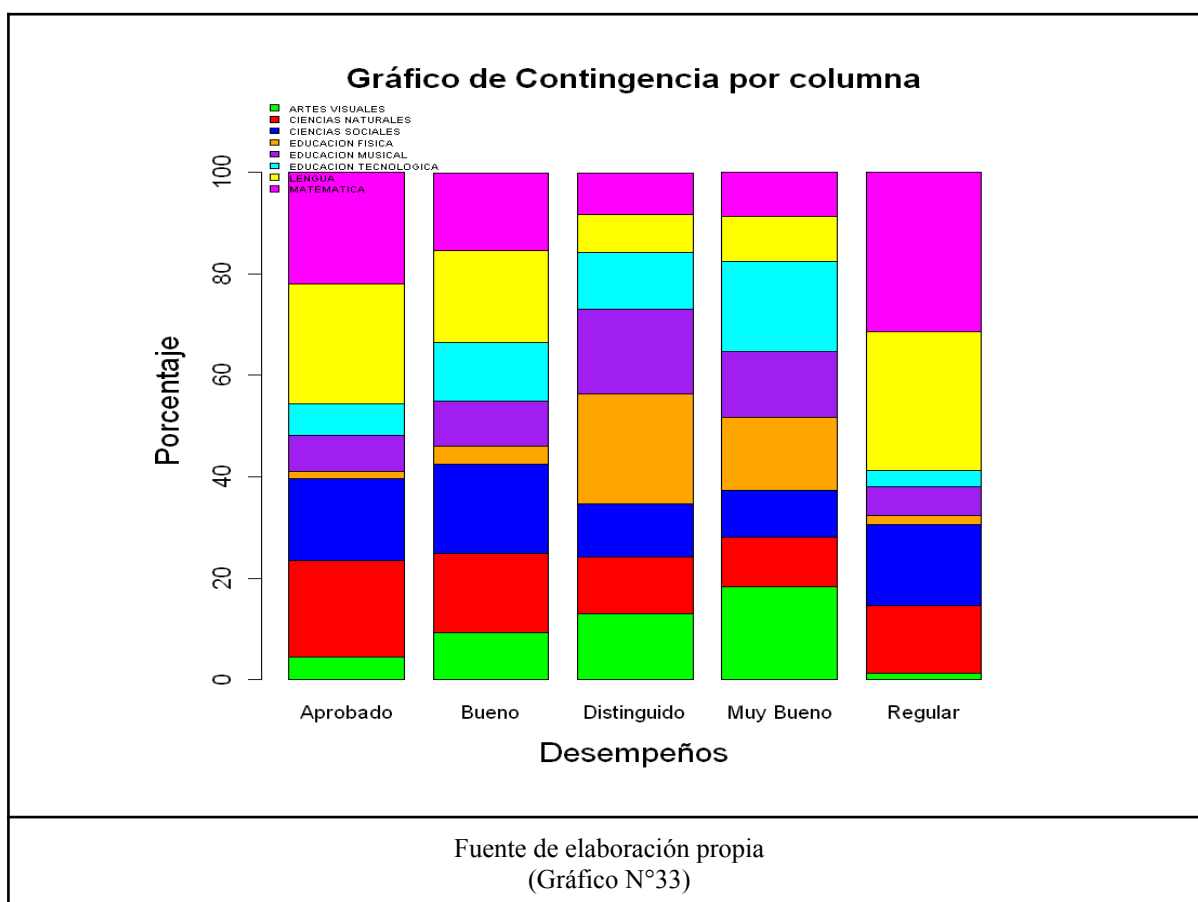
Por eso basaremos el mismo análisis pero con el test de Friedman, donde compararemos los grupos con la mediana. La salida fue de un p-value de 0.2122, por lo que seguimos tomando a  $H_0$  como la hipótesis.

Si vemos las salidas del test de Pairwise, vemos que no hay diferencias entre los trimestres.

	Primer Trimestre	Segundo Trimestre
Segundo Trimestre	0.31	-
Tercer Trimestre	0.60	0.37

Fuente de elaboración propia  
(Tabla N°13)

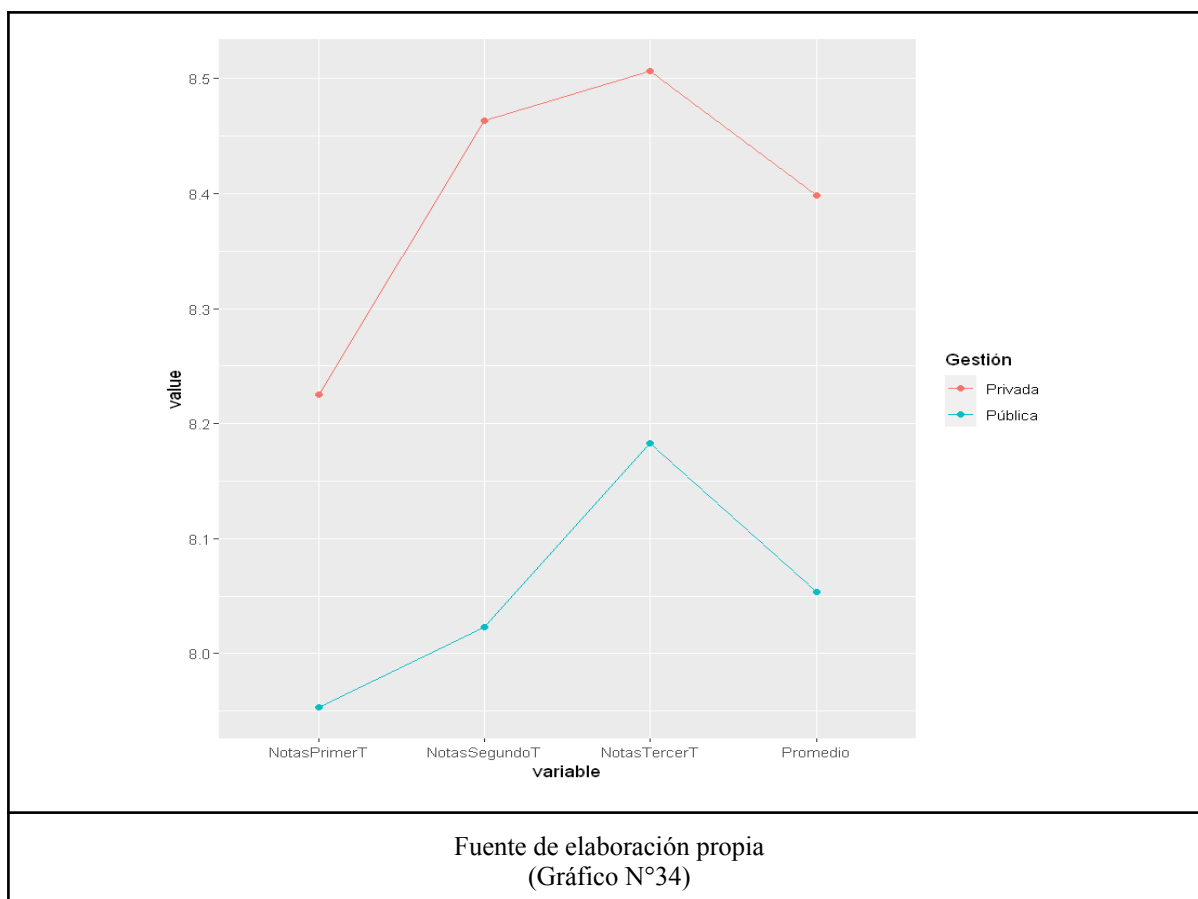
En las tablas de contingencia seguimos encontrando una asociación entre los desempeños y las materias. En este grado, lo que nos llamó la atención es que en la tabla de contingencia por columna aumentaron los porcentajes de desempeños regulares en Ciencias Sociales y Naturales, mientras que Matemática y Lengua siguen teniendo un porcentaje alto de desempeños regulares. Estos resultados se reflejan en el siguiente gráfico:



Antes de pasar al siguiente análisis queremos decir que vamos a asumir que las medias de las variables se aproximan a una distribución normal multivariada si tomamos el teorema del límite central multivariado.

Aclarado lo anterior nos resultó interesante saber si los vectores medios entre los tipos de gestión son iguales o difieren según el tipo de gestión en este grado.

El siguiente gráfico resume totalmente la salida del test de Hotelling, donde nos dice que si hay diferencias entre los vectores medios del tipo de gestión.



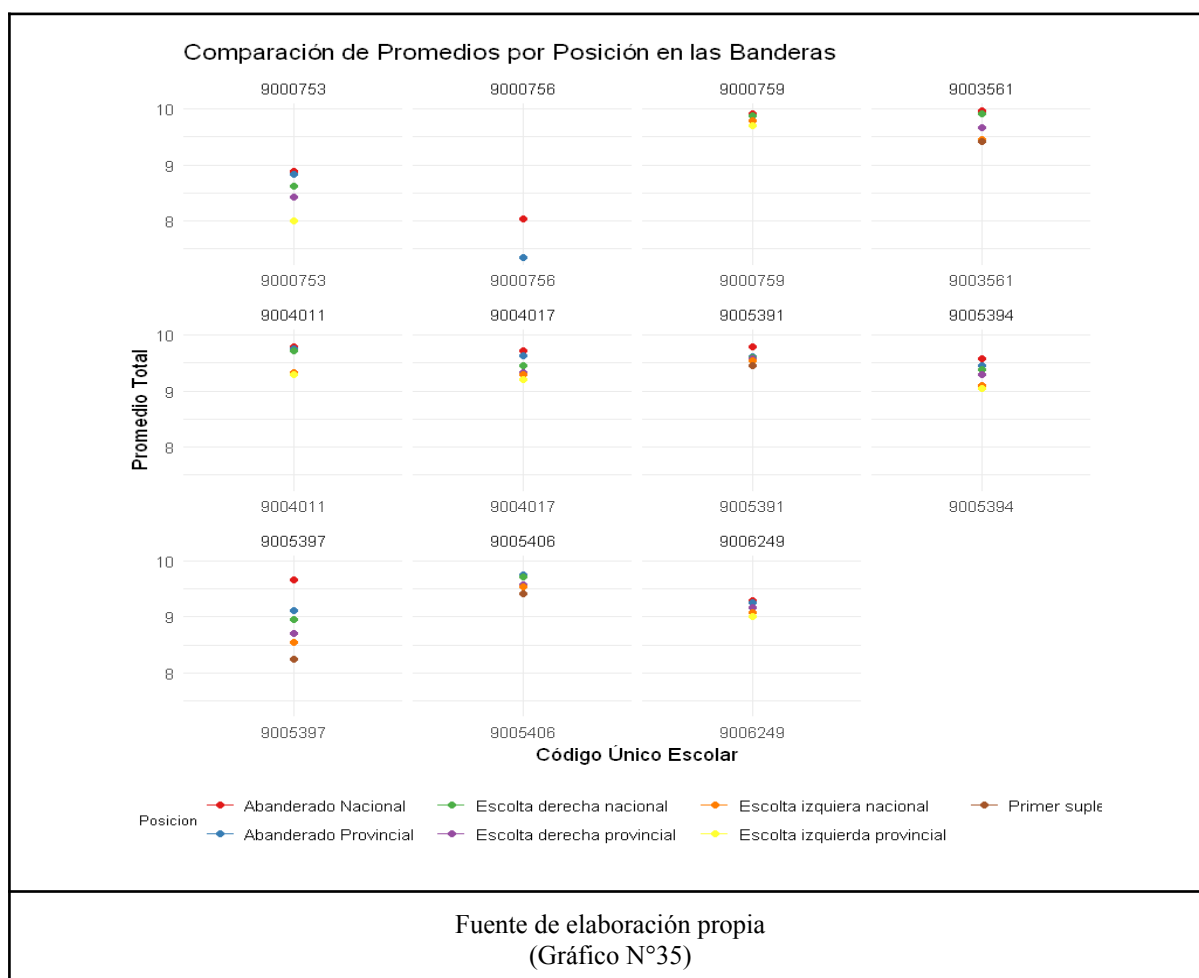
Para finalizar con sexto grado, decidimos “premiar” a los mejores estudiantes de este grado. Para eso lo que hicimos fue muy similar a lo realizado para los estudiantes de quinto grado, pero esta vez destacaremos a los abanderados que a final de año entregan los símbolos patrios.

Algunos de los abanderados obtenidos son:

CodigoUnicoEscolar	idalumno	Documento	Gestión	turno	Promedio_Total	Posición
9000759	549728	51116503	Pública	MAÑANA	9.92	Abanderado Nacional
9000759	454880	51020174	Pública	MAÑANA	9.88	Abanderado Provincial
9000759	564663	52143634	Pública	MAÑANA	9.88	Escolta derecha nacional

9000759	549725	51283976	Pública	MAÑANA	9.79	Escolta derecha provincial
9000759	937437	51503135	Pública	TARDE	9.79	Escolta izquierda nacional
9000759	564674	51284070	Pública	MAÑANA	9.71	Escolta izquierda provincia
Fuente de elaboración propia (Tabla N°14)						

En el siguiente gráfico, se mostrarán los promedios totales conseguidos por los estudiantes y su posición, separados por escuela:



De este gráfico también podemos sacar información como que la escuela mostrada en la tabla N° X tiene los mejores promedios totales, siendo una escuela pública. Una escuela privada como lo es la N°9005391, tiene los abanderados levemente por debajo que la escuela N° 9000759.

---

## Secundaria

Ahora nos vamos a centrar en el análisis de los datos obtenidos para los alumnos de **colegios secundarios** del departamento Victoria. En esta sección, abordaremos diversos aspectos del rendimiento académico y la distribución de los estudiantes, utilizando diferentes métodos de análisis. Para esta segunda parte del informe, planteamos una estructura temática que permitirá una visión integral y detallada de los datos, facilitando comparaciones transversales. Los temas que se cubrirán incluyen un análisis del rendimiento académico en diferentes materias y ciclos educativos, el impacto del turno escolar, la influencia del tipo de gestión (público/privado) en la condición final de los alumnos, y comparaciones a nivel del desempeño. Cada subsección ofrecerá un análisis exhaustivo y gráficos ilustrativos para apoyar nuestras observaciones y conclusiones.

Lo primero que presentamos es una tabla que resume la diversidad de instituciones educativas secundarias en el departamento Victoria, ofreciendo una visión general de su composición y distribución. Aquí se detallan aspectos clave:

Tipo de escuela	Cantidad
Escuelas secundarias totales	20
Escuelas secundarias públicas	17
Escuelas secundarias privadas	3
Escuelas secundarias que tienen turno mañana	16
Escuelas secundarias que tienen turno tarde	11
Escuelas secundarias que tienen turno completo	2
Escuelas secundarias que tienen turno rotativo	1
Escuelas secundarias comunes	20
Escuelas secundarias Técnicas	5
Escuelas secundarias Agrotécnicas	1
Fuente de elaboración propia (Tabla N°15)	

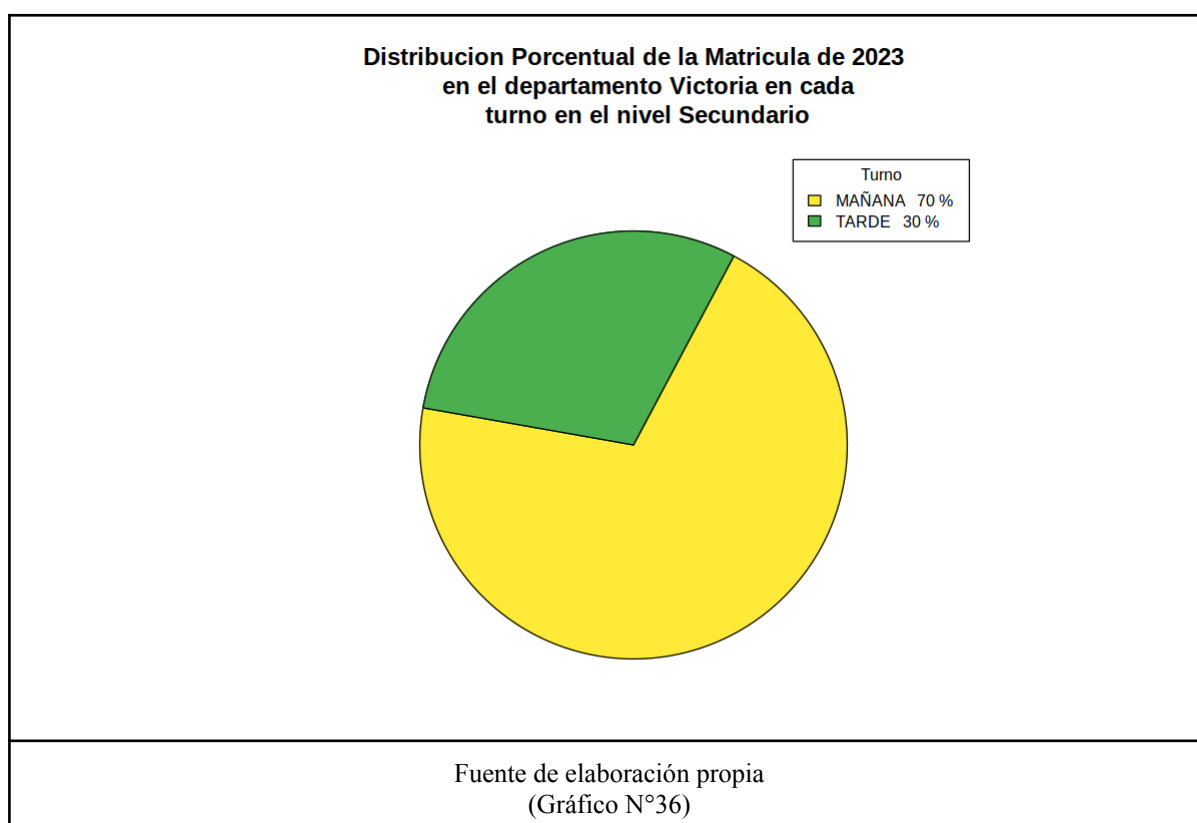


Continuando, a modo de introducción a esta segunda parte, vamos a plantearnos algunas preguntas generales de esta sección:

- ¿En qué turno cursa la mayoría de los estudiantes?
- ¿Cómo se distribuye la modalidad de enseñanza?
- ¿La cantidad de aprobados finales es alto, bajo, está bien?
- ¿Cómo se distribuyen los colegios en públicos o privados?
- ¿Cómo es el desempeño de los estudiantes?
- ¿Cuál es la cantidad de matriculados totales? ¿Y en cada año? ¿Y en cada ciclo?

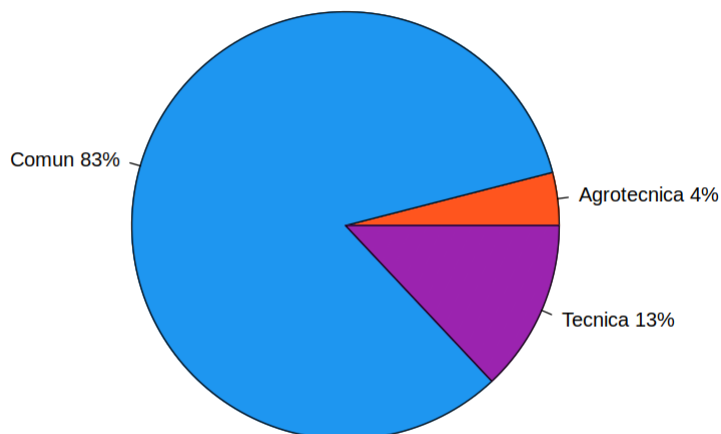
Con la ayuda de diversos gráficos iremos respondiendo una por una estas preguntas, para tener un mejor panorama del nivel secundario.

Lo primero que podemos observar, es que el mayor porcentaje de matriculados cursan la secundaria en el turno **mañana**. Siguiendo a ella el turno tarde. Aclarar que para la confección de este gráfico, no tuvimos en cuenta los turnos noche y rotativo por corresponderse con colegios para adultos.



Luego podemos observar que la mayoría de los colegios secundarios (un 83%) se adecuan a un modo de enseñanza **común**. El resto se corresponde con una educación más especializada (siendo un 13% escuelas técnicas y un 4% escuelas agrotécnicas).

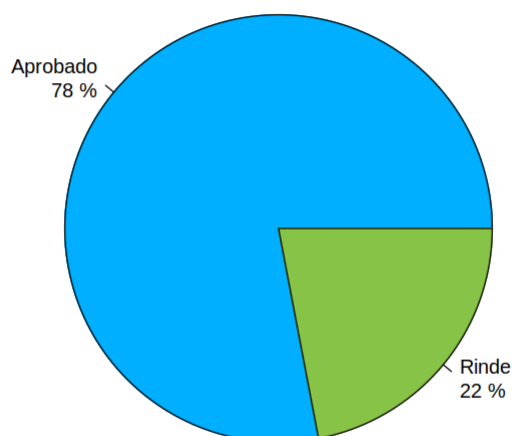
**Distribucion Porcentual de la Matricula de 2023  
en Entre Ríos por modo de enseñanza en nivel secundario**



Fuente de elaboración propia  
(Gráfico N°37)

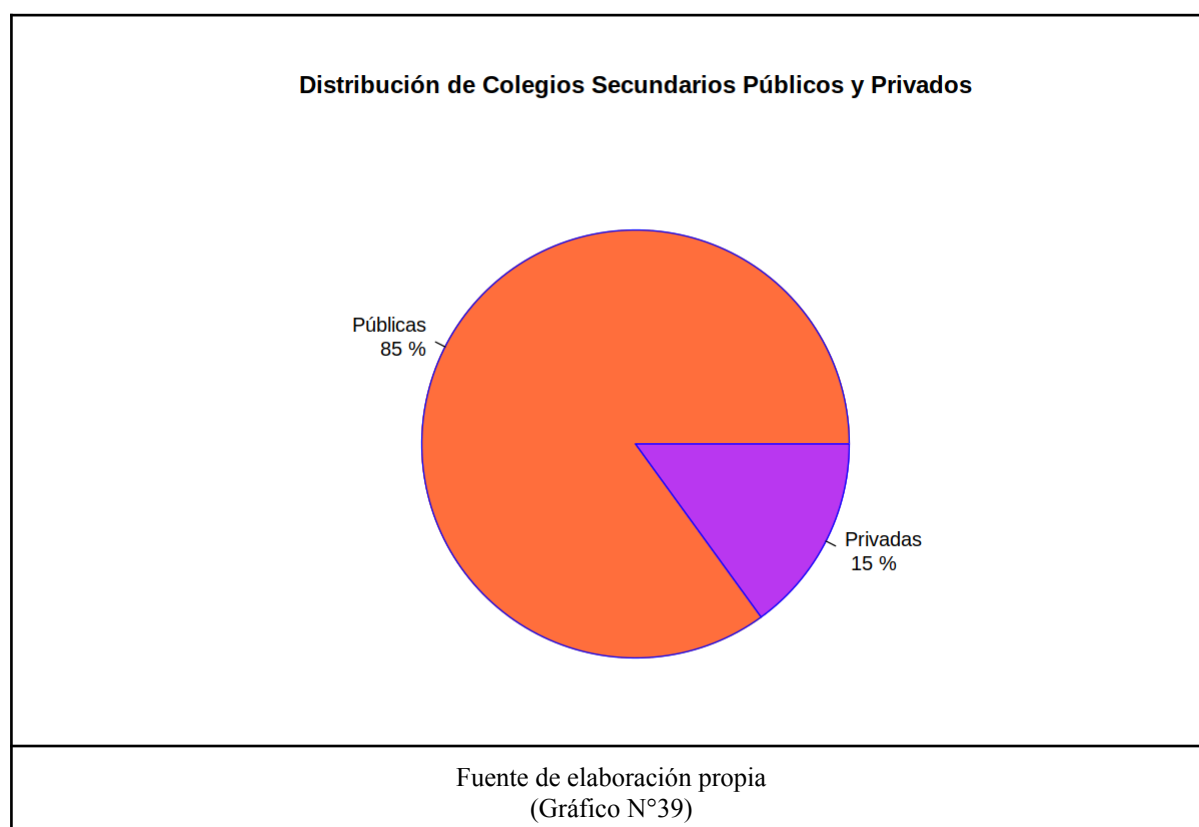
Siguiendo con la cantidad de aprobados totales, observar que en los colegios secundarios en el departamento Victoria, un 78% de los alumnos totales finalizan el curso en condición de **Aprobado**, mientras que el restante 22% tiene que rendir en las instancias de recuperación para poder aprobar la asignatura correspondiente.

**Distribucion Porcentual de la Matricula de 2023 en  
Entre Ríos por condición final en nivel secundario**



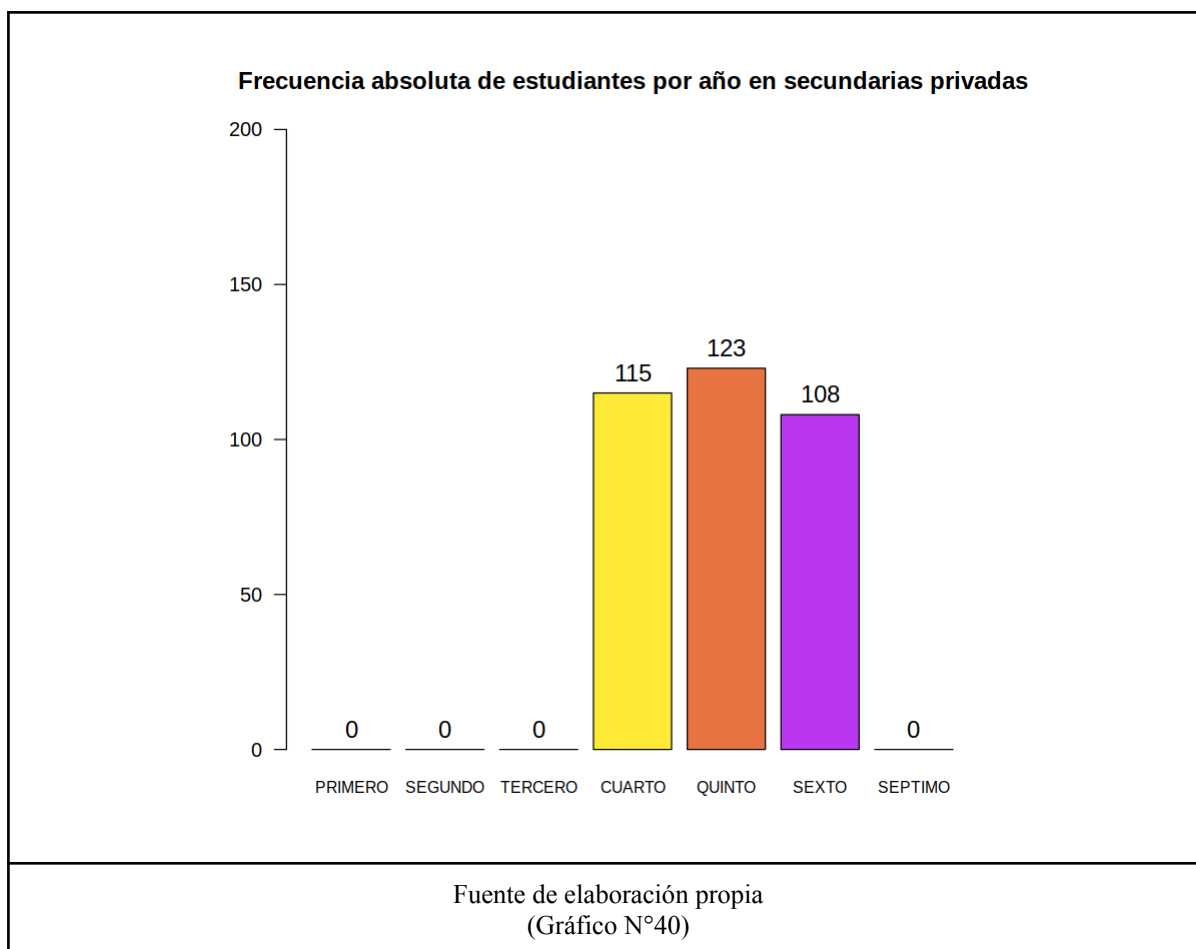
Fuente de elaboración propia  
(Gráfico N°38)

Si miramos en el gráfico N° X como es la gestión de los colegios secundarios en el departamento Victoria, notamos que el 85% de las secundarias son de gestión pública, mientras que solamente el 15% son de gestión privada. Claramente la distribución apunta a que la mayoría de los colegios son de **gestión pública**. Más precisamente, si volvemos a la tabla N°15, 17 son de gestión pública y 3 de gestión privada.



Al analizar más detenidamente la distribución de estudiantes por año en colegios secundarios privados, hemos observado una discrepancia notable: solo encontramos matrículas en cuarto, quinto y sexto año.

Este hallazgo nos ha llamado la atención y ha generado múltiples intentos de revisar nuestra etapa de limpieza de datos para asegurarnos de que no estamos perdiendo información relevante debido a filtros o procesos incorrectos. A pesar de nuestros esfuerzos, hasta el momento no hemos podido identificar una fuente clara del problema. Igualmente nos vemos en la necesidad de comentar estos datos, ya que esta discrepancia podría tener implicaciones importantes para nuestro análisis y conclusiones. En el gráfico N°X podemos ver esto que recién comentamos.

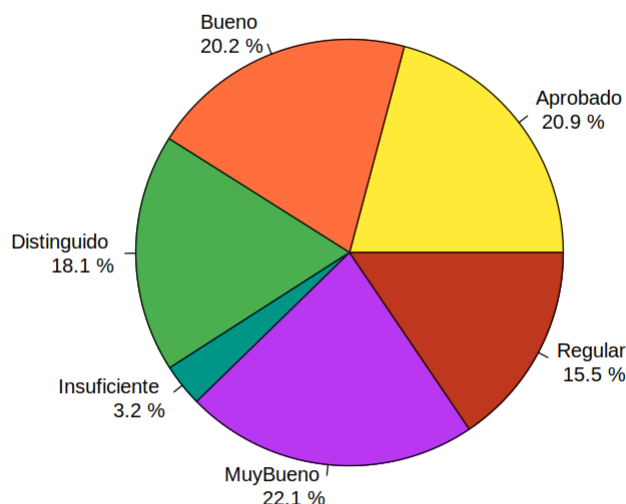


Para evaluar el rendimiento académico de los estudiantes en los colegios secundarios del Departamento Victoria, se creó la variable Desempeño a partir del promedio de calificaciones (Promedio). Esta nueva variable clasifica a los estudiantes en seis categorías distintas, basadas en sus promedios, según los siguientes criterios:

- Insuficiente: Promedio menor a 4.
- Regular: Promedio entre 4 y menos de 6.
- Aprobado: Promedio entre 6 y menos de 7.
- Bueno: Promedio entre 7 y menos de 8.
- Muy Bueno: Promedio entre 8 y menos de 9.
- Distinguido: Promedio entre 9 y 10.

Debajo, en el Gráfico N°X se puede ver cómo es la distribución del desempeño logrado por los estudiantes.

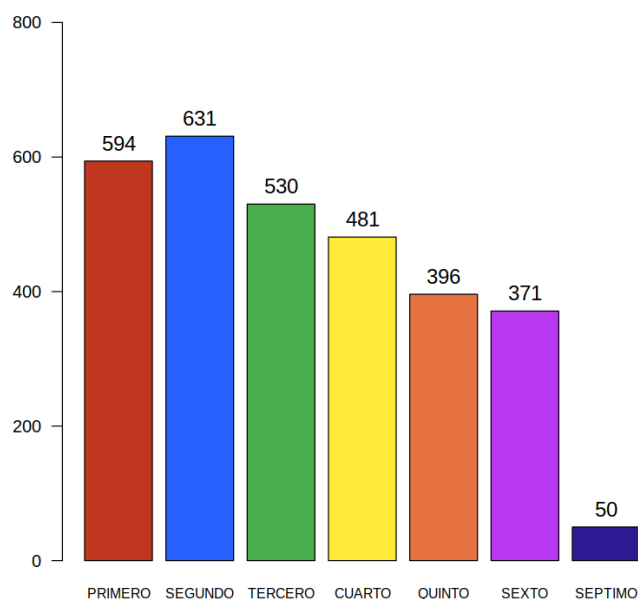
**Distribución Porcentual del Desempeño logrado por Estudiantes secundarios en el departameto Victoria, Entre Rios.**



Fuente de elaboración propia  
(Gráfico N°41)

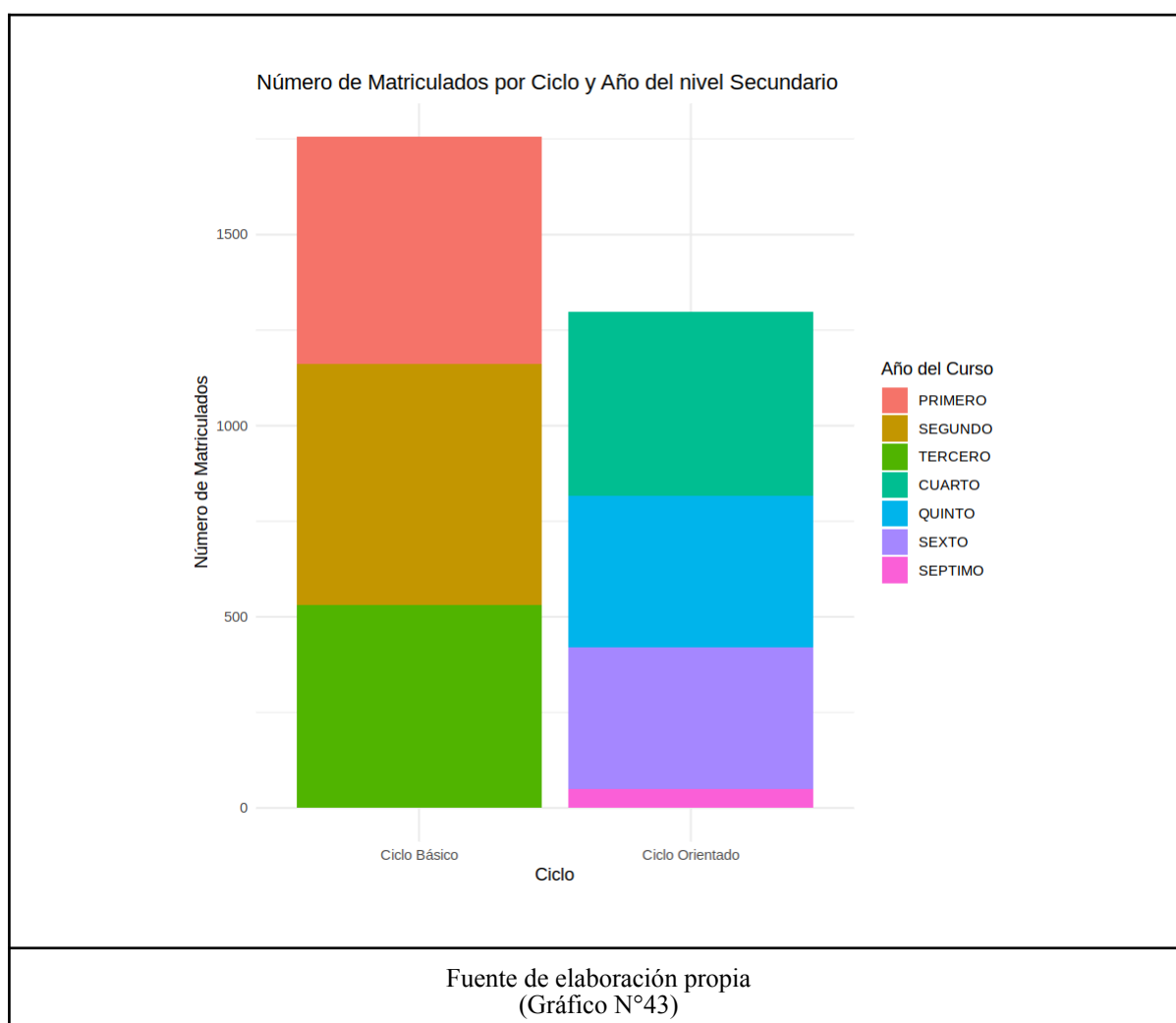
Ya finalizando esta primera parte, se muestra debajo la cantidad de matriculados por año en secundaria. Como se ve en el gráfico, la distribución por ciclo parece bastante uniforme. La diferencia se observa en el séptimo año, ya que la mayoría de los colegios secundarios cursan de primero a sexto año.

**Frecuencia absoluta de estudiantes por año en secundaria**



Fuente de elaboración propia  
(Gráfico N°42)

Por último, es interesante observar cómo se distribuyen las matrículas por ciclo, estando el Ciclo Básico compuesto por 1755 alumnos y el Ciclo Orientado/Especializado por 1298 alumnos. Abajo un gráfico de cantidad de matriculados por ciclo: Consideramos ciclo básico a los 3 primeros años en todos los colegios secundarios, ciclo orientado en secundarias comunes (4to a 6to) y ciclo especializado en colegios secundarios técnicos y agrotécnicos (4to a 7mo).



## Análisis Exploratorio del rendimiento en asignaturas troncales.

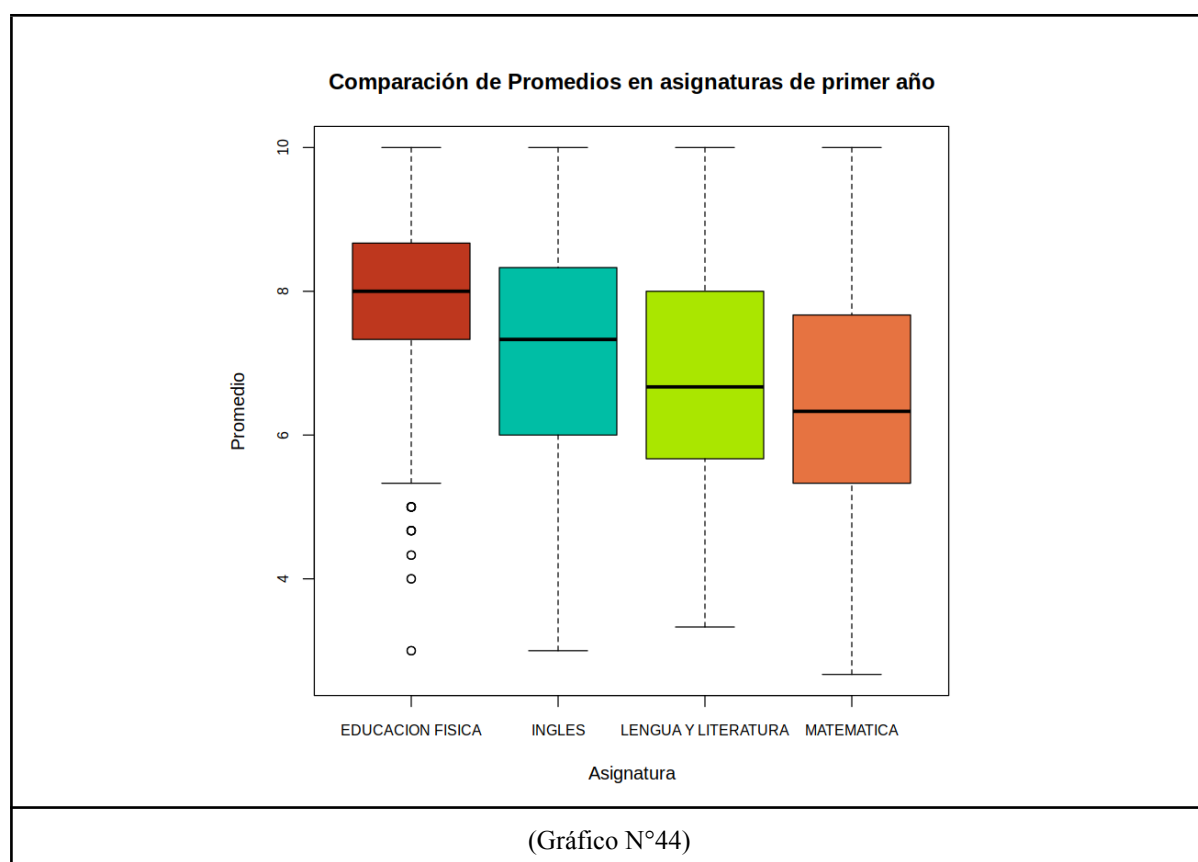
En esta primera parte, realizaremos un análisis exploratorio mediante boxplots comparativos de las asignaturas troncales: Matemática, Lengua y Literatura, Educación Física e Inglés. Estas materias son fundamentales en el plan de estudio y tienen continuidad en los años posteriores, lo que nos permite obtener una visión general del rendimiento académico de los

estudiantes en áreas clave. A través de estos gráficos, podremos comparar promedios e identificar tendencias y aspectos destacados en el rendimiento académico.

### **Primer Año**

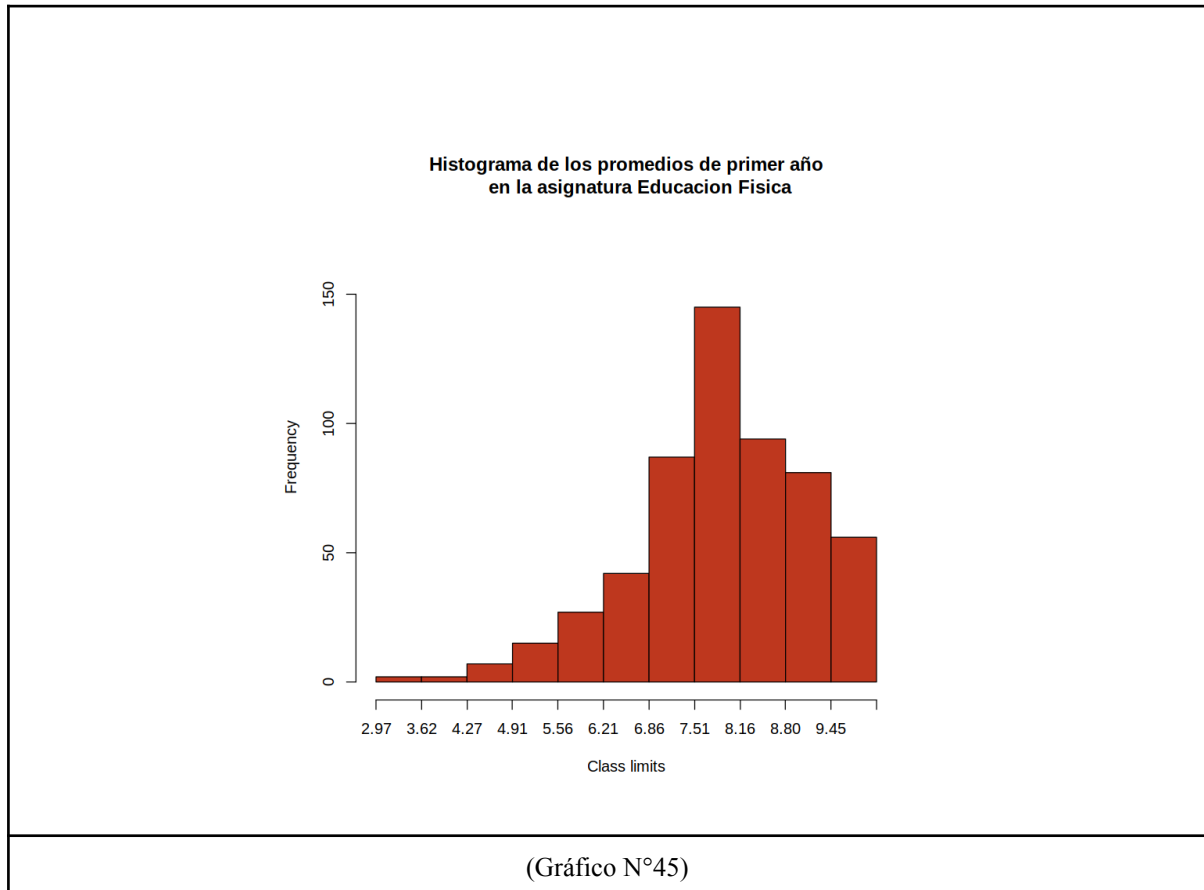
Comenzando con el primer año del nivel secundario, podemos observar en el gráfico N°28, que este tiene 594 matriculados en todo el departamento Victoria. A continuación, información que surge de un análisis del gráfico N°29 donde se comparan los distintos boxplots de promedios:

- ❖ Educación Física es la asignatura con mayor rendimiento educativo. Más del 50% de los alumnos obtuvieron un promedio de 8 o superior.
- ❖ Matemática es la asignatura con peor rendimiento, la mitad de los alumnos tiene un promedio igual o inferior a 6,33.
- ❖ Educación física presenta algunos outliers, que son promedios anormalmente bajos. Esto podría ser debido a un rendimiento muy deficiente en la asignatura o a problemas individuales que afectan a algunos estudiantes.



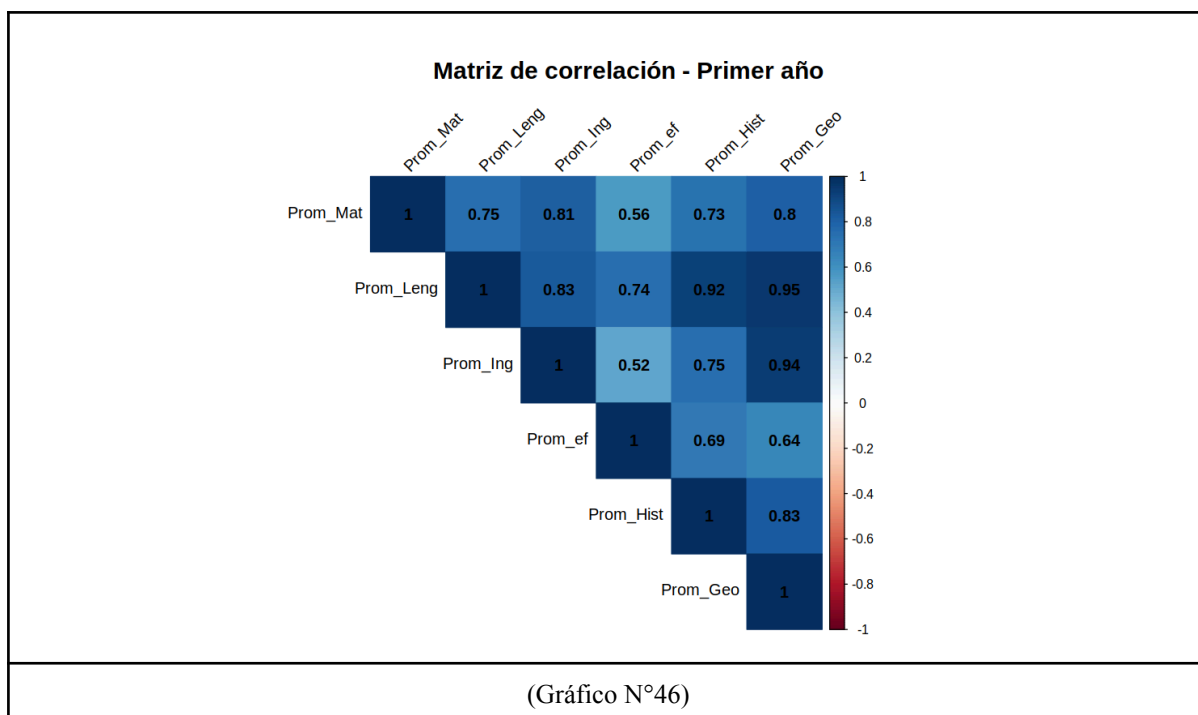
Para complementar con el gráfico N°29, hicimos el histograma (Gráfico N°30) de los promedios de la asignatura Educación Física, donde podemos constatar el alto rendimiento en

Educación Física, ya que la distribución está sesgada hacia la izquierda. Esto significa que hay más valores concentrados en el lado derecho de la media, mientras que algunos valores extremadamente bajos tiran de la media hacia la izquierda (los outliers que se ven en el boxplot). Un sesgo a la izquierda se traduce en mejores notas.

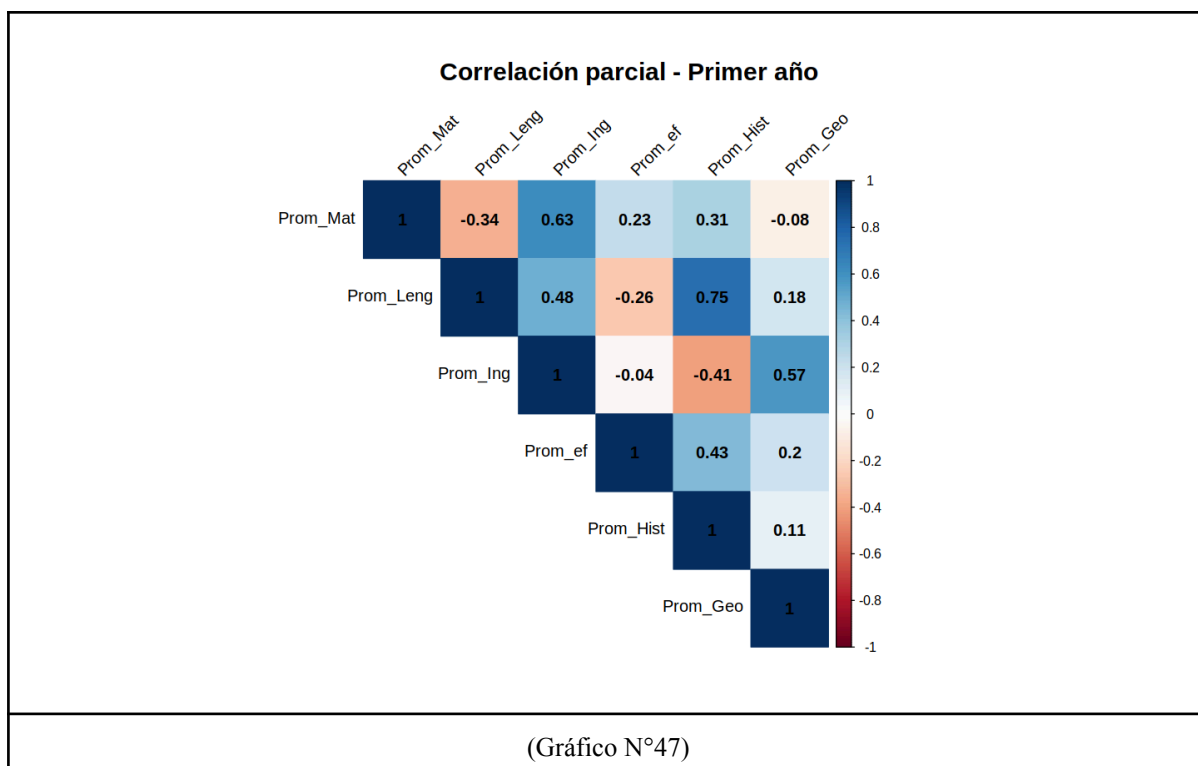


Ahora veremos un pequeño diagrama con las materias más importantes del nivel secundario, donde nos enfocaremos en las correlaciones.





Como vemos en el gráfico N°31 tenemos muchas correlaciones altas y algunas pocas moderadas. Pero ahora si calculamos las correlaciones parciales para descartar posibilidades de que alguna otra asignatura intervenga en estas correlaciones ¿Se mantendrá la dirección y fuerza de las correlaciones?



Justamente, como podemos ver en el gráfico N°32, las correlaciones bajaron considerablemente. Lo que obtenemos es lo siguiente:

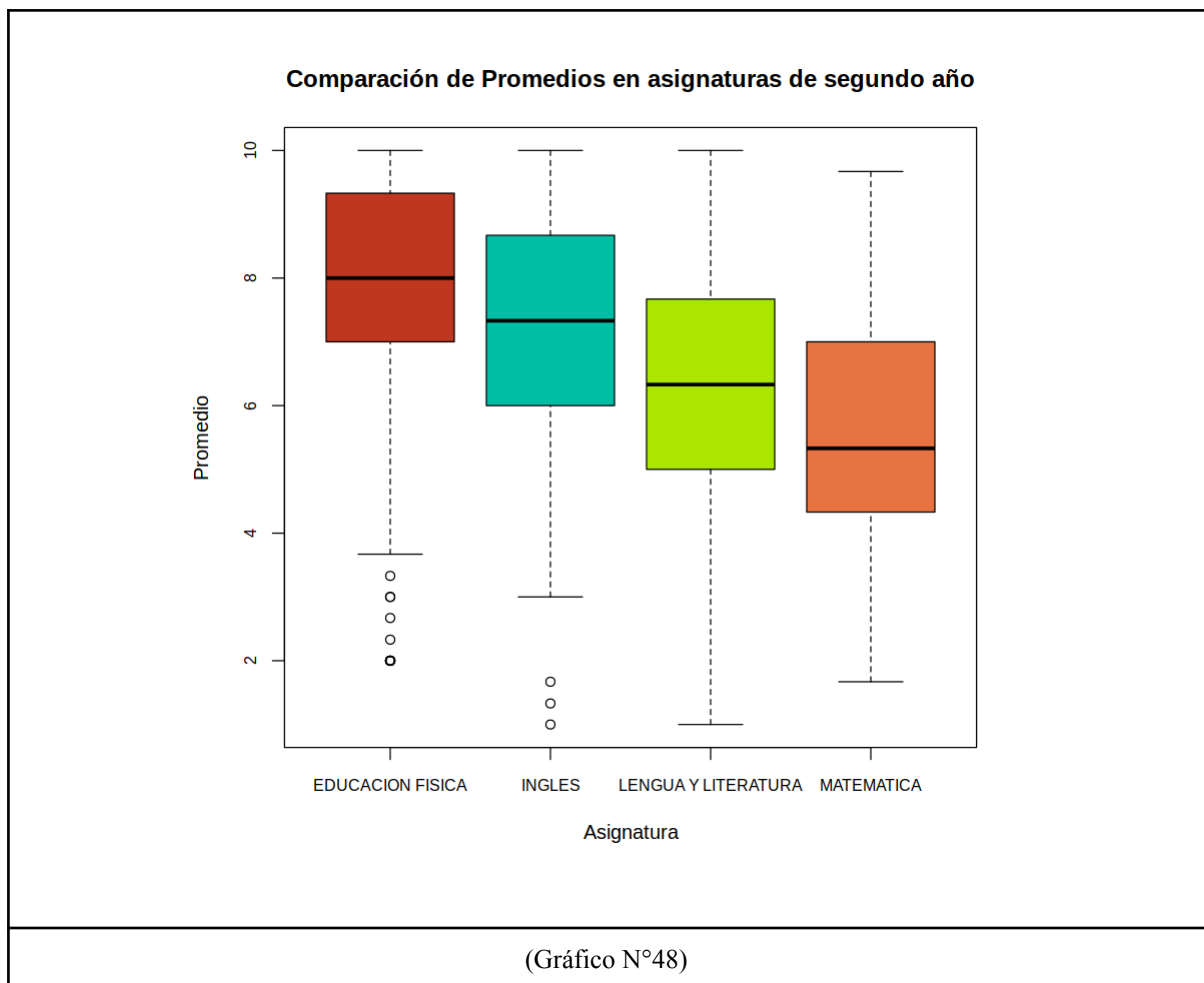
- ❖ Tenemos una fuerte correlación positiva de 0.75 entre el promedio de lengua y el promedio de historia. Este valor de correlación sugiere que, en promedio, cuando el promedio de lengua de un estudiante aumenta, es probable que también aumente su promedio de historia.
- ❖ Hay una correlación moderada negativa de -0.41 entre el promedio de inglés y el promedio de historia, lo que sugiere que el desempeño en una asignatura está inversamente relacionado de manera significativa con el desempeño en la otra asignatura. Aunque no tan fuerte como una correlación cercana a -1, -0.41 todavía indica una relación significativa entre las dos variables.
- ❖ Otras correlaciones moderadas a destacar son:
  - Una correlación de 0.63 entre el promedio de matemáticas y el promedio de inglés.
  - Una correlación de 0.57 entre el promedio de inglés y el promedio de geografía.

## **Segundo año**

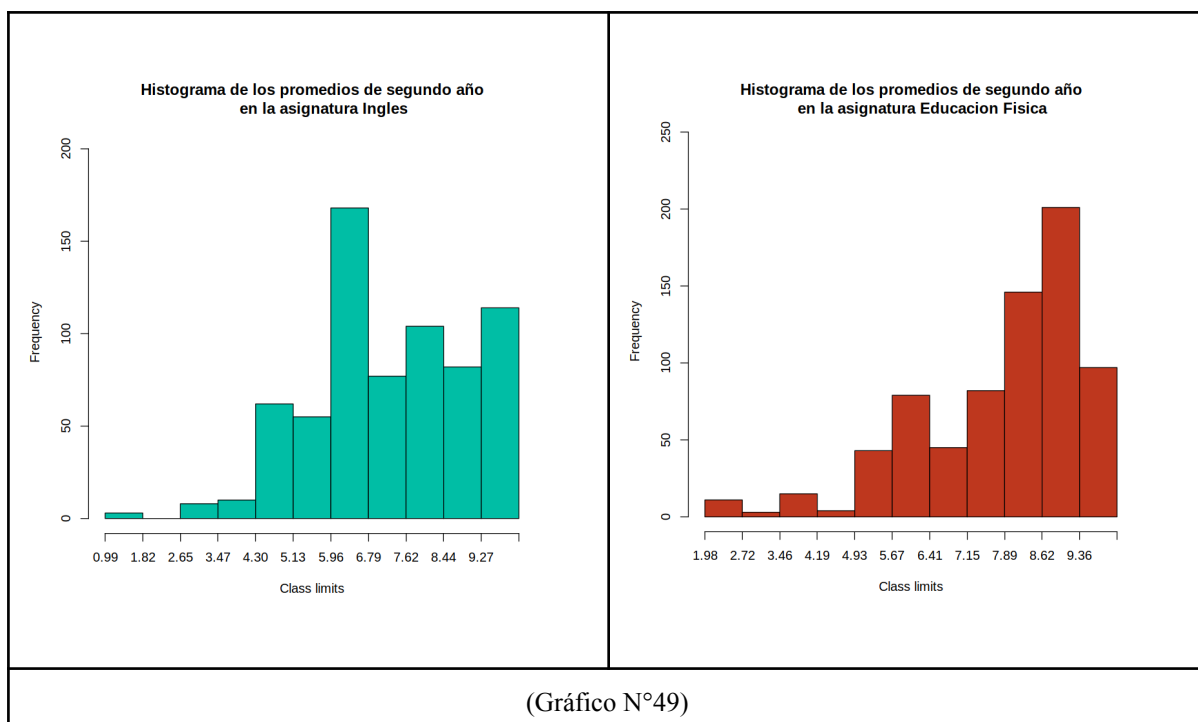
Continuando el análisis para el segundo año de los colegios secundarios, primero destacar que en total hay unos 631 alumnos, siendo este el año con mayor matrícula de todo el nivel secundario.

Para luego poder realizar una cierta relación, se continúan analizando las mismas asignaturas troncales. Entonces, al igual que hicimos antes, podemos extraer información que relevante del gráfico N°33:

- ❖ Nuevamente Educación Física es la asignatura con mejor rendimiento (la mitad de los alumnos tienen promedio 8 o superior). También en este año sigue presentando valores anormales (outliers)
- ❖ Matemática continua siendo la que posee peor rendimiento de las 4.
- ❖ La segunda con mejor rendimiento es inglés, con un 50% de sus matriculados con promedio 7,33 o mayor y un 25% de sus alumnos con promedio 9 o más. Esta tiene algunos alumnos atípicos, con promedios menores a 2.



En este año solamente hicimos los histogramas de Inglés y Educación Física(gráfico N°34), ya que nos pareció interesante poder verificar una distribución sesgada hacia la izquierda en las asignaturas. Justamente estas poseen varios outliers en las notas más bajas, lo que hace que tiren de la media hacia la izquierda.

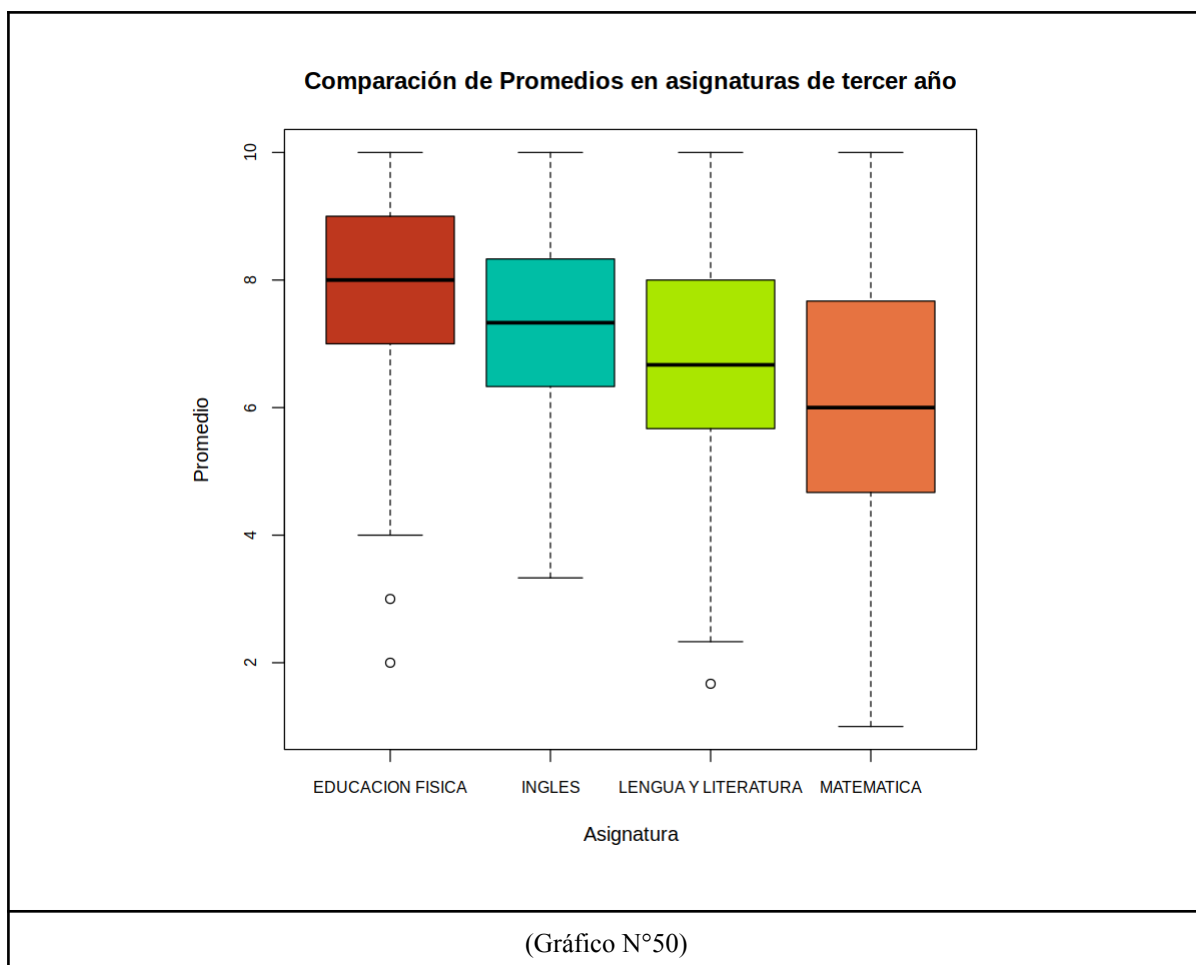


### **Tercer año**

Ahora en el tercer año del ciclo secundario, hay una totalidad de 530 alumnos inscritos.

Nuevamente comenzamos haciendo un pequeño análisis de los promedios obtenidos por el estudiantado en las asignaturas troncales:

- ❖ Se sigue repitiendo el patrón de los años posteriores, donde Educación Física es la materia con mejor rendimiento. A esta altura ya parece cotidiano que la mitad de los estudiantes tengan promedio 8 o más. Aproximadamente menos del 25% tiene una calificación inferior a 6.
- ❖ También se vuelve a repetir que Matemática es la asignatura con el rendimiento más bajo, donde solamente la mitad de los inscriptos aprobaron (promedio mayor o igual a 6). Aunque este año, hay una mayor dispersión de los datos, observando que aproximadamente el 50% de los alumnos tienen una calificación entre 5 y 8.
- ❖ Inglés y Lengua siguen teniendo un rendimiento aceptable, donde más de la mitad de sus alumnos tienen promedio igual o mayor a 6.

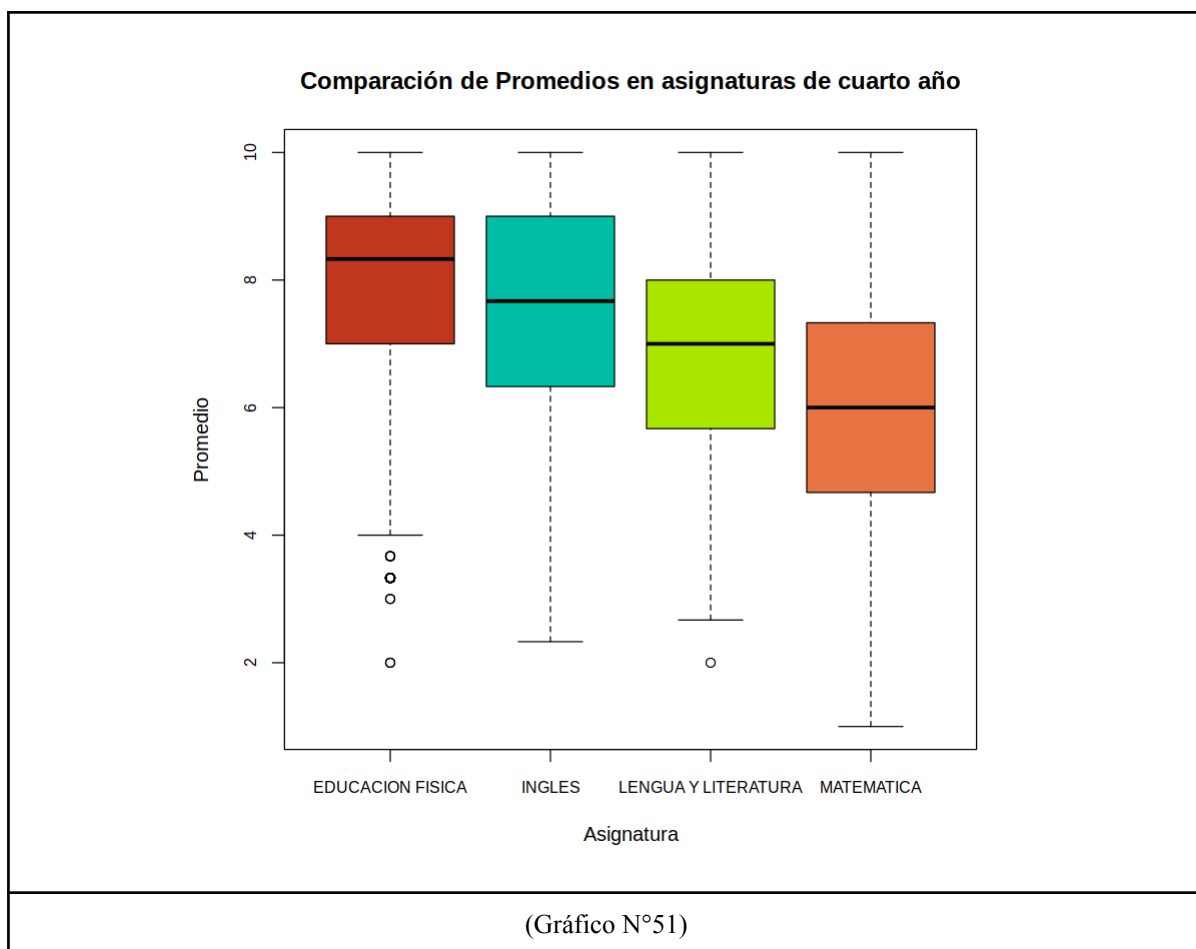


No presentamos histogramas de asignaturas, ya que no vimos que aporten información extra, se sigue repitiendo el patrón de años anteriores.

#### **Cuarto año**

El cuarto año del ciclo secundario tiene una totalidad de 481 matriculados. Algunos de los puntos a destacar del gráfico N°37 son:

- ❖ Matemática continua siendo la asignatura de peor rendimiento de las 4. Pero mantiene la misma mediana que Tercer año, donde la mitad de sus alumnos tienen promedio de 6 o más (es decir, están aprobados).
- ❖ Educación Física se destaca como la de mejor rendimiento. A diferencia de tercer año, aumentó la mediana, teniendo más de la mitad de sus alumnos con promedio 8.33 o superior. Y vale destacar que más del 75% de los inscriptos aprueban.
- ❖ Inglés y Lengua y Literatura mantienen las tendencias de años previos. En este año, el 25% de los alumnos de Lengua obtuvieron una nota de 8 o superior.



### **Quinto año**

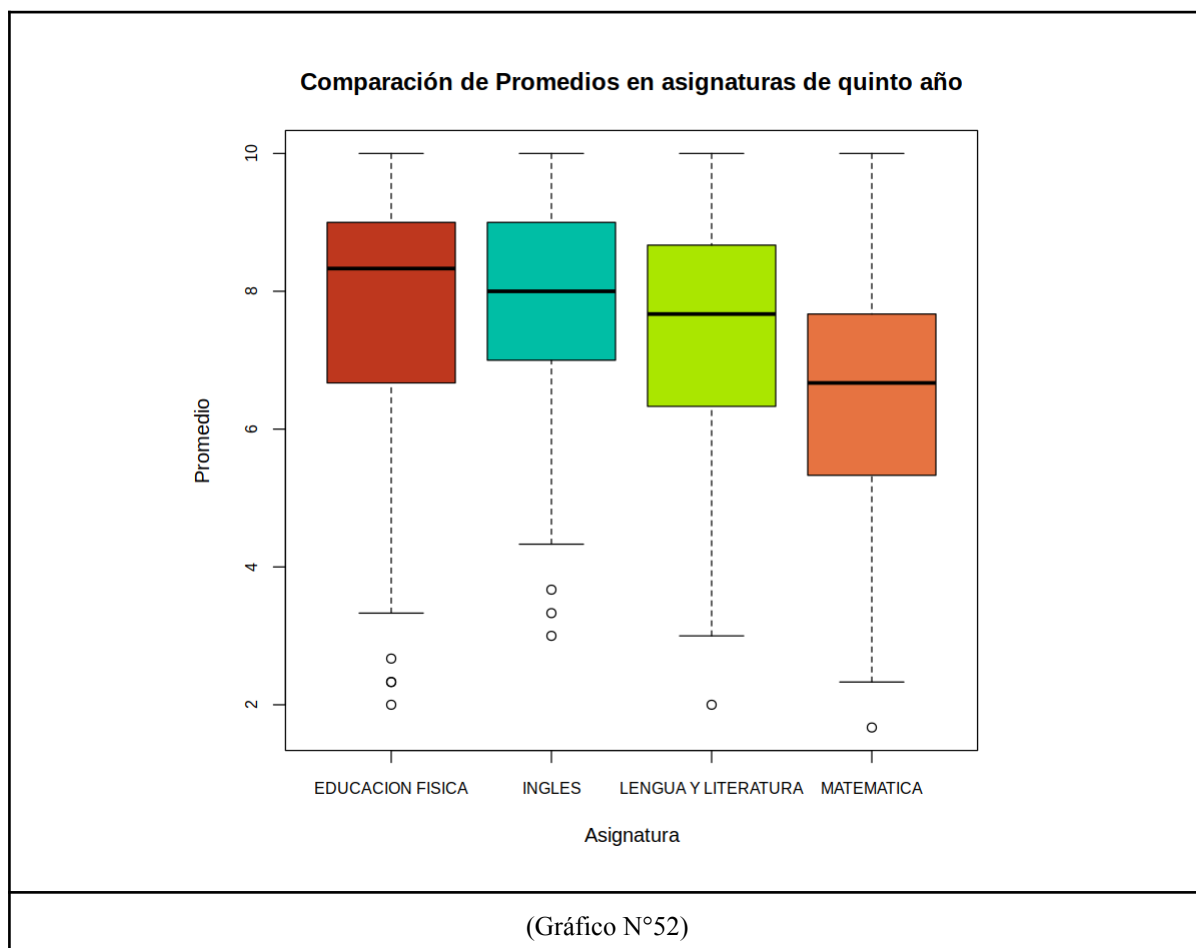
El quinto año del ciclo secundario cuenta con un total de 396 matriculados. Se puede ver a lo largo de los años, cómo va decayendo el total de alumnos.

Como la información con la que contamos no es secuencial, es decir no tenemos los datos de la trayectoria de los mismos estudiantes, no podríamos asegurar que esto se debe a que hubo deserción por parte de los estudiantes. Tampoco podemos encontrar, con los datos que tenemos, una explicación del porqué de este suceso, simplemente resta comentarlo y mencionarlo en este informe.

Continuando con el análisis de los promedios de las asignaturas troncales, destacamos la siguiente información de este año:

- ❖ Educación Física sigue siendo la asignatura de mayor rendimiento, aunque este año se ve como los promedios se dispersan en un mayor rango, donde la caja central que representa el 50% de los datos incrementó su tamaño.
- ❖ La materia con peor rendimiento de las 4 (Matemática) tiene más del 50% de sus alumnos con promedio igual o mayor a 6,77. Se nota un incremento de la mediana en un 0.77 respecto al año anterior.

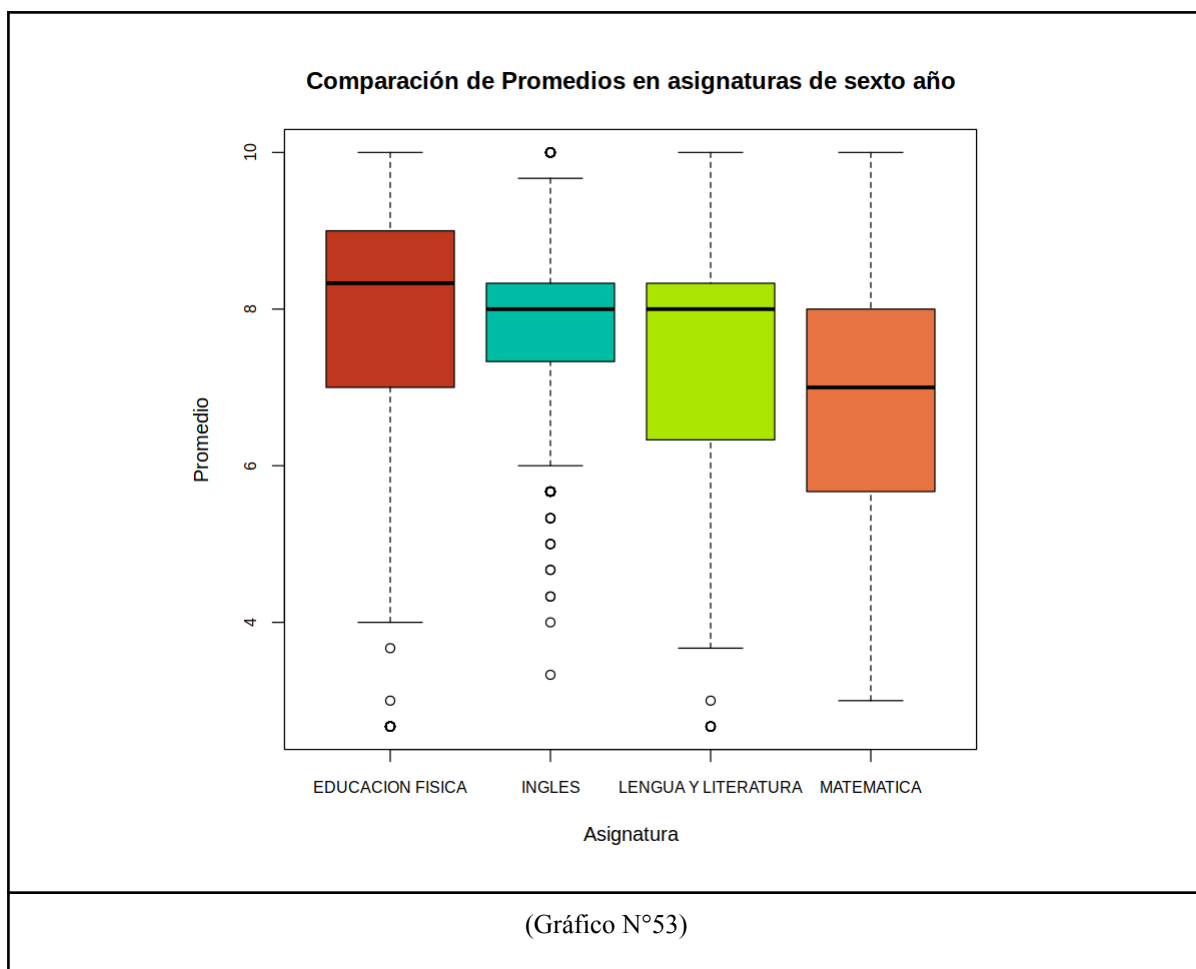
- ❖ Destacar que más de la mitad de los alumnos de Lengua y Literatura tiene un promedio de 8 o superior. En el caso de Educación Física e Inglés, más del 25% de los estudiantes tienen promedio 9 o superior.



### **Sexto año**

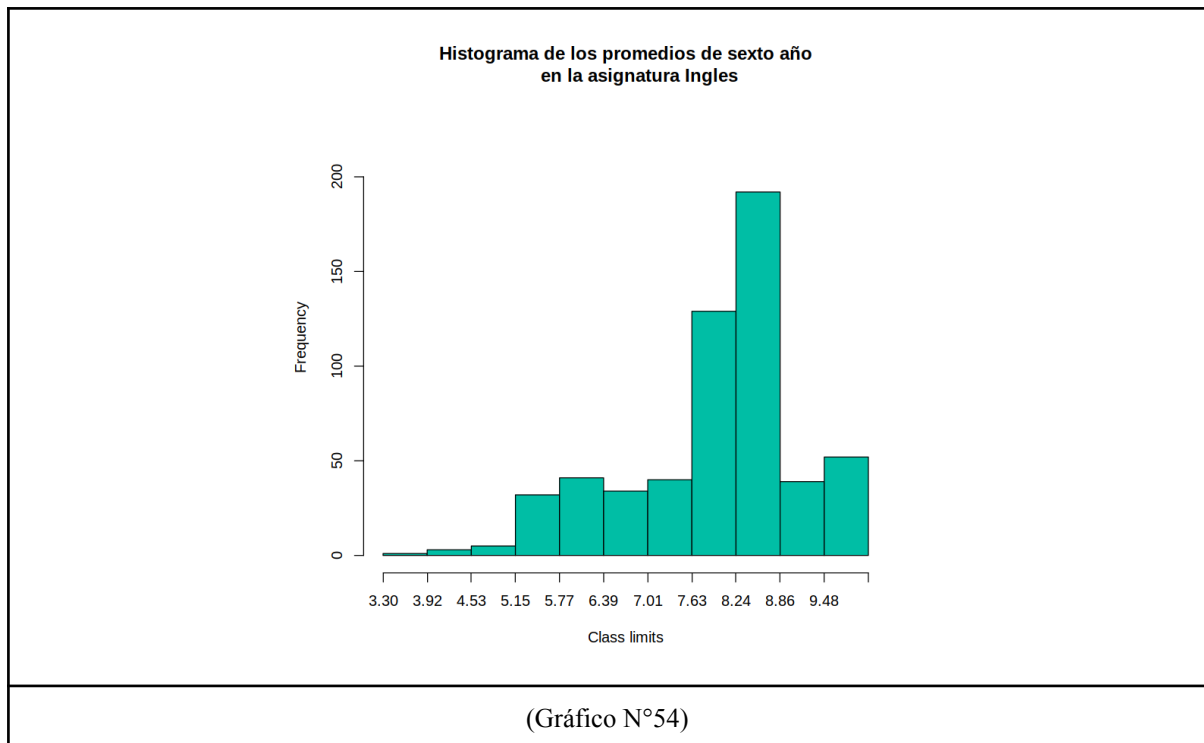
El Sexto año de los colegios secundarios en el departamento Victoria tiene un total de 371 matriculados. Una posible lectura del gráfico N°41 sería:

- ❖ Se mantiene Educación Física como la asignatura con mejor rendimiento.
- ❖ Inglés tiene muy poca dispersión en los promedios de sus alumnos, pero presenta muchos valores atípicos, que están debajo del promedio 6.
- ❖ A pesar de ser la de peor rendimiento de las 4, Matemática vuelve a registrar un incremento en sus notas respecto a Quinto. La mitad de sus estudiantes tiene una nota mayor o igual a 7. Registrando sexto año, el mejor rendimiento en Matemática respecto a los demás.



En el gráfico N°42 mostramos el histograma de la asignatura Inglés, es interesante destacar como se ve en el histograma la baja distribución de los datos. En el boxplot vemos que el 50% de los datos centrales están muy comprimidos y en el histograma podemos ver que la moda se corresponde con la ubicación de la caja central, entre 8.24 y 8.86.



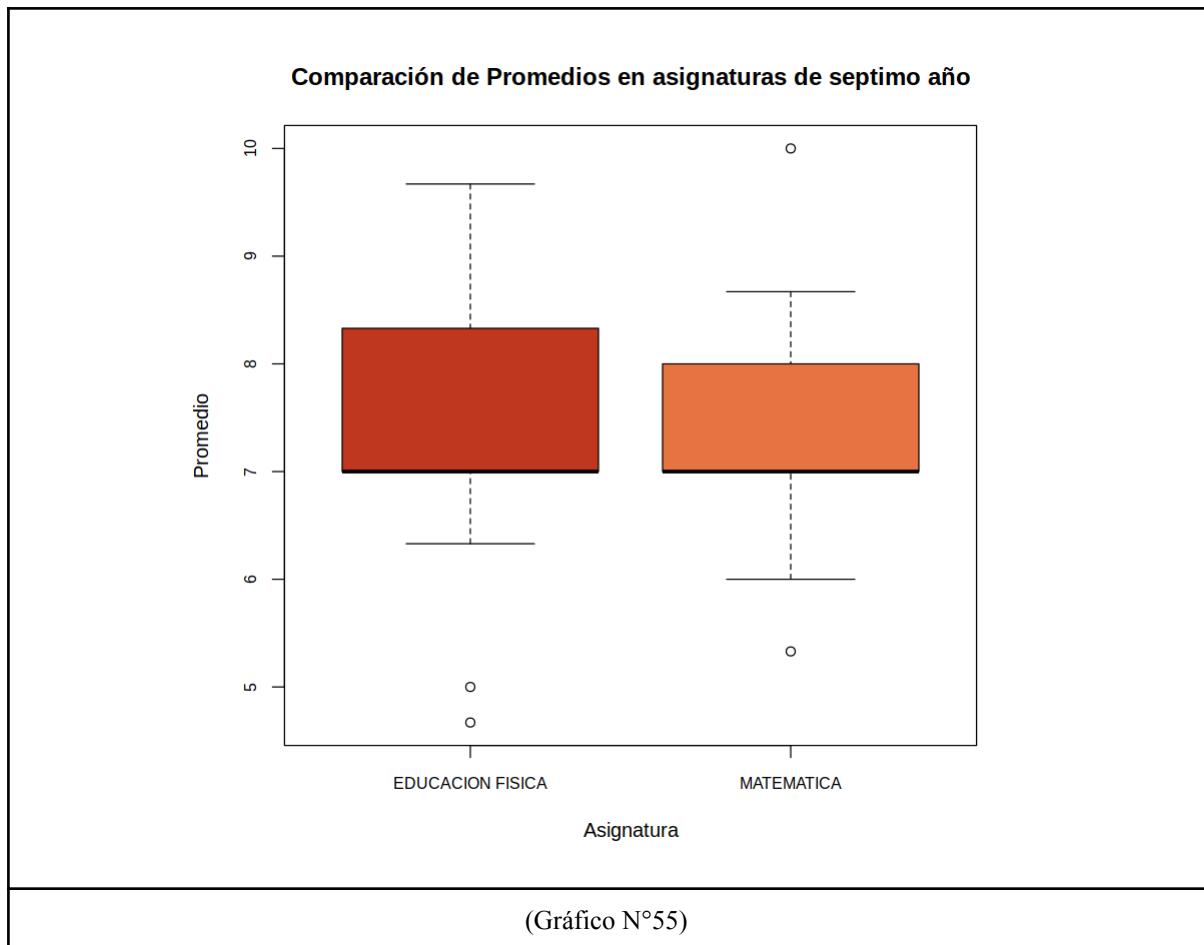


### **Séptimo año**

Este año en particular, registra una baja en los matriculados, teniendo solamente 50 alumnos. Esto puede deberse a que solamente los colegios técnicos y agrotécnicos tienen siete años. Además el 83% son colegios comunes y solamente el 17% se corresponde a colegios técnicos/agrotécnicos.

Ahora bien, la observación del gráfico N°43 nos permite leer ciertas cosas:

- ❖ Ambas asignaturas tienen una mediana de 7, es decir, más de la mitad de sus estudiantes obtuvieron un promedio de 7 o superior.
- ❖ El boxplot solamente tiene dos lados, esto significa que la mediana (o segundo cuartil) coincide con uno de los extremos. Esto puede ocurrir cuando hay pocos datos, o cuando están muy agrupados alrededor de la mediana (solamente tenemos 50 datos)

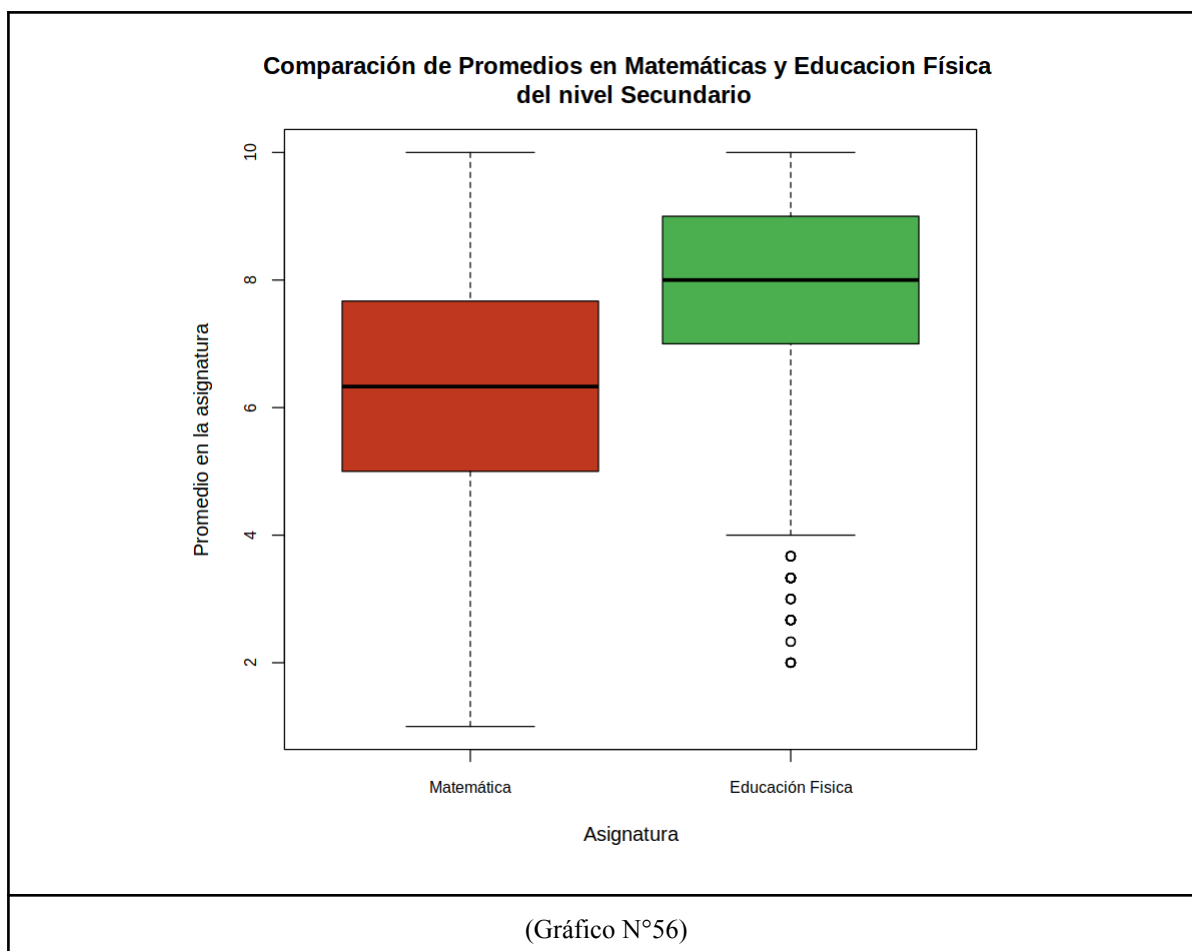


### Comparación del rendimiento entre Matemática y Educación Física.

En este apartado, vamos a intentar responder a la pregunta de si hay una diferencia en el rendimiento entre Matemática y Educación Física. La idea es respaldar lo visto en los boxplots presentados en el apartado anterior y lograr justificarlo con algún test estadístico.

Lo vamos a analizar a nivel general, teniendo en cuenta los promedios de ambas asignaturas de 1ero a 6to año.

Lo primero que hacemos es presentar un boxplot comparativo, de los promedios de ambas asignaturas. En el gráfico N°X podemos ver esta comparación. A simple vista podemos notar que el rendimiento en Educación Física es superior al rendimiento en Matemática.



Para comparar los promedios de Educación Física y Matemática de los alumnos del nivel secundario, se utilizó el test de Wilcoxon ya que los datos no tenían una distribución normal. Verificamos la normalidad con el test de Kolmogorov-Smirnov. Los resultados de las pruebas fueron los siguientes:

- ➔ Para los datos de Matemática, el p-value fue menor a  $2.2e-16$ , lo que es inferior al nivel de significancia  $\alpha = 0.05$ . Esto nos lleva a rechazar la hipótesis nula ( $H_0$ ) y concluir que los datos no siguen una distribución normal.
- ➔ Para los datos de Educación Física, el p-value también fue menor de  $2.2e-16$ , lo cual es inferior al nivel de significancia  $\alpha = 0.05$ . Por lo tanto, también rechazamos la hipótesis nula ( $H_0$ ) para esta asignatura y concluimos que los datos no siguen una distribución normal.

Estos resultados sugieren que los datos de rendimiento en ambas asignaturas no se distribuyen normalmente, lo que conlleva a que usemos el test no paramétrico de Wilcoxon. Las hipótesis de este test son:

- H0: La mediana de las notas en Matemática es mayor que la mediana de las notas en Educación Física
- H1: La mediana de las notas en Matemática es menor que la mediana de las notas en Educación Física

Las salidas de este tests fue conciso, el p-value nos dió  $< 2.2e-16$ , indicando que hay evidencia suficiente para rechazar la hipótesis nula, dando entender que el rendimiento en Matemática es menor que el rendimiento en Educación Física.

## Variación del rendimiento en Ciclo Básico y Ciclo Orientado.

En esta sección, analizamos si existen diferencias significativas en el rendimiento académico dentro de cada ciclo. Tomamos Matemática como referencia, ya que es una asignatura que se dicta todos los años y es de las más importantes.

Primero antes de comenzar con el análisis, cabe aclarar que no consideramos séptimo año, ya que la muestra es muy pequeña (tiene 21 observaciones, en comparación con los 600 en promedio de 1ro a 6to). Además tomamos en conjunto los ciclos orientado (secundarias comunes de 4to a 6to) y especializado (secundaria técnicas y agrotécnicas 4to a 7mo).

Se verificó, como en el apartado anterior, la normalidad de los grupos con el test de Kolmogorov-Smirnov. Los p-values obtenidos en los 6 grupos fueron:

Grupo	p-value
Primero	p-value = 5.983e-05
Segundo	p-value = 0.00751
Tercero	p-value = 0.001581
Cuarto	p-value = 0.1212
Quinto	p-value = 3.504e-06
Sexto	p-value = 8.119e-05
Fuente de elaboración propia (Tabla N°16)	

En cuarto el p-value  $> 0.05$ , lo que lleva a un no rechazo de H0, por lo tanto la distribución de los datos es normal. En todos los otros grupos, el p-value  $< 0.05$  lo que lleva a un rechazo de H0, entonces la distribución de los datos no es normal. En este caso, como solo cuarto se distribuye normalmente, usamos un test no paramétrico que no requiera asumir normalidad.

Entonces, para evaluar si el rendimiento académico en Matemática varía significativamente a lo largo del Ciclo Básico (primeros tres años de secundaria), realizamos un análisis utilizando la prueba no paramétrica de Kruskal-Wallis, debido a que los datos no siguen una distribución normal y los tres grupos son independientes. Esta prueba nos permite comparar las distribuciones de notas entre los tres cursos (Primero, Segundo y Tercero). Las hipótesis son las siguientes:

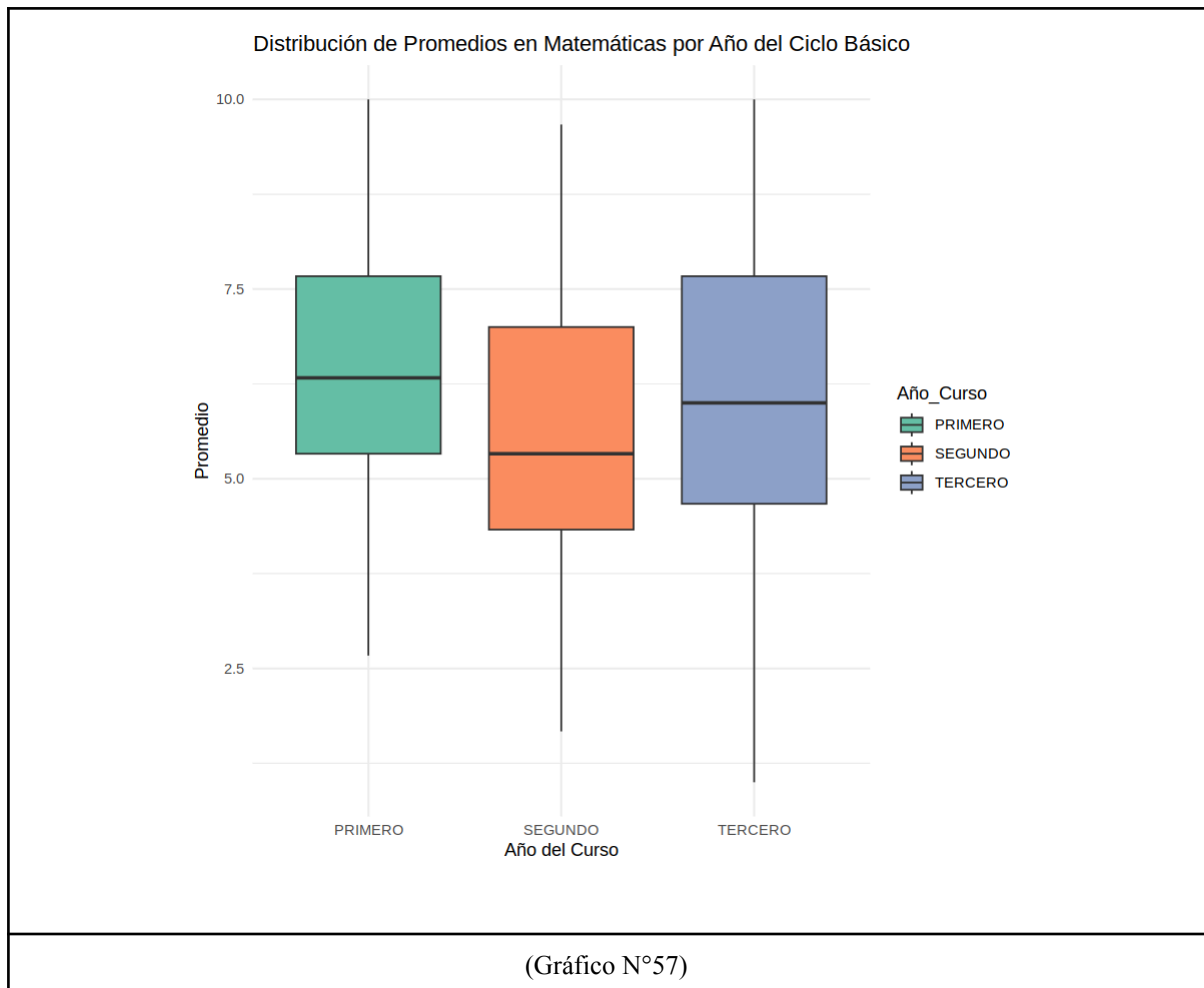
- H0: Las distribuciones de las notas de los cursos "PRIMERO", "SEGUNDO" y "TERCERO" son iguales.
- H1: Al menos una de las distribuciones de las notas es diferente.

Obtenemos un  $p\text{-value} < 2.2e-16 < (0.05)$ , entonces rechazamos H0, lo que me lleva a concluir que “al menos una de las distribuciones de las notas es diferente”. Como encontramos que existe una diferencia, lo que resta es hacer el test post-hoc de Pairwise Wilcoxon. Los resultados de las comparaciones post-hoc fueron los siguientes:

- Entre Primero y Segundo:  $p\text{-valor} < 2e-16$ , indicando una diferencia significativa.
- Entre Primero y Tercero:  $p\text{-valor} = 0.046$ , indicando una diferencia significativa\*.
- Entre Segundo y Tercero:  $p\text{-valor} = 1.3e-11$ , indicando una diferencia significativa

\*aunque debido a su proximidad al umbral de significancia (0.05), la evidencia no es concluyente y podría dar lugar a que no haya una diferencia significativa.

Justamente esto lo podemos verificar con el gráfico N°X de abajo, donde se ve que las distribuciones de promedios de Primero (*mediana*=6.330) y Tercero(*mediana* = 6.00) son bastantes similares, siendo la que tiene mayor diferencia Segundo(*mediana* = 5.330). Mirando el boxplot y teniendo en cuenta los resultados del test de hipótesis, podríamos decir que el rendimiento no es constante en el ciclo básico, siendo primero y tercero muy similares entre sí (*mediana de 6.330 y 6.00*) y segundo muestra una caída del resto (*mediana de 5.330*).

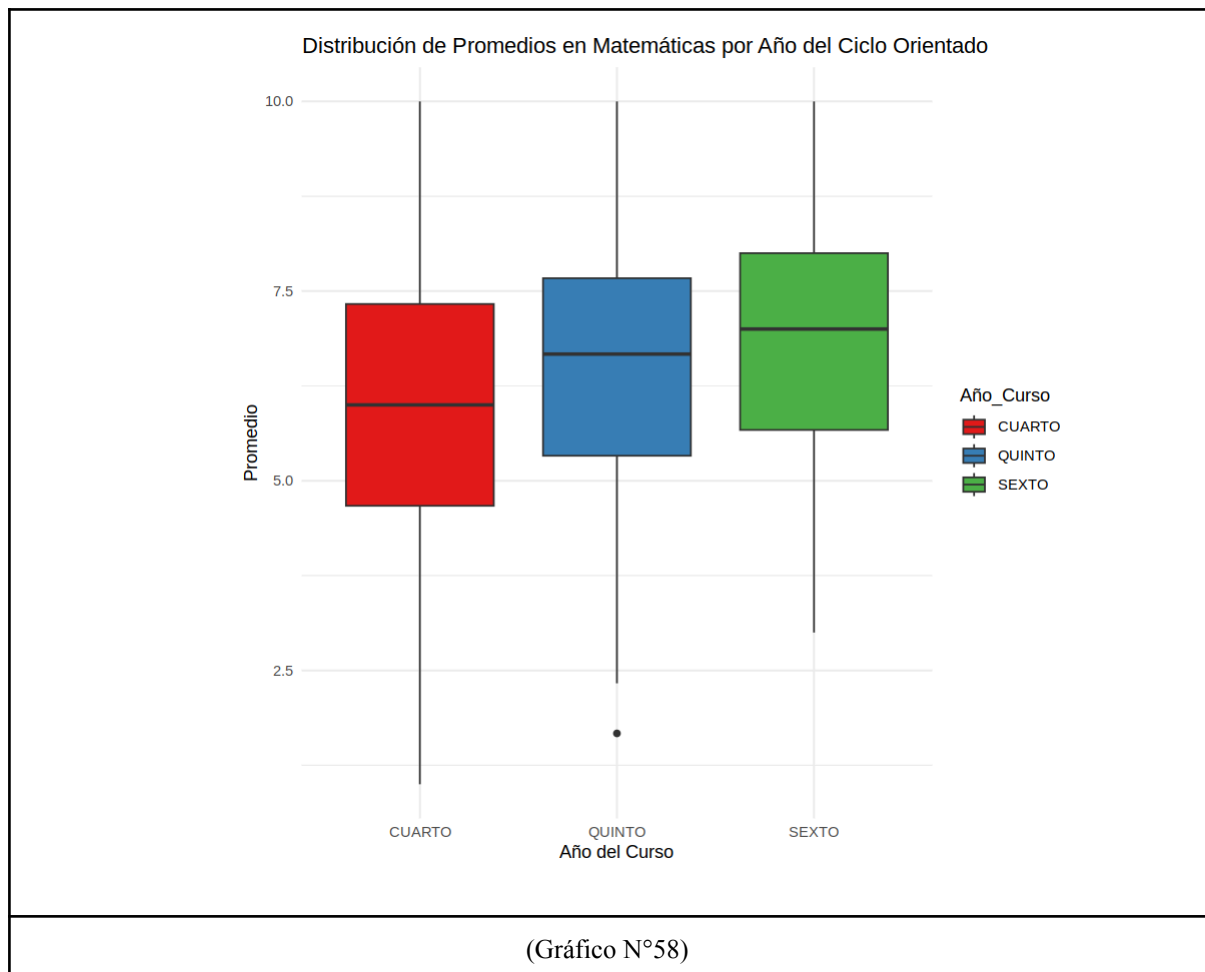


Repetiendo el análisis para el Ciclo Orientado (últimos tres años de secundaria), realizamos un análisis utilizando la prueba no paramétrica de Kruskal-Wallis, y obtenemos un  $p\text{-value} < 2.2e-16 < (0.05)$ , entonces rechazamos  $H_0$ , lo que me lleva a concluir que “al menos una de las distribuciones de las notas es diferente”. Como encontramos que existe una diferencia, lo que resta es hacer el test post-hoc de Pairwise Wilcoxon. Los resultados de las comparaciones post-hoc fueron los siguientes:

- Entre Cuarto y Quinto:  $p\text{-valor} = 2.8e-09$ , indicando una diferencia significativa.
- Entre Quinto y Sexto:  $p\text{-valor} = 0.016$ , indicando una diferencia significativa.
- Entre Segundo y Tercero:  $p\text{-valor} < 2e-16$ , indicando una diferencia significativa.

Justamente esto lo podemos verificar con el gráfico N°X de abajo, donde se ve que las distribuciones de promedios cuarto(*mediana* = 6.000), quinto(*mediana* = 6.670) y sexto(*mediana* = 7.00) difieren. Mirando el boxplot y teniendo en cuenta los resultados del test de hipótesis, podríamos decir que el rendimiento en Matemática no es constante en el

ciclo orientado. Lo que sucede es que el rendimiento en Matemática se incrementa a lo largo de los años.



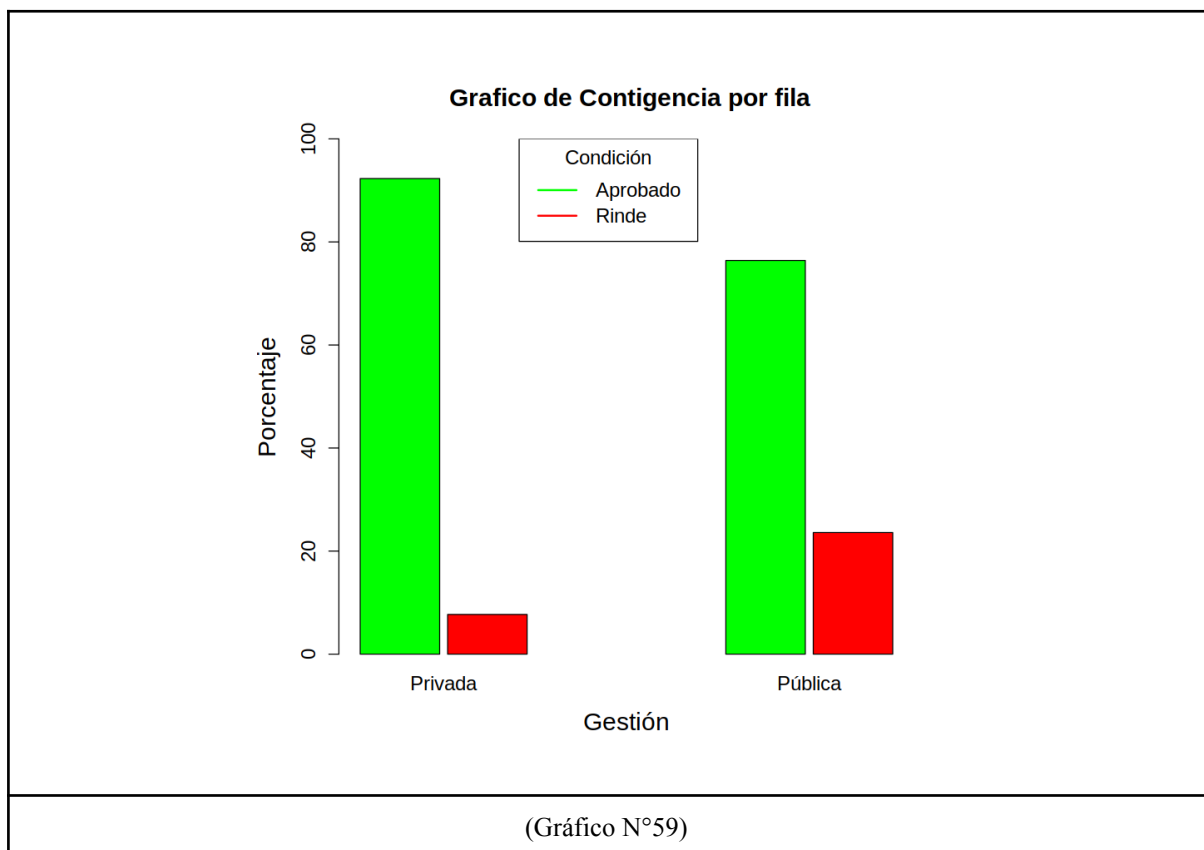
## Impacto del Tipo de Gestión en la Condición Final del Alumno

En esta sección, se analiza si el tipo de gestión de las escuelas (pública o privada) tiene un impacto en la condición final de los alumnos (aprobado o rinde). Este análisis busca identificar posibles diferencias en los resultados académicos de los estudiantes en función del tipo de gestión de sus escuelas, proporcionando una visión más profunda sobre cómo la gestión educativa puede influir en el rendimiento académico. En este caso, el análisis lo hacemos a nivel general, no hurgamos año por año.

Para realizar el análisis bivariado cualitativo, obtuvimos la tabla ordenada y con sus totales marginales tanto en filas como en columnas, esta quedó conformada de la siguiente manera:

	Aprobado	Rinde	Total
Pública	35102	10862	45964
Privada	5279	443	5722
Total	40381	11305	51686
Fuente de elaboración propia (Tabla N°17)			

La tabla de contingencia con los porcentajes por fila fue representada con un gráfico de barras adosadas:



Utilizamos la prueba de Chi-cuadrado de Pearson para evaluar la relación entre el tipo de gestión y la condición final del alumno. Los resultados de esta prueba indicaron una diferencia significativa entre las distribuciones de gestión y condición final ( $p\text{-value} < 2.2e-16$ ). Esto sugiere que existe una asociación estadísticamente significativa entre el tipo de gestión de la escuela y la condición final de los estudiantes.



Para medir la fuerza de esta asociación, calculamos el coeficiente V de Cramer, que resultó ser 0.121. Este valor indica una asociación débil entre el tipo de gestión y la condición final de los alumnos. Aunque la relación es estadísticamente significativa, la magnitud de la asociación es relativamente pequeña.

Estos resultados sugieren que, si bien el tipo de gestión tiene un impacto en la condición final de los alumnos, como la magnitud de la asociación es débil, otros factores probablemente desempeñan un papel más importante en determinar el éxito académico de los estudiantes.

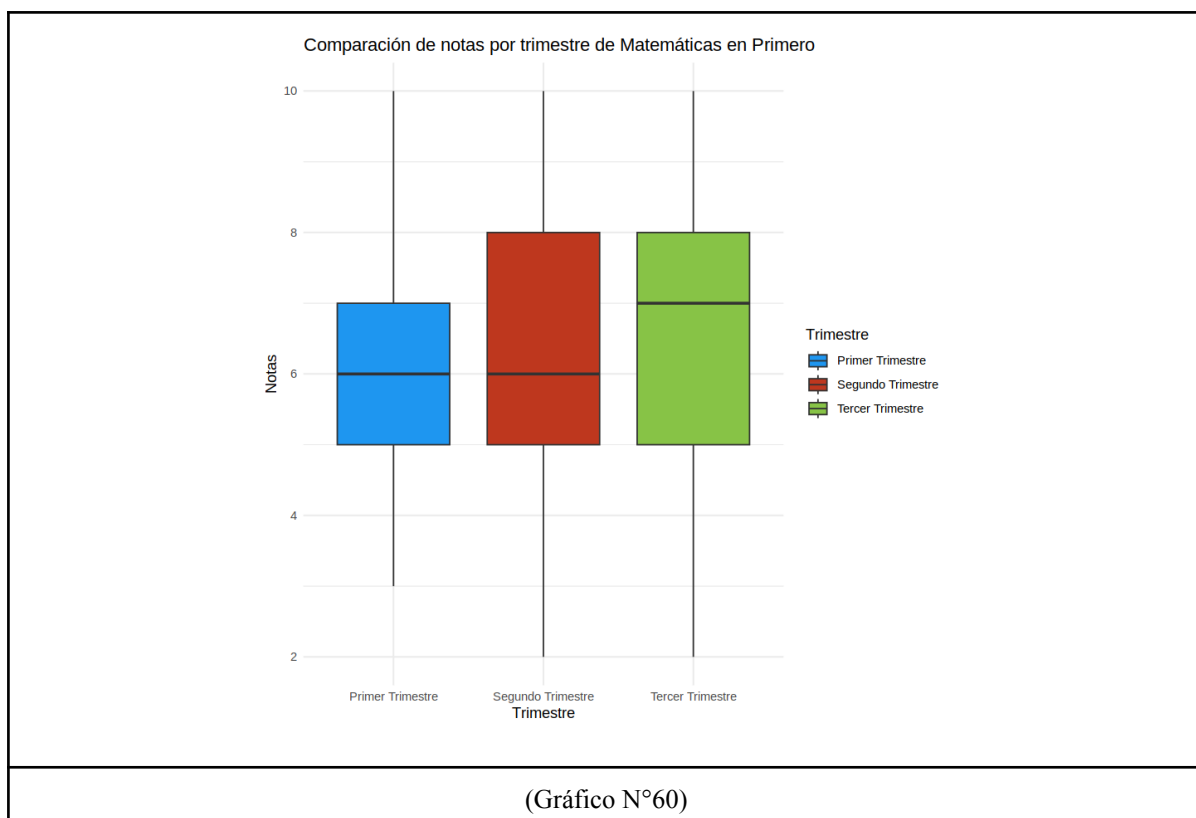
---

### Comparación del Rendimiento en los Tres Trimestres (Año por Año)

En esta sección, se comparará el rendimiento académico de los estudiantes a lo largo de los tres trimestres del año académico. Este análisis se realiza utilizando el test de Friedman, que es adecuado para comparar tres o más medias en muestras pareadas. Dado que estamos trabajando con la misma muestra de estudiantes en distintos momentos del año, este test es apropiado para evaluar si existen diferencias significativas en el rendimiento a lo largo del tiempo.

Para llevar a cabo este análisis, utilizamos las notas obtenidas en la asignatura de Matemática. Matemática es una asignatura fundamental y su rendimiento puede proporcionar una visión clara de la evolución académica de los estudiantes a lo largo del año. Al comparar los tres trimestres, podremos identificar patrones de mejora o deterioro en el rendimiento académico.

#### **Primero**



A simple vista parece ser que el tercer trimestre tiene una distribución diferente a los otros dos trimestres. Vamos a comprobarlo con un test. Primero, verificamos la normalidad de los residuos utilizando el test de Shapiro-Wilk. El test de Shapiro-Wilk muestra que los residuos no se distribuyen normalmente ( $p\text{-value} < 2.2e-16$ ), por lo que es apropiado usar el test no paramétrico de Friedman, el cual tiene como hipótesis:

- H0: No hay diferencias significativas en la condición final del alumno entre los trimestres.
- H1: Al menos un trimestre tiene una diferencia significativa en la condición final del alumno.

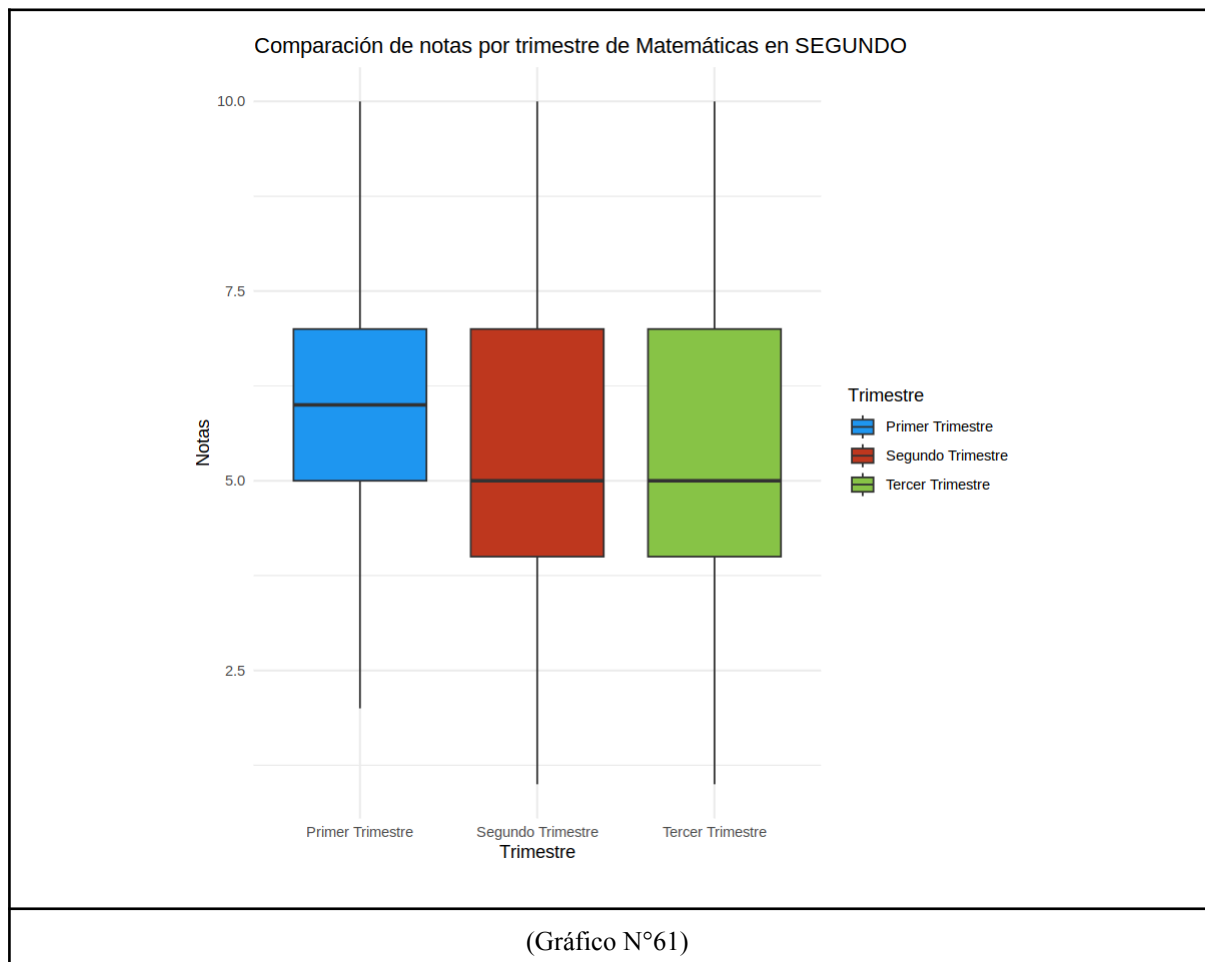
Los resultados del test de Friedman indican que no hay diferencias significativas entre los trimestres ( $p\text{-value} = 0.2244$ ). Calculamos la potencia del test para evaluar la robustez de nuestros resultados, la cual fue alta, reforzando la validez de estos hallazgos.

Si vemos las salidas del test de Pairwise, vemos que no hay diferencias entre los trimestres.

	Primer Trimestre	Segundo Trimestre
Segundo Trimestre	0.32	-
Tercer Trimestre	0.53	0.94
Fuente de elaboración propia (Tabla N°18)		

En conclusión, podemos afirmar que los alumnos mantuvieron un rendimiento estable en Matemática durante los tres trimestres del primer año escolar, sin diferencias significativas que indiquen variaciones en su desempeño académico.

## Segundo



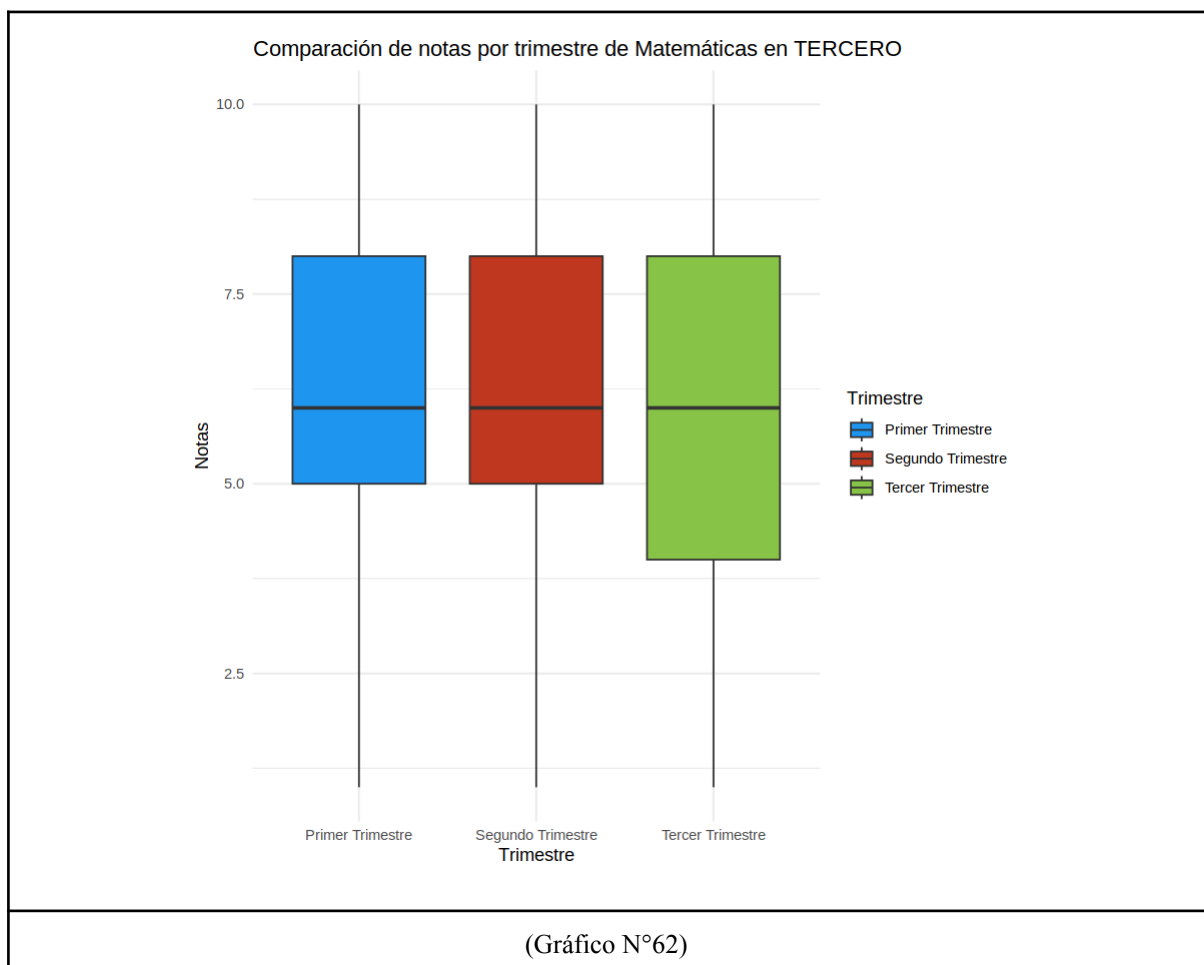
Mirando el gráfico de arriba, podríamos intuir que el rendimiento es diferente en el primer trimestre del segundo año.

Al igual que en primero, se verificó que los residuos no se distribuyen normalmente, por lo tanto es adecuado continuar con el test no paramétrico. El test de Friedman nos dice que si hay diferencias entre las medianas de los grupos ya que devuelve un  $p\text{-value} = 1.056e-15$ . Este resultado es confiable ya que su test de potencia es de 1, siendo muy fuerte. Además, al examinar las comparaciones de pares utilizando el test de Wilcoxon, encontramos que todas las comparaciones tienen  $p\text{-values}$  significativamente bajos, lo que respalda la presencia de diferencias significativas en el rendimiento entre los trimestres.

	Primer Trimestre	Segundo Trimestre
Segundo Trimestre	$< 2e-16$	-
Tercer Trimestre	$< 2e-16$	1.6e-11
Fuente de elaboración propia (Tabla N°19)		

Dado que la mediana de las notas en el primer trimestre es 6 y la mediana en el segundo y tercer trimestre es 5, podemos observar una disminución en el rendimiento promedio de los estudiantes a lo largo del año académico. Aunque las pruebas estadísticas mostraron diferencias significativas entre los grupos, esta diferencia puede no ser lo suficientemente grande como para considerarla significativa en la práctica. Sí podemos afirmar que el rendimiento es superior en el primer trimestre y luego decae.

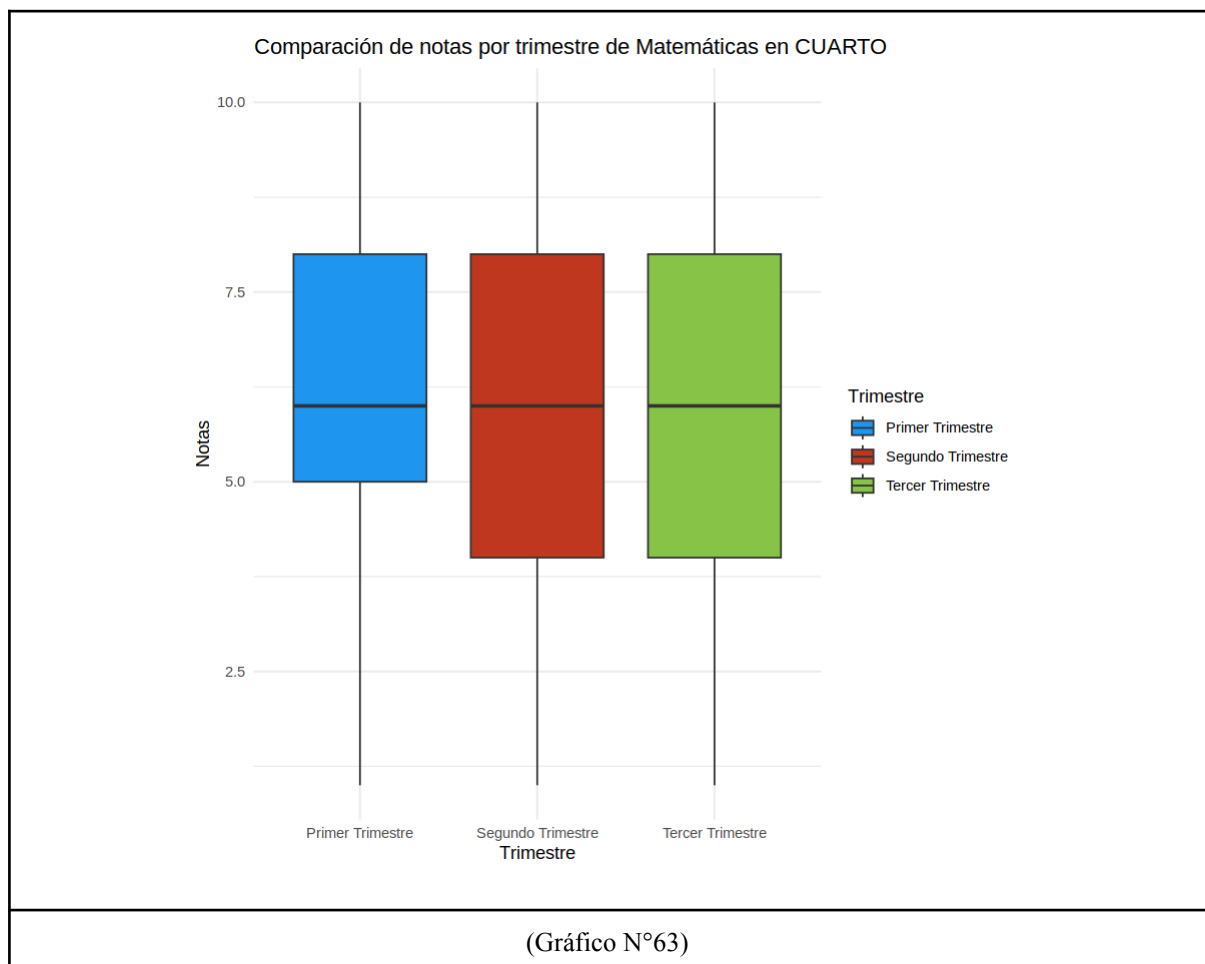
### **Tercero**



Una primera observación de los boxplots, es que el rendimiento parece parejo, teniendo todos los trimestres una mediana de 6. Vamos a ver que dicen los test:

Lo primero es que la prueba de Shapiro-Wilk mostró una distribución no normal de los residuos, por eso procedemos con el test de Friedman, el cual no mostró diferencias significativas entre los grupos (trimestres), dándonos un  $p\text{-value} = 0.05709$ . Con este  $p\text{-value}$  no rechazamos  $H_0$  y concluimos que no hay diferencias significativas. En términos generales los trimestres de tercer año no parecen influir en el rendimiento de Matemáticas.

#### Cuarto



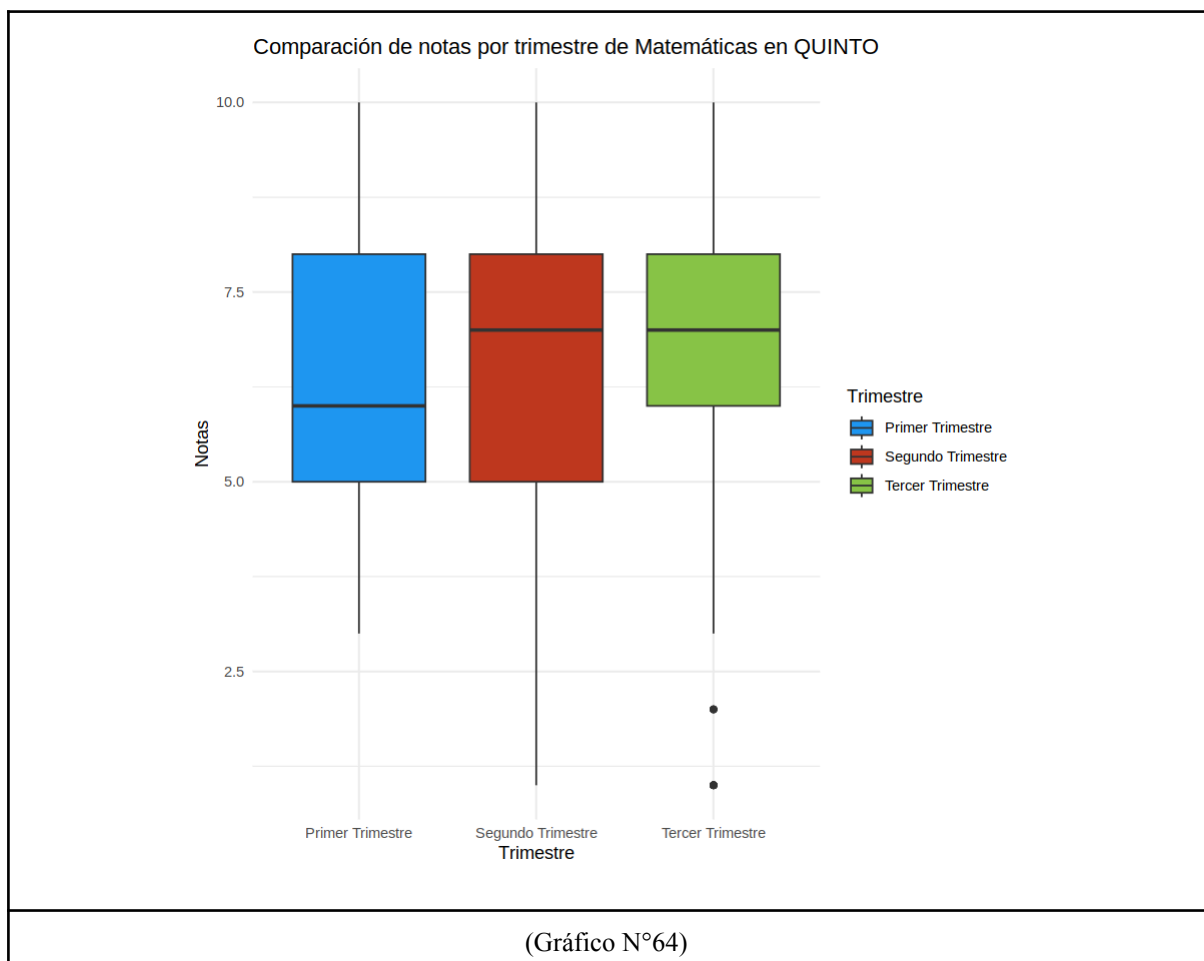
Nuevamente, una primera observación de los boxplots arroja que el rendimiento parece parejo, teniendo todos los trimestres una mediana de 6. Vamos a ver que dicen los test:

El análisis de Friedman mostró diferencias significativas en el rendimiento entre los tres trimestres ( $p\text{-valor} = 1.588e-05 < 0.05$ ). Esto sugiere que el trimestre sí tiene un impacto en el rendimiento académico de los estudiantes de cuarto año. Las pruebas post hoc revelaron diferencias significativas (todos los  $p\text{-values} < 0.05$ ) en las comparaciones entre el primer y segundo trimestre, así como entre el primer y tercer trimestre.

	Primer Trimestre	Segundo Trimestre
Segundo Trimestre	2.9e-05	-
Tercer Trimestre	2.3e-07	0.001
Fuente de elaboración propia (Tabla N°20)		

Finalmente, aunque las medianas sean iguales (6), el test de Friedman detecta diferencias significativas entre los grupos basándose en los rangos de los datos, lo que significa que aunque las medianas pueden ser similares, la distribución de los datos puede ser diferente. En conclusión, esto no nos alcanza para decir que el rendimiento de matemática en cuarto año varía dependiendo el trimestre.

### Quinto



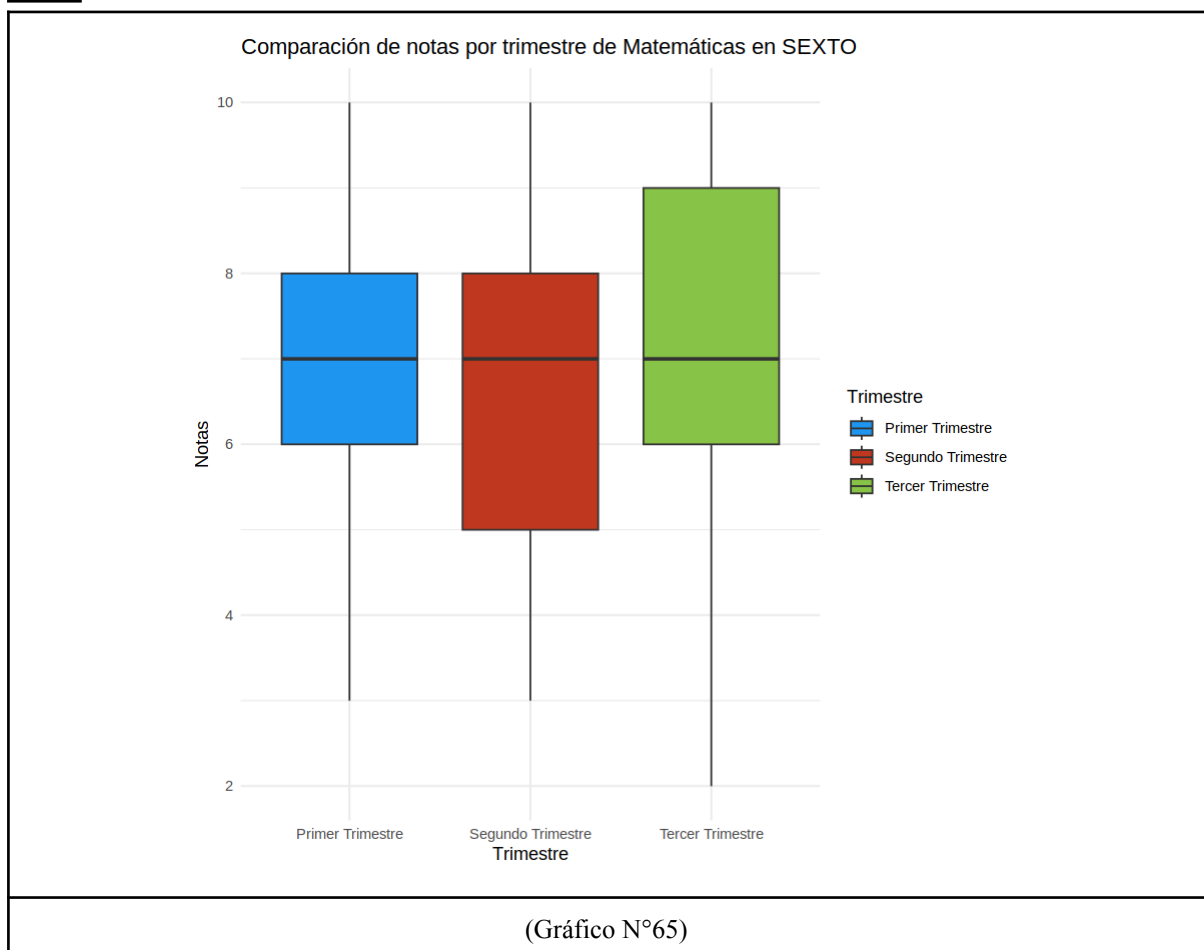
Mirando el gráfico a simple vista vemos diferencias significativas, podemos detectar un incremento del rendimiento. Avanzando con las pruebas, el test de Friedman muestra diferencias significativas entre los trimestres, con un valor de  $p < 2.2e-16$ . Sin embargo, es

importante observar que la potencia del test es muy alta(0.99) lo que nos da un resultado muy confiable. Las post-hoc detectaron diferencias significativas en la distribución de las notas entre los trimestres.

	Primer Trimestre	Segundo Trimestre
Segundo Trimestre	3.5e-11	-
Tercer Trimestre	< 2e-16	8.1e-07
Fuente de elaboración propia (Tabla N°21)		

Entonces como el primer trimestre tiene una mediana de 6 y el segundo y tercer trimestre una mediana de 7, esto sumado al resultado del test, concluimos que en quinto año, el rendimiento en matemática aumenta a medida que avanza el año.

### Sexto



Por último, sexto año. Observamos una misma mediana de 7 en los 3 trimestres, pero pareciera que el último trimestre tiene un mejor desempeño. El test de Friedman también muestra diferencias significativas entre los trimestres, con un p-value de 0.0001197. La

potencia del test es alta (1), lo que indica que el tamaño de la muestra es suficiente para detectar estas diferencias.

	Primer Trimestre	Segundo Trimestre
Segundo Trimestre	0.0490	-
Tercer Trimestre	0.0084	0.3109
Fuente de elaboración propia (Tabla N°22)		

Al realizar las comparaciones post-hoc, encontramos una diferencia significativa entre el primer trimestre y el segundo trimestre ( $p\text{-value} = 0.0490$ ), así como entre el primer trimestre y el tercer trimestre ( $p = 0.0084$ ). Sin embargo, no se encontraron diferencias significativas entre el segundo y tercer trimestre ( $p = 0.3109$ ). Esto, sumado a mirar el gráfico, nos lleva a concluir que el rendimiento de los alumnos en sexto año es más bajo en el primer trimestre y luego aumenta en el segundo trimestre, manteniéndose hasta el tercero.



## **Discusión**

Los hallazgos obtenidos en la base de datos de primaria son importantes analizarlos detenidamente porque se encontraron resultados que explican la situación que atraviesa la educación entrerriana en la actualidad.

Los datos indican que solo el 6% de los estudiantes desaprobaron el año académico. Sin embargo, la mayoría de las materias en todos los grados presentan una nota mínima de 5. Esto podría estar relacionado con la asignación numérica de la calificación ‘EPAC’.

Lo que es importante destacar, es que en la mayoría de los análisis de distribución, nos encontramos con asimetría hacia la izquierda, lo que se traduce como que hay más estudiantes que lograron un promedio por encima de la media. Solamente en Lengua de sexto grado, se encontró una leve asimetría hacia la derecha.

En el siguiente cuadro se va a mostrar los coeficientes de asimetría para las 4 materias que venimos utilizando como troncales para los grados de primaria en los distintos grados.

Grado	Matemática	Lengua	Cs. Naturales	Cs. Sociales
Primero	-0.41	-0.33	-0.64	-0.58
Segundo	-0.52	-0.41	-0.68	-0.65
Tercero	-0.36	-0.25	-0.42	-0.41
Cuarto	-0.29	-0.39	-0.5	-0.48
Quinto	-0.14	-0.21	-0.29	-0.28
Sexto	-0.03	0.02	-0.24	-0.18
Fuente de elaboración propia (Tabla N°23)				

Como se puede ver, al avanzar de grado (excepto en segundo y cuarto grado) se puede ver como la asimetría hacia la izquierda se va acercando a cero, llegando a un caso donde se invierte y da positivo. Esto lo interpretamos como que mientras más avances a nivel educativo, cada vez el conseguir mejores notas es más difícil o que se va perdiendo el interés de ir a clases.

Otro de los hallazgos más destacados es la correlación constante en los distintos grados entre Matemática ~ Lengua, solo tercer grado fue la excepción donde no se encontró una correlación moderada como mínimo. Se esperaban más correlaciones entre materias que por lo general, están correlacionadas, pero este departamento no es el caso.

Mientras que las comparaciones entre los grupos divididos por turnos siempre daban un resultado igual. Mientras se tomaban los promedios de Matemática y/o Lengua, en la mayoría de los casos, el turno mañana y tarde tenían medias y medianas similares, muy por encima de las que tenían el turno completo.

Por otro lado, las comparaciones entre los promedios de escuelas públicas y privadas, en asignaturas como Matemática y Lengua se ve una clara diferencia de performance. Durante todos los grados, los rendimientos de los estudiantes de escuelas privadas, son muy superiores a los de su par de escuelas públicas.

En el análisis bivariado, encontramos resultados que se repetían en todos los años. Las materias tienen un efecto en el desempeño final de cada estudiante en esa materia. Si bien, la fuerza de asociación no pasa de débil. Entre las distintas tablas y gráficos mostrados, se puede ver como en materias de índole más prácticas como lo son Educación Tecnológica, Educación Física y/o Artes Visuales, los porcentajes se concentran en los desempeños más altos, mientras que en desempeños bajos tienen casi nula observaciones.

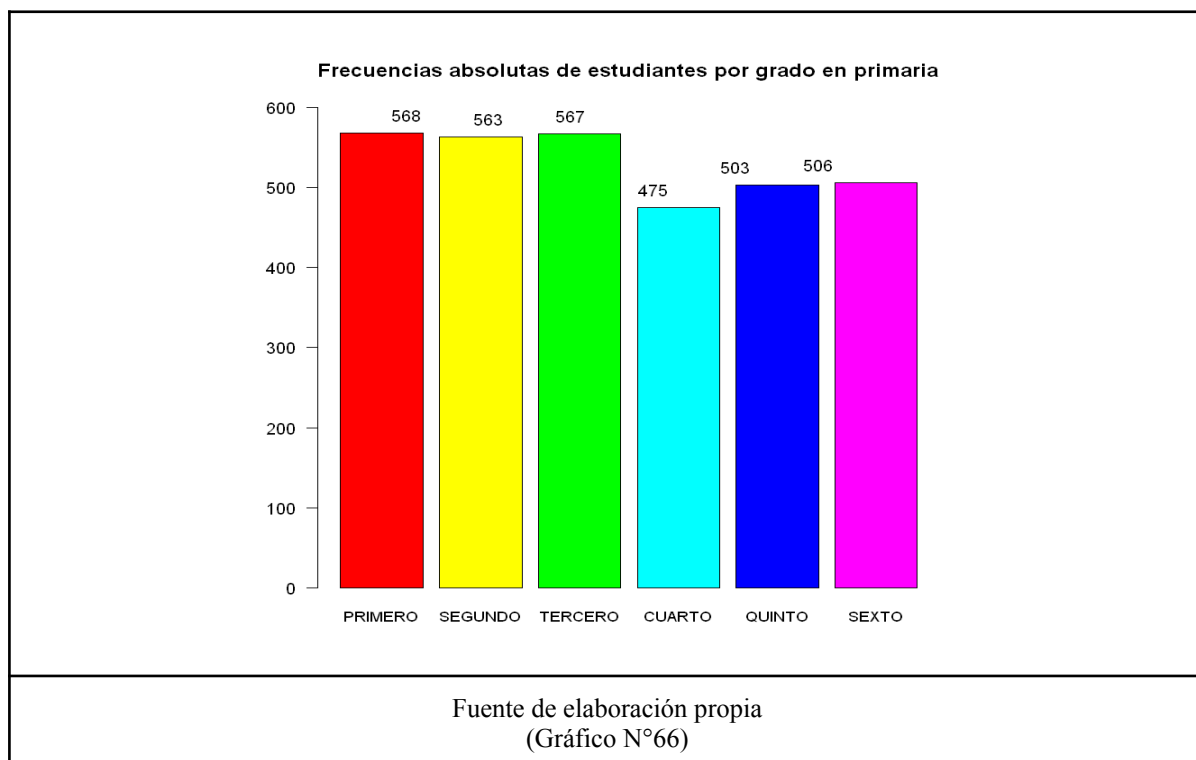
Esto es totalmente diferente en las materias que son más teóricas, como son el caso de Lengua, Matemática y en algunos casos Ciencias Sociales y Naturales. En estas áreas, los porcentajes están bien distribuidos, pero comparados a las demás materias, tienen muchas observaciones en los desempeños regulares y aprobados.

En los distintos cursos se midió la distancia de Mahalanobis, en donde los resultados rondaban siempre entre 35 y 50, o sea que tomando las variables numéricas como las notas de los tres trimestres y el promedio, algunos registros eran tomados como outliers multivariados.

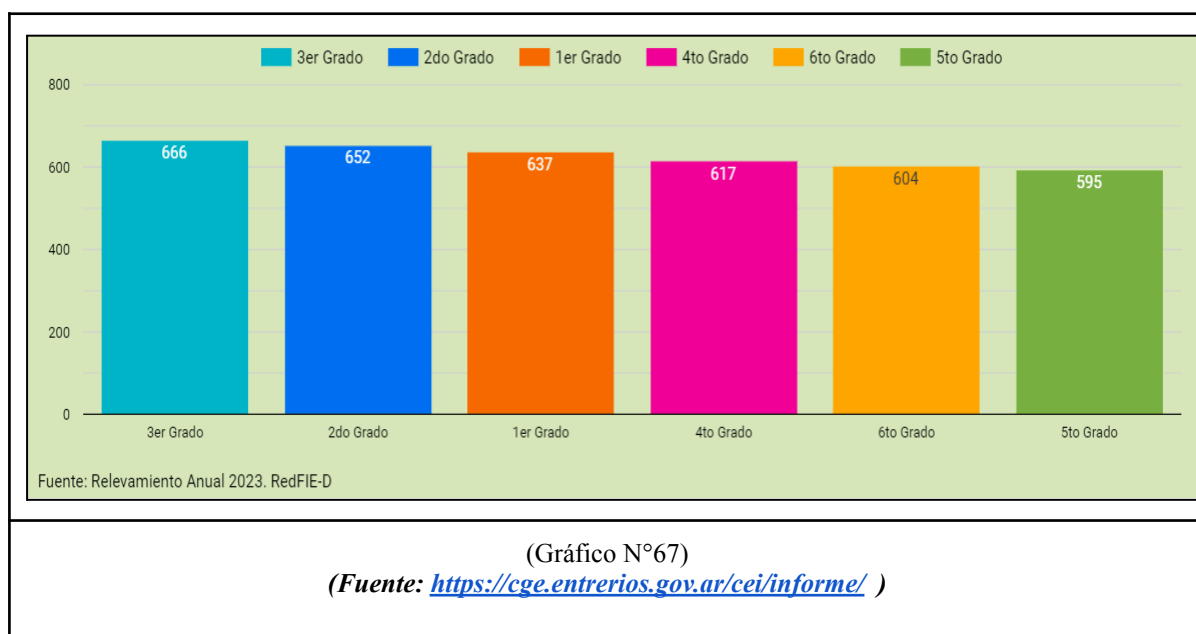
En algunos grados, se aplicaba la comparación de dos poblaciones multivariadas independientes para saber las diferencias entre vectores medios de los tipos de gestión. En otras se aplica la comparación de más de dos poblaciones multivariadas independientes, comparando los vectores medios entre los distintos turnos presentes en nuestra base de datos. Esto fue posible gracias al Teorema del Límite Central Multivariado, ya que podíamos asumir que las medias de nuestras variables provienen de una distribución normal multivariada.

Para finalizar la investigación en primaria, queríamos destacar a los mejores estudiantes. Por eso nos interesó armar los grupos de abanderados, tanto como los salientes como los entrantes.

La cantidad de matriculados por grado en primaria, como lo muestra el gráfico N°66, parece ser uniforme la distribución.



Estos datos obtenidos los podemos comparar con un gráfico proporcionado por la página web del consejo general de educación de Entre Ríos. En el gráfico se pueden observar los datos obtenidos en el relevamiento del año 2023 en el departamento Victoria.



Por otra parte, los datos obtenidos en el **nivel secundario** son más “realistas”, ya que no se puede opacar el bajo rendimiento de los estudiantes, por la manera de calificar de los

docentes. Se basan en el sistema tradicional de calificación, obteniendo una nota final entre 1 y 10, donde un 6 o superior representa un estudiante aprobado.

En el gráfico N°38 presentado anteriormente, vemos que el 78% de los alumnos tienen una condición final de “Aprobado” mientras que el restante 22% debe rendir en otra instancia de evaluación.

En la tabla de abajo se puede ver una diferencia evidente en el porcentaje de alumnos aprobados, lo que podría deberse al distinto modo de evaluar de los docentes.

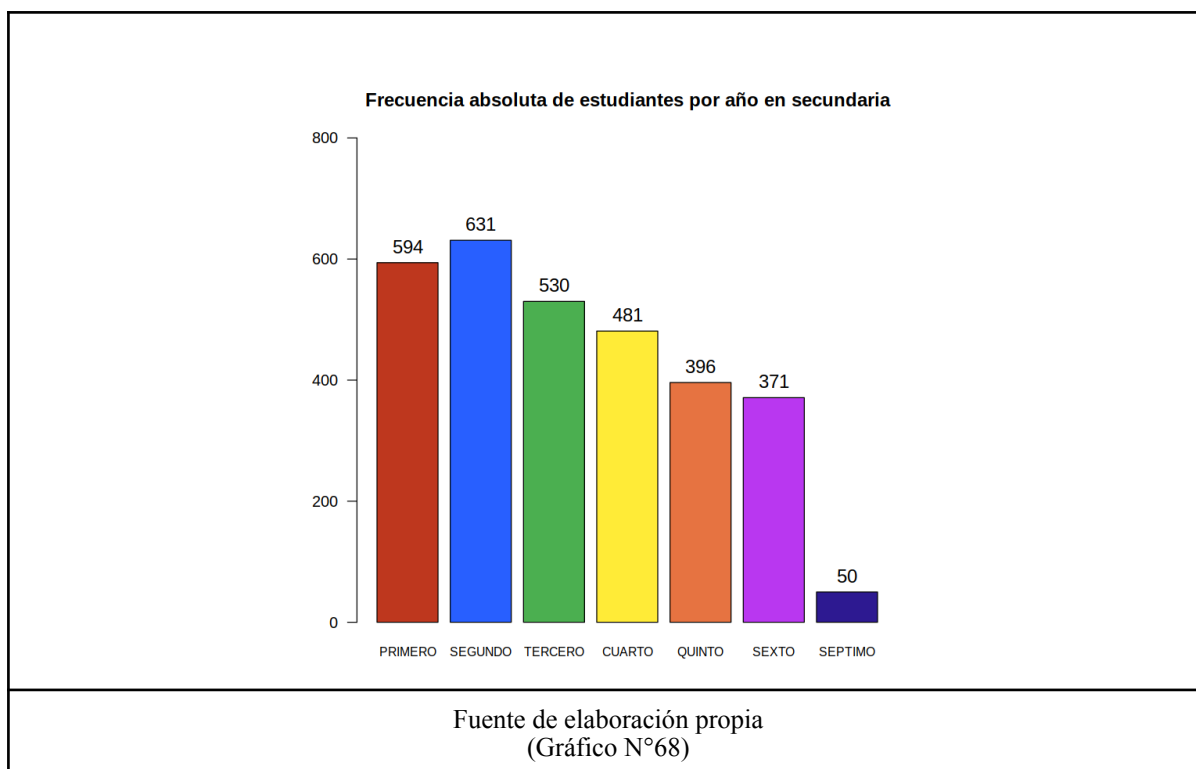
	Aprobado	Rinde
Primaria	94%	6%
Secundaria	78%	22%
Fuente de elaboración propia (Tabla N°24)		

El análisis comparativo del rendimiento entre Matemática y Educación Física reveló hallazgos que se corresponde con lo que se venía viendo en el análisis exploratorio donde comparamos los Boxplots. Los resultados obtenidos a través del test de Wilcoxon, debido a la falta de normalidad en la distribución de los datos, encontraron evidencia significativa ( $p < 2.2e-16$ ) para rechazar la hipótesis nula, lo que indica que el rendimiento en Matemática es inferior al rendimiento en Educación Física en todo el nivel secundario.

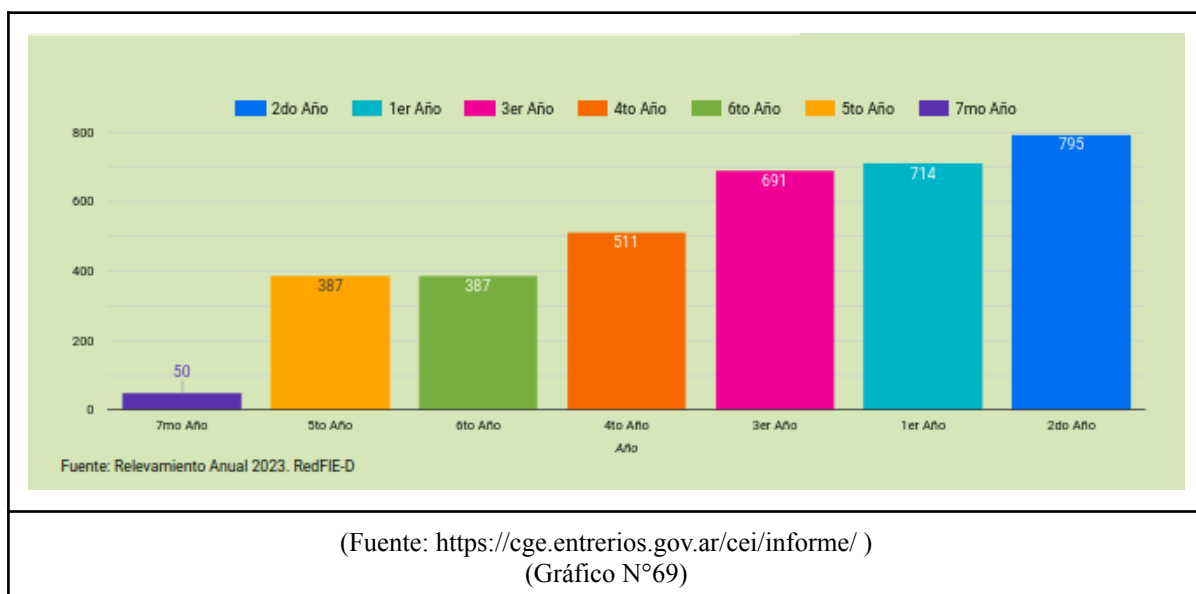
El análisis de la variación del rendimiento en Matemática durante el Ciclo Básico y el Ciclo Orientado revela diferencias significativas entre los cursos. En el Ciclo Básico, se observa una fluctuación en el rendimiento, con el Segundo año mostrando una disminución notable en comparación con Primero y Tercero, como sugieren los resultados de Kruskal-Wallis y las comparaciones post-hoc. En contraste, en el Ciclo Orientado se evidencia un aumento progresivo en el rendimiento a medida que avanzan los años, indicando una mejora constante en el rendimiento en Matemática.

El análisis de la comparación entre el tipo de gestión escolar (pública o privada) y la condición final del alumno (aprobado o rinde) revela como resultado en la prueba de Chi-cuadrado de Pearson una asociación estadísticamente significativa ( $p\text{-value} < 2.2e-16$ ), aunque el coeficiente V de Cramer de (0.121), indica que la fuerza de esta relación es débil.

Por último, podemos realizar la misma comparación que en primaria. Si recordamos el gráfico N°68 donde tenemos la cantidad total de estudiantes por año:



Estos datos obtenidos los podemos comparar nuevamente con el gráfico que nos brinda la página web del consejo general de educación de Entre Ríos. En el gráfico se pueden observar los datos obtenidos en el relevamiento del año 2023 en el departamento Victoria.



Nuevamente, como sucedió en primaria, si comparamos los datos, podemos concluir que nuestros datos están aptos para analizarlos ya que no hay tanta diferencia con los datos reales brindados por el consejo. Podemos ver que tenemos la misma tendencia, donde 2do año tiene la mayor cantidad de matrículas, seguido por 1er año, 3er año, 4to año, 5to/6to año y finalmente 7mo año.

## **Conclusiones**

Para finalizar podemos concluir que en primaria, los datos de las notas no nos deja extraer demasiada información, dificultando la obtención de resultados donde se conozcan los problemas reales del nivel primario. Sin embargo, los resultados obtenidos son significativos. Con distintas estrategias de análisis, donde se iba probando y descartando los resultados no relevantes. Creemos haber realizado un buen trabajo, describiendo la situación educativa en el departamento Victoria durante el año 2023.

Se ha demostrado a lo largo de todo el informe que los promedios de cada estudiante son mejores o peores según con qué grupo se lo compare. Ya que hay grandes diferencias entre los promedios de materias en los estudiantes de escuelas públicas y de escuelas privadas.

Así como también, esto se ve reflejado en el turno en el que asiste al colegio el matriculado. Ya que en la mayoría de los casos, el turno mañana contiene los mejores promedios.

A partir del análisis realizado de los colegios primarios se podría resumir la información más relevante en los siguientes puntos:

- El 45% de los estudiantes cursan la primaria en el turno de la mañana.
- El 81% de las escuelas en el departamento son estatales.
- Un 32% de las materias tienen un desempeño distinguido por parte de los estudiantes.
- El 94% de las materias fueron aprobadas en tiempo y forma.
- Los promedios de Lengua y Matemática están correlacionadas en 5 de los 6 grados, lo cual se interpreta como una relación académica importante.
- Se nota una tendencia a la baja de las calificaciones si empezamos de primer grado y terminamos en sexto grado. Lo que subraya la necesidad de estrategias educativas efectivas para mantener el rendimiento de los estudiantes.
- Se identifican áreas de mejora en materias como Lengua y Matemática, lo que indica la necesidad de implementar estrategias para fortalecer el desempeño en estas áreas.
- Se destaca un rendimiento positivo en Ciencias Sociales y Naturales, lo que puede ser una fortaleza a nivel educativo en el departamento. Así como también se destacan otras asignaturas como Catequesis, Informática, Artes Visuales y Educación Física.

A partir del análisis que hicimos de los colegios secundarios podríamos resumir lo que encontramos en los siguientes puntos:

**Turno y tipo de escuela:**

La mayoría de los estudiantes de secundaria cursan en el turno de la mañana, seguido por el turno tarde.

La mayoría de las escuelas secundarias siguen una modalidad de enseñanza común, con una pequeña proporción dedicada a la educación técnica y agrotécnica.

La mayoría de las escuelas secundarias en el departamento son de gestión pública, con solo una pequeña proporción de escuelas de gestión privada.

**Aprobados finales:**

El 78% de los estudiantes totales de secundaria finalizan el curso con la condición de aprobado, lo que sugiere un nivel aceptable de éxito académico. El 22% restante necesita rendir en instancias de recuperación para aprobar las asignaturas correspondientes.

**Tendencias generales del rendimiento académico:**

Educación Física es consistentemente la asignatura con el mejor rendimiento en todos los años del ciclo secundario. La mayoría de los estudiantes obtienen calificaciones altas en esta materia.

Matemática tiende a ser la asignatura con el rendimiento más bajo en la mayoría de los años, con una proporción significativa de estudiantes que tienen calificaciones más bajas.

**Distribución de calificaciones:**

Se observa una dispersión en la distribución de calificaciones en todas las asignaturas, pero especialmente en Matemáticas, donde la mediana tiende a ser más baja y hay una variabilidad más amplia en los puntajes.

Se identifican outliers en algunas asignaturas, lo que sugiere la presencia de un pequeño número de estudiantes con rendimiento excepcionalmente alto o bajo en comparación con el resto de la población (la mayoría corresponde a estudiantes con rendimiento muy inferior al resto).

**Rendimiento a lo largo de los tres trimestres:**

En el primer año, no se encontraron diferencias significativas entre los trimestres, sugiriendo un rendimiento estable a lo largo del año. Sin embargo, en el segundo año, se observó una disminución gradual en el rendimiento a medida que avanzaba el año. En tercer

año, no se detectaron diferencias significativas entre los trimestres. En cuarto año, aunque las medianas fueron similares, se encontraron diferencias significativas en el rendimiento entre los trimestres, pero esto puede deberse a variaciones en la distribución de los datos más que a diferencias en las medianas. En quinto año, se observó un aumento progresivo en el rendimiento a lo largo del año. Finalmente, en sexto año, se encontraron diferencias significativas entre los trimestres, indicando un rendimiento más bajo en el primer trimestre seguido de mejoras en los trimestres posteriores.

### **Correlaciones entre asignaturas:**

Se encuentran correlaciones significativas entre ciertas asignaturas, como la correlación positiva entre Lengua e Historia, y la correlación negativa entre Inglés e Historia.

Al calcular correlaciones parciales para descartar la influencia de otras asignaturas, se observa una disminución en la magnitud de algunas correlaciones, lo que sugiere que la influencia de otras asignaturas puede estar afectando las relaciones entre ciertas materias.

En resumen, estos resultados pueden ser importantes a la hora de saber la situación educativa a nivel primario y secundario en el departamento Victoria, ya que pueden ser utilizados como punto de referencia para la distribución correcta de recursos a las áreas a mejorar, mejorar la calidad de enseñanza y la toma de medidas educativas provinciales para el beneficio de la educación.

### **Recomendaciones**

Sería interesante también fortalecer la integración curricular y la colaboración entre docentes de las asignaturas donde se observan correlaciones significativas, permitiendo mejorar la comprensión interdisciplinaria.

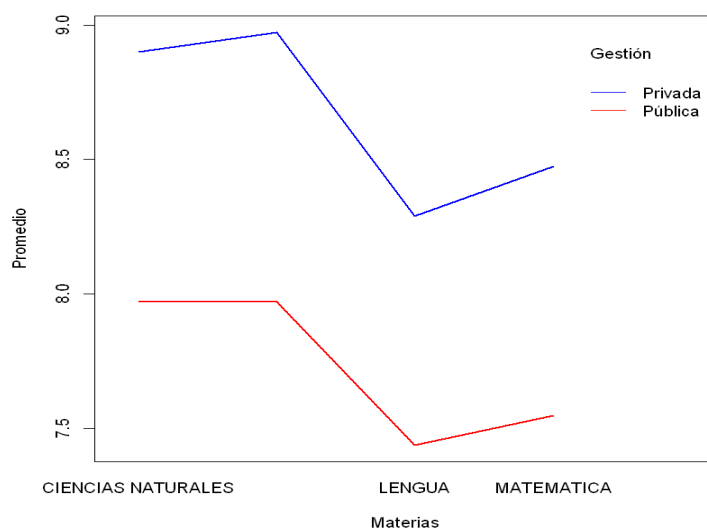


## **Referencias**

- Consejo General de Educación Gobierno de Entre Ríos. (s/f). *Información Educativa*.  
<https://cge.entrerios.gov.ar/cei/planeamiento/>
- R CHARTS. (s/f). *Paletas de colores en R*.  
<https://r-charts.com/es/paletas-colores/#discretas>

## Anexo

Para el anova de dos factores utilizamos los promedios conseguidos por los estudiantes en primer grado en las materias troncales. Llegamos a hacer un gráfico comparativo (Gráfico N°70), pero a la hora de constatar que estábamos trabajando con datos normales para seguir con el análisis, el supuesto no se cumple, por lo que el test no se siguió trabajando.



Fuente de elaboración propia  
(Gráfico N°70)

### Escuelas

<b>9000291</b>	<b>9000753</b>	<b>9000756</b>	<b>9000759</b>
<b>9001791</b>	<b>9003555</b>	<b>9003561</b>	<b>9004011</b>
<b>9004017</b>	<b>9004020</b>	<b>9005391</b>	<b>9005394</b>
<b>9005397</b>	<b>9005403</b>	<b>9005406</b>	<b>9006249</b>

Fuente de elaboración propia

(Gráfico N°71)

Efecto de las variables del gráfico N°71	
Artículo modificado	Var
Altura de la cara	CodigoUnicoEscolar
Ancho de cara	ARTES VISUALES
Estructura del rostro	CIENCIAS NATURALES
Altura de boca	CIENCIAS SOCIALES
Ancho de boca	EDUCACION FISICA
Sonriendo	EDUCACION MUSICAL
Altura de los ojos	EDUCACION TECNOLOGICA
Ancho de los ojos	LENGUA
Altura del cabello	MATEMATICA
Ancho del cabello	CodigoUnicoEscolar
Estilo de cabello	ARTES VISUALES
Altura de la nariz	CIENCIAS NATURALES
Ancho de la nariz	CIENCIAS SOCIALES
Ancho de oreja	EDUCACION FISICA
Altura de oreja	EDUCACION MUSICAL
Fuente de elaboración propia (Tabla N°25)	