



**UNIVERSIDAD NACIONAL DE ENTRE  
RÍOS FACULTAD DE INGENIERÍA**

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

**MATERIA: Exploración de Datos Multivariados**

**Nombre de la Actividad: Trabajo Final Parte 1**

**Fecha de Entrega: 25/04/2023**

**Tema: Informe Escuelas Primarias y Secundarias  
Comunes del Departamento de Diamante 2022**

**Profesoras: Fernández, Melisa**

**Alumnos: Colignon, Sabrina - Samaniego Francisco**

## Tabla de contenido

Informe Escuelas Primarias y Secundarias Comunes del Departamento de Diamante 2022. ....	2
Autoría .....	2
Resumen .....	3
Métodos.....	4
Resultados.....	5
Análisis .....	13
Conclusiones .....	31
Anexo .....	32
Imágenes: .....	32

# Informe Escuelas Primarias y Secundarias Comunes del Departamento de Diamante 2022.

## Autoría

Colignon, Sabrina

[sabrina.colignon@ingenieria.uner.edu.ar](mailto:sabrina.colignon@ingenieria.uner.edu.ar)

Samaniego, Francisco

[fmsamaniego@gmail.com](mailto:fmsamaniego@gmail.com)

# Resumen

En el siguiente informe se realizó con el objetivo de estudiar la educación primaria y secundaria en el departamento de Diamante durante el año 2022, a fin de recabar información para la implementación de programas complementarios. Se tienen en cuenta escuelas primarias y secundarias comunes, enfocándonos en materias troncales tales como Matemática, Lengua, Geografía, Historia, Biología, Física, Química, inglés, Educación Física y su respectiva transformación al nivel primario. Para esto se trabajó sobre una base de datos brindada por la docente, luego de todo el procesamiento obtuvimos los resultados haciendo foco en los promedios de las asignaturas Matemática y Lengua de los doce cursos, con este análisis pudimos concluir que el rendimiento para las escuelas diamantinas fue correcto, aunque mejorable en ciertas áreas como el primer ciclo de secundaria en Matemática. Por otra parte, obtuvimos que existe una tendencia de mejores calificaciones en el sector privado frente a la gestión pública, sumado a que existe una inclinación por parte de los estudiantes de obtener buenos resultados en ambas materias y viceversa.

# Métodos

Se utiliza como herramienta principal Jupyter Notebook y empleando las librerías readxl, tidyverse, stringr, readxl y scales para análisis univariado, y PerformanceAnalytics, nortest, corrplot, ppcor y pwr para análisis bivariado. Comenzamos cargando el dataset en un notebook with R nuevo, comprendiendo cada variable y sus alcances, teniendo en cuenta que el dataset principal no está dividido en escuelas primarias y secundarias. El tamaño del dataset es de 10475550 registros, con 419022 filas y 25 columnas. Podemos diferenciar en él variables de tipo cuantitativas tales como “anio\_lectivo”, “Curso” y “Nota” y variables cualitativas como “CUE”, “EsPrivada”, “Documento”, entre otras. Luego del análisis preliminar realizamos la limpieza y acondicionamiento de nuestra base de datos modificando el nombre de aquellas columnas que lo requieran como “anio\_lectivo” por “año”, y eliminando variables como “observaciones” y “motivo ausente” que contenían datos NA, es decir, datos perdidos, omitidos o que por alguna razón son faltantes.

Procedemos a crear variables que nos ayudan a discriminar la información como “Promedio” en la cual almacenamos el valor promedio de las notas según la materia y la variable “Curso” que contiene en una escala el número de cursos. Procedemos a normalizar las variables a mayúscula y luego arreglar aquellas que están mal escritas, como “Matematica” por “MATEMÁTICA”, entre otras. Una vez que el dataset está en condiciones lo dividimos en primaria y secundaria, cabe aclarar que la normalización de las materias se tuvo que realizar prestando atención a las materias troncales que compartían primaria y secundaria y su respectiva transformación ya que en primaria no está presente la materia “Biología” pero está “Ciencias Naturales”. Como filtro para este primer análisis utilizamos Matemática y Lengua. Aplicando los filtros correspondientes, nos quedamos con un total de 106 escuelas, las cuales 88 son primarias y 18 secundarias. La modalidad de enseñanza de todas estas es común, decidimos prescindir de las escuelas técnicas y agrotécnicas ya que eran muy pocas y no nos aportan información relevante con respecto a lo que queremos analizar en este informe. Además, nos quedamos con solo 12 columnas de las 25 originales:

- CUE
- Es Privada
- Turno
- Curso
- Período Evaluatorio
- Año (previamente llamada anio\_lectivo)
- DNI (previamente llamada Documento)
- Materia (previamente llamada asignatura)
- NotasPT (notas primer trimestre)
- NotasST (notas segundo trimestre)
- NotasTT (notas tercer trimestre)
- Promedios

# Resultados

Antes de trabajar nuestra base de datos, debemos ver si esta es representativa o no, esto se puede calcular para primaria sumando todos los registros de las matrículas de escuelas comunes, no NINA, del departamento de Diamante, que nos da un resultado de 5396. A este resultado lo trasladamos a una calculadora de representatividad y utilizando como tamaño de la población 5396, nivel de confianza de 95% y un error del 5%, nuestra muestra debería ser mayor a 359. Como muestra de primaria supera dicho valor, es representativa.

Ahora bien, en el dataset de secundaria, las matrículas están cargadas por curso, por ende, debemos cargar cada curso, sumando las matrículas de cada uno, pasar dichos valores a la calculadora utilizando los mismos valores de nivel de confianza y error, y ver si son representativos. Para una mejor visualización de los datos por curso y su representatividad, consideramos que una tabla permite comparar de forma más directa la información obtenida.

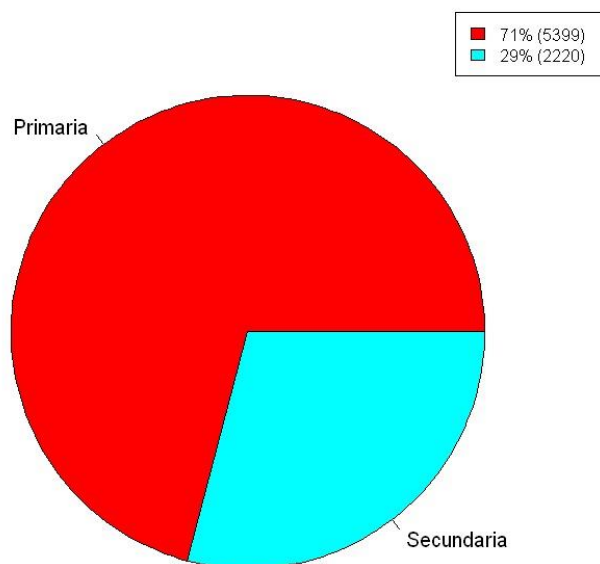
Curso	1RO	2DO	3RO	4TO	5TO	6TO
MATRICULA REAL	1125	1070	947	838	700	680
MUESTRA SUGERIDA	287	283	274	264	249	246
MATRICULA CARGADA EN NOTA	474	415	340	356	299	331
ES REPRESENTATIVO?	SI	SI	SI	SI	SI	SI

*Tabla 1- Representatividad escuelas secundarias del departamento de Diamante*

Con un total de 7619 de estudiantes en primaria y secundaria, tenemos un 70,86% en primaria (5399 alumnos) y un 29,14% en secundaria (2220), por ende, la distribución en primaria es mucho mayor a la de secundaria. Es esperable esta diferencia ya que durante la secundaria existe una mayor probabilidad de abandono, dicha diferencia la podemos ver con claridad en la imagen número 1.

Teniendo en cuenta los registros, tenemos un total de 72252, distribuidos en un 37% en secundaria y un 63% en primaria. Podemos ver que los porcentajes entre cantidad de registros y cantidad de alumnos (DNI) no son similares, por ende, para un mejor análisis estadístico, continuaremos usando la cantidad de alumnos.

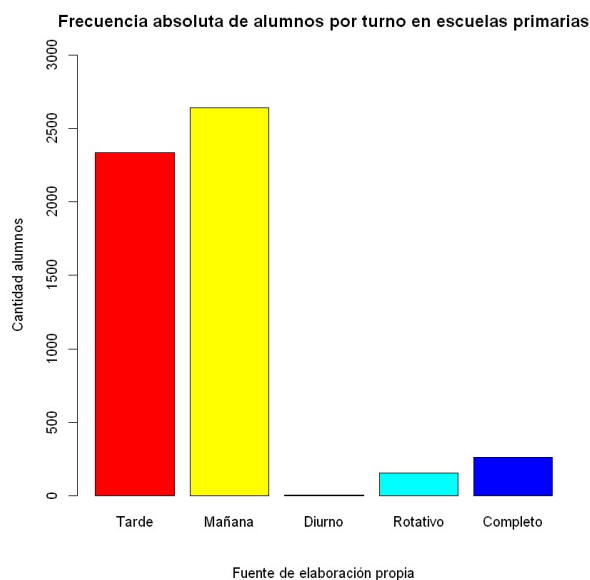
### Distribución de los alumnos por Nivel de Enseñanza



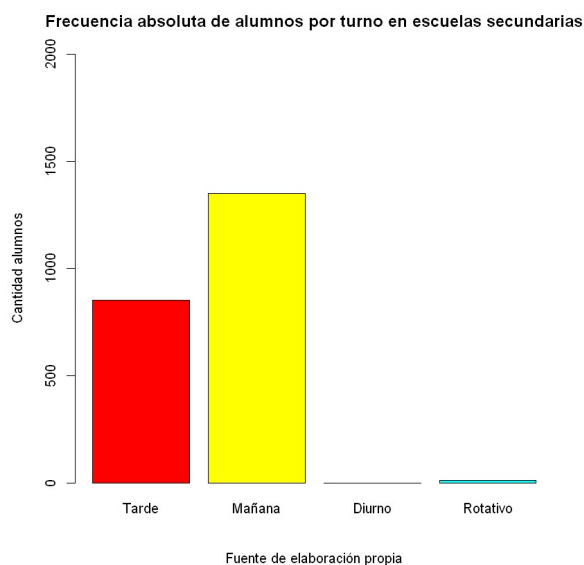
Fuente de elaboración propia

*Ilustración 1- Distribución de los alumnos por Nivel de Enseñanza en el departamento de Diamante*

Realizando un análisis particular para escuelas primarias y secundarias en el año 2022, separando por turnos, dentro de primaria, ilustración 2, encontramos que el turno “completo” representa un 4.51%, “diurno” un 0.09%, “mañana” un 49.16%, “rotativo” 3.16% y por último “tarde” un 43.08%. En cambio, en secundaria, ilustración 3, encontramos que el turno “diurno” un 0.59%, “mañana” un 56.23%, “rotativo” 1.12% y por último “tarde” un 42.06%. En ambos casos los turnos que contienen mayor cantidad de alumnos son los de “mañana” y “tarde”, así mismo, los turnos menos solicitados son “diurno” y “rotativo” en ambos casos.



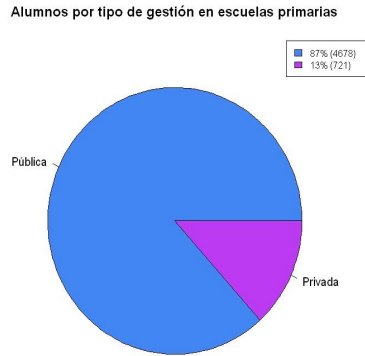
*Ilustración 2- Distribución de los alumnos de primaria por turno en el departamento de Diamante*



*Ilustración 3- Distribución de los alumnos de secundaria por turno en el departamento de Diamante*

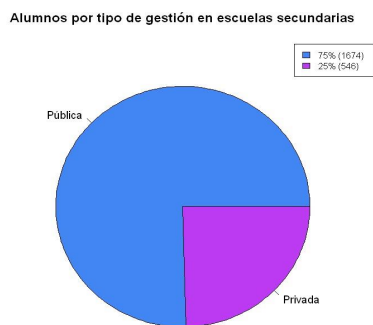
Segmentando la distribución en sector público y privado en el año 2022, en ambos casos la gestión pública es mucho mayor que la gestión primaria, siendo en primaria del 86.65% y en secundaria del 89.4%, en las ilustraciones 4 y 5 podemos ver los gráficos de torta correspondientes a primaria y secundaria respectivamente.





Fuente de elaboración propia

*Ilustración 4- Distribución de los alumnos de primaria por tipo de gestión en el departamento de Diamante*

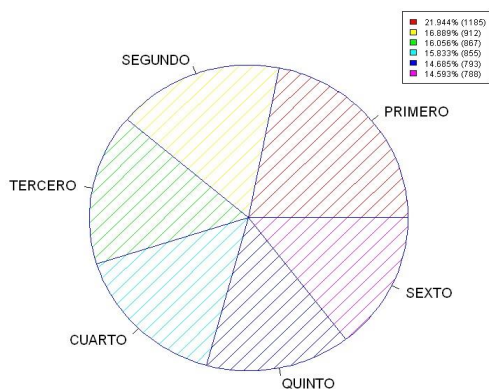


Fuente de elaboración propia

*Ilustración 5- Distribución de los alumnos de secundaria por tipo de gestión en el departamento de Diamante*

Segmentando la distribución por cantidad de estudiantes por año de cursado en el 2022, en primaria podemos ver una distribución casi uniforme de alumnos en cada grado, como podemos ver en la ilustración 6, de igual forma en secundaria, la distribución también es casi uniforme, visible en la ilustración 7.

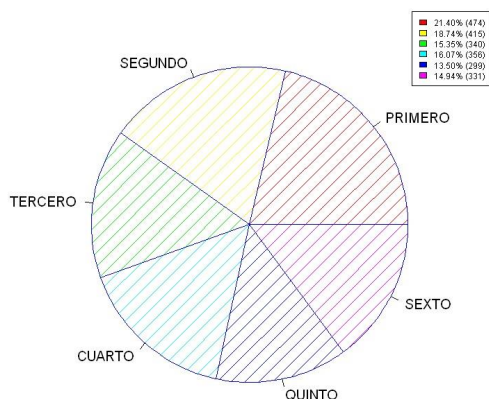
Distribución de los alumnos por curso en escuelas primarias



Fuente de elaboración propia

*Ilustración 6- Distribución de los alumnos de primaria por curso en el departamento de Diamante*

Distribución de los alumnos por curso en escuelas secundarias



Fuente de elaboración propia

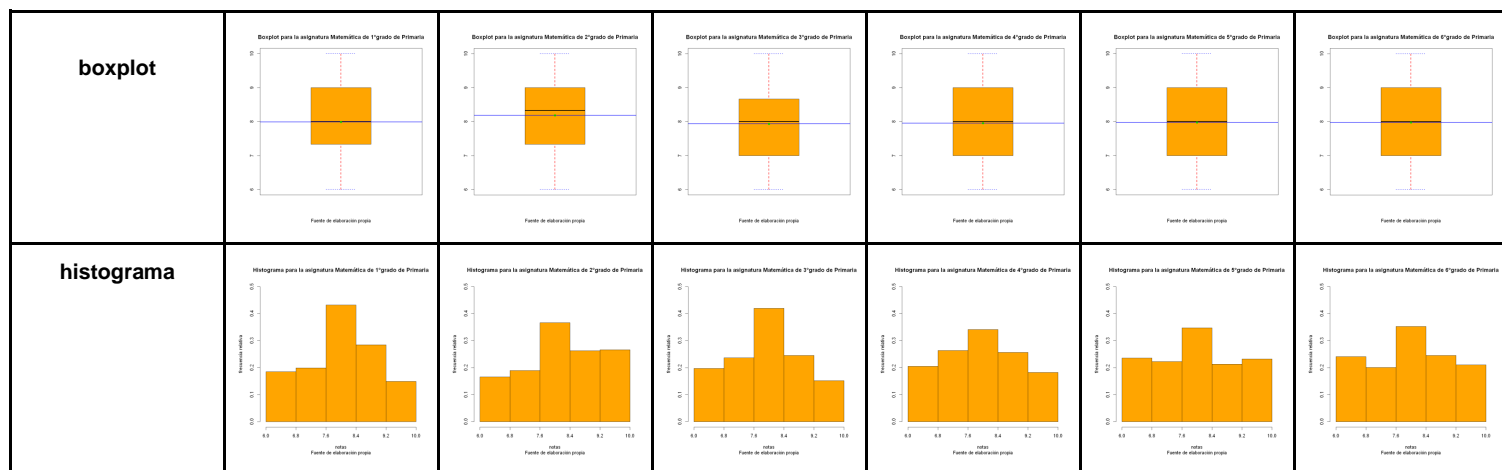
*Ilustración 7- Distribución de los alumnos de secundaria por tipo de gestión en el departamento de Diamante*

Como hemos visto hasta el momento, las distribuciones entre primaria y secundaria son similares, tanto en la distribución de alumnos por turno, gestión y curso.

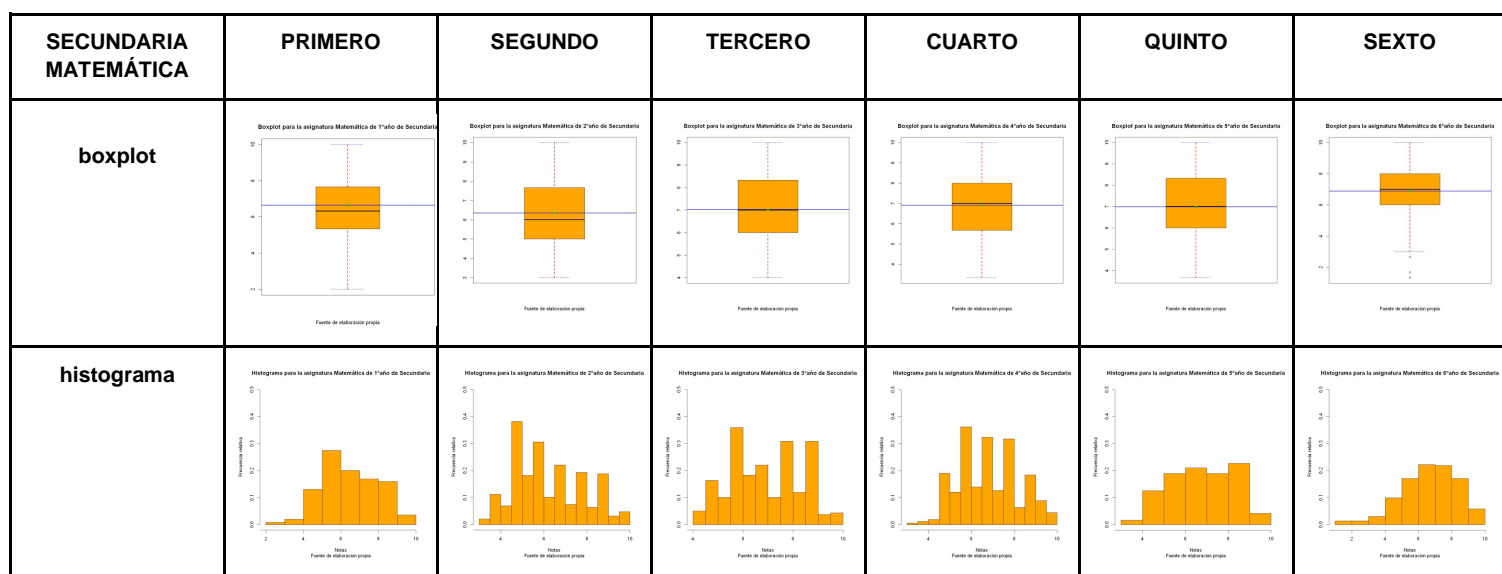
Partiendo de la base de datos de primaria y secundaria, realizaremos un análisis para las asignaturas de matemática y lengua durante el año 2022. Para poder visualizar los datos de mejor manera, realizaremos una tabla en la cual discriminamos por curso, según la materia, para comparar cada gráfico.

Resultados obtenidos para la asignatura de matemática:

PRIMARIA MATEMÁTICA	PRIMERO	SEGUNDO	TERCERO	CUARTO	QUINTO	SEXTO
------------------------	---------	---------	---------	--------	--------	-------

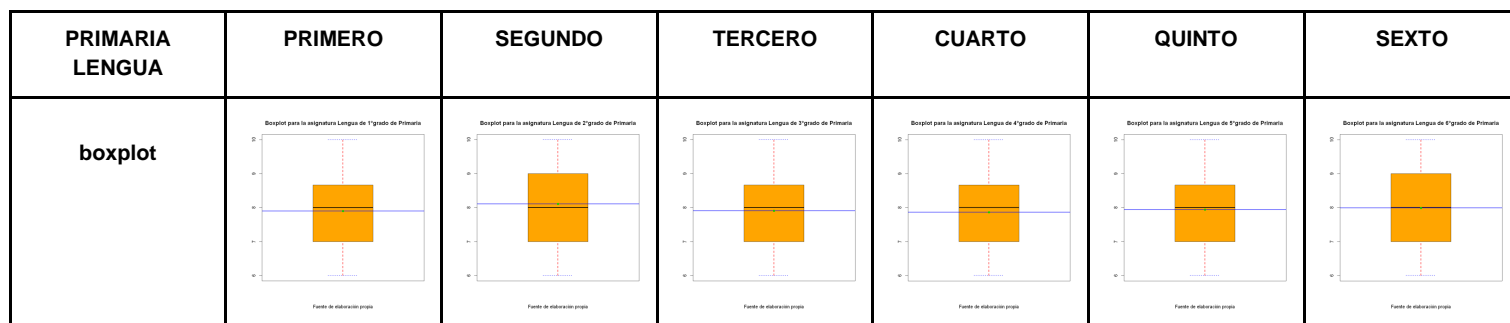


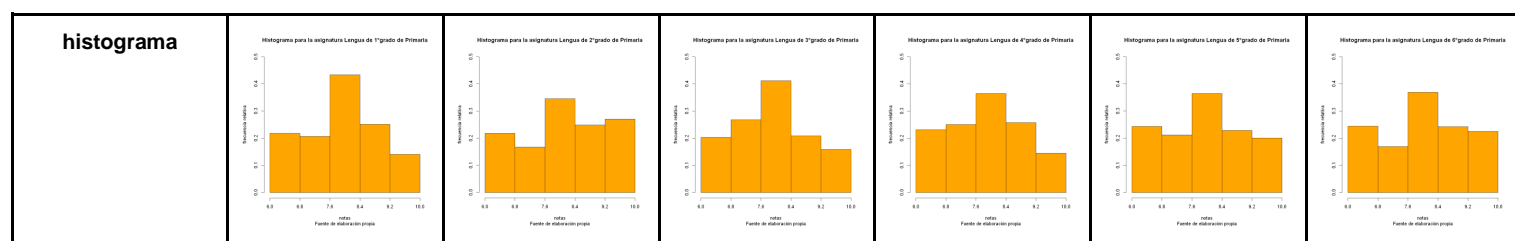
*Tabla 2- Boxplots e histogramas obtenidos para la asignatura de matemática de las escuelas de primarias del departamento de Diamante*



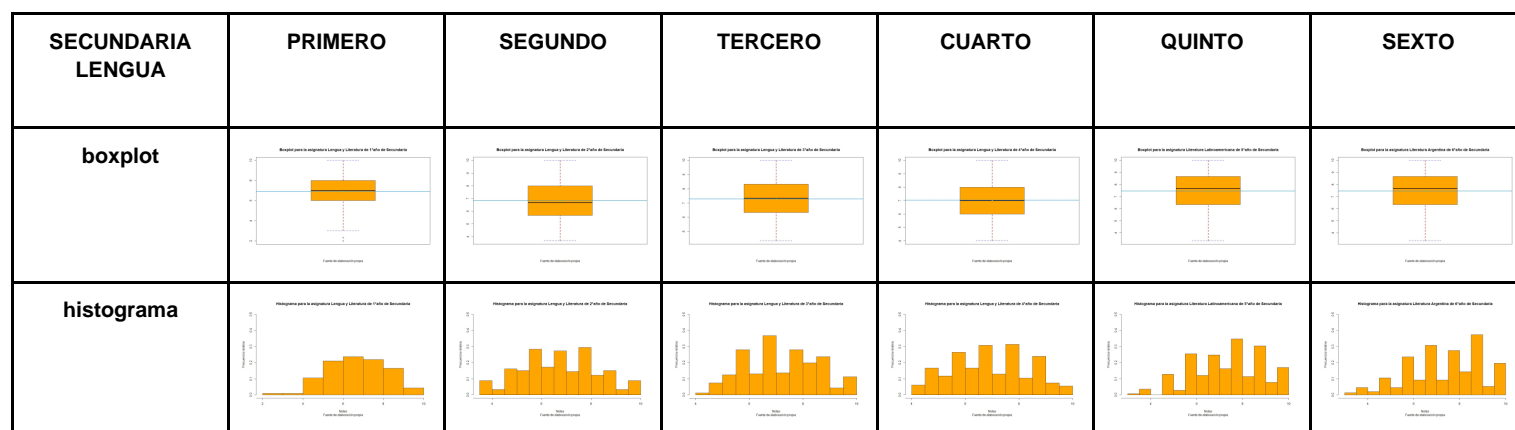
*Tabla 3- Boxplots e histogramas obtenidos para la asignatura de matemática de las escuelas de secundarias del departamento de Diamante*

Resultados obtenidos para la asignatura de lengua:





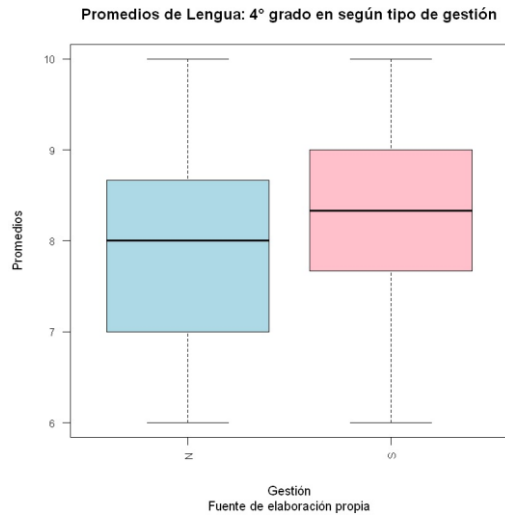
*Tabla 4- Boxplots e histogramas obtenidos para la asignatura de lengua de las escuelas de primarias del departamento de Diamante*



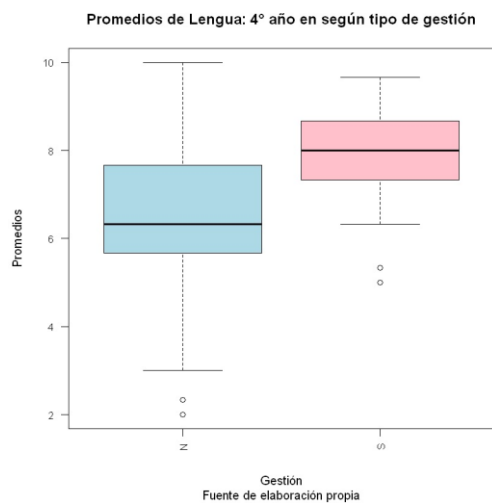
*Tabla 5- Boxplots e histogramas obtenidos para la asignatura de lengua de las escuelas de secundarias del departamento de Diamante*

Ahora tenemos la oportunidad de avanzar en nuestro análisis y abordar cuestiones más interesantes para la investigación. Esto nos permitirá encontrar explicaciones, predecir el comportamiento de algunas características basándonos en el conocimiento de otras y contrastar hipótesis que relacionan mis dos variables. En resumen, en esta etapa de la investigación nos proponemos analizar la relación entre dos variables para un mismo grupo de individuos utilizando correlación ya que nuestras variables son de tipo numéricas.

Para analizar la correlación, optamos por gráficos que nos permitirán de forma sencilla y fácil interpretar y demostrar la existencia de una relación o no entre ambas variables. A continuación, el gráfico de tipo boxplot obtenido para el análisis de correlación de cuarto grado y año para la asignatura de lengua de primaria según el tipo de gestión, público o privado, para el 2022. En el primer gráfico, ilustración número 8, encontramos los promedios para cuarto grado de primaria para lengua, podemos ver en la caja rosada cómo se comportan los datos para la gestión privada y en celeste para la gestión pública. En contraposición encontramos el gráfico para cuarto año de secundaria, ilustración número 9, para la asignatura de lengua, guiándonos por los colores anteriores, podemos ver en rosado el boxplot correspondiente para la gestión privada y en celeste para la gestión pública.



*Ilustración 8- Boxplot de promedios de lengua de 4to grado según tipo de gestión*



*Ilustración 9- Boxplot de promedios de lengua de 4to año según tipo de gestión*

Nuestro objetivo es investigar si al cambiar el valor de una variable, el valor de otra variable también cambia, es decir, estamos interesados en la variación conjunta de dos variables numéricas. Dos variables numéricas están correlacionadas si se observa una modificación en los valores de una variable al aumentar o disminuir los valores de la otra variable.

# Análisis

Comenzaremos analizando la representatividad de la base de datos, para afirmar o no dicho caso, una muestra es representativa siempre y cuando posea las mismas propiedades que la población a la que representa, en nuestro caso la representatividad de la base de datos de Diamante debe ser representativa de la base de datos primaria, la cual no ha sido dividida por departamentos. Teniendo en cuenta el valor que logramos obtener, lo comparamos y llegamos a la conclusión de que tanto las muestras de las escuelas de primaria y secundaria de Diamante son representativas de la base de datos primaria.

Otro elemento a tener en cuenta es la cantidad de estudiantes que hay en las escuelas del departamento. En nuestro caso, al ser la distribución de alumnos en primaria mucho mayor que en secundaria, podemos evaluar aquellos factores que están llevando a que los alumnos abandonen sus estudios antes o en el transcurso de la secundaria.

La segunda parte de nuestro análisis se basa en la diferencia de las distribuciones entre las escuelas primarias y secundarias de Diamante durante el año 2022. Comenzamos analizando los turnos a los que asisten los estudiantes, en los gráficos 2 y 3, podemos ver que los turnos más solicitados son los de mañana y tarde tanto de escuelas primarias como de secundarias, los turnos completo, diurno y rotativo tienen muy poca cantidad de alumnos. En la distribución de tipo de gestión en las escuelas es donde podemos ver por primera vez una diferencia entre las escuelas de primaria y secundaria, la gestión pública de las escuelas primarias es mucho mayor que la privada, un 87% vs un 13%, mientras que, en secundaria, dicha diferencia se disminuye y la gestión privada representa un cuarto dentro del gráfico de gestión. La distribución de alumnos por curso dentro de las escuelas primarias y secundarias es casi uniforme, en ambos casos podemos ver que el curso con más alumnos es primero y el con menos estudiantes es sexto en la primaria y quinto en la secundaria, pero más allá de esta diferencia, si ponemos los gráficos de forma superpuesta, apenas son visibles las diferencias gráficas, mientras que la mayor diferencia se encuentra en la cantidad de alumnos. En ambos casos primero es el curso con mayor cantidad de alumnos, siendo que en primaria son alrededor de mil alumnos y en secundaria no llegan a los 500. Hablando en proporciones, los gráficos son similares, pero no debemos olvidarnos que la cantidad de alumnos en este análisis es un factor determinante. En resumen, para el departamento que estamos estudiando podemos destacar que la mayoría de familias prefiere el sector público para la educación de sus hijos, dejando al sector privado con menor cantidad. Podemos resaltar que en nuestro caso particular disponemos de la información de 16 escuelas secundarias públicas y solo 2 privadas, mientras que para primaria la diferencia es bastante más notoria ya que hay registro de 85 primarias públicas y 3 privadas. Consideramos este un factor importante a destacar ya que, a la hora de implementar programas para mejorar los rendimientos, es el Estado el que debería encargarse de aplicarlos.

Encaminándonos en el estudio particular de las asignaturas de matemática y lengua de ambos datasets durante el año 2022, decidimos realizar dos tipos de gráficos para la visualización de los datos. Primero optamos por un boxplot o diagrama de cajas, en el cual podemos comparar la distribución en conjuntos de datos numéricos, organizándolos dentro de una “caja”, que nos proporciona información sobre la mediana, rangos y valores extremos, lo que nos da una rápida comprensión de la distribución del conjunto

de esos datos. Como este tipo de gráfico no es bueno para el análisis detallado de los datos, realizamos además un histograma para poder obtener información sobre la forma y simetría de su distribución.

La principal diferencia dentro de este análisis se radica en el desempeño académico de los alumnos, particularmente en primaria los alumnos no pueden sacar notas menores a 6, por ende, los rangos de notas no son tantos ni tan variados, mientras que la escala de notas de secundaria va de 1 a 10, es decir, hay más variabilidad dentro de los resultados posibles dentro de las asignaturas.

Prestando particular énfasis en lo descrito, comenzaremos con las escuelas primarias. Como vemos en la tabla 3 y 4, podemos ver tanto en los boxplots como en los histogramas de matemática y lengua que las distribuciones son muy similares. Nos enfocaremos en análisis en dos ciclos, el primer ciclo comprenderá los cursos de primero, segundo y tercero, mientras que el segundo ciclo comprenderá los cursos de cuarto, quinto y sexto. Dicha distribución la utilizaremos para las asignaturas de matemática y lengua de tanto primaria como secundaria. En la parte superior de la tabla encontramos los boxplots y en la inferior el histograma correspondiente.

Para el análisis del primer ciclo dentro de la asignatura de matemática, antes de graficar realizamos una tabla de frecuencia utilizando la asignatura matemática y el promedio de las notas. Para todos los cursos de primaria optamos por utilizar cinco rangos de notas que abarquen todos los datos, estos son: 6, 6.8, 7.6, 8.4, 9.2 y 10.

Luego, comenzaremos analizando los elementos del gráfico:

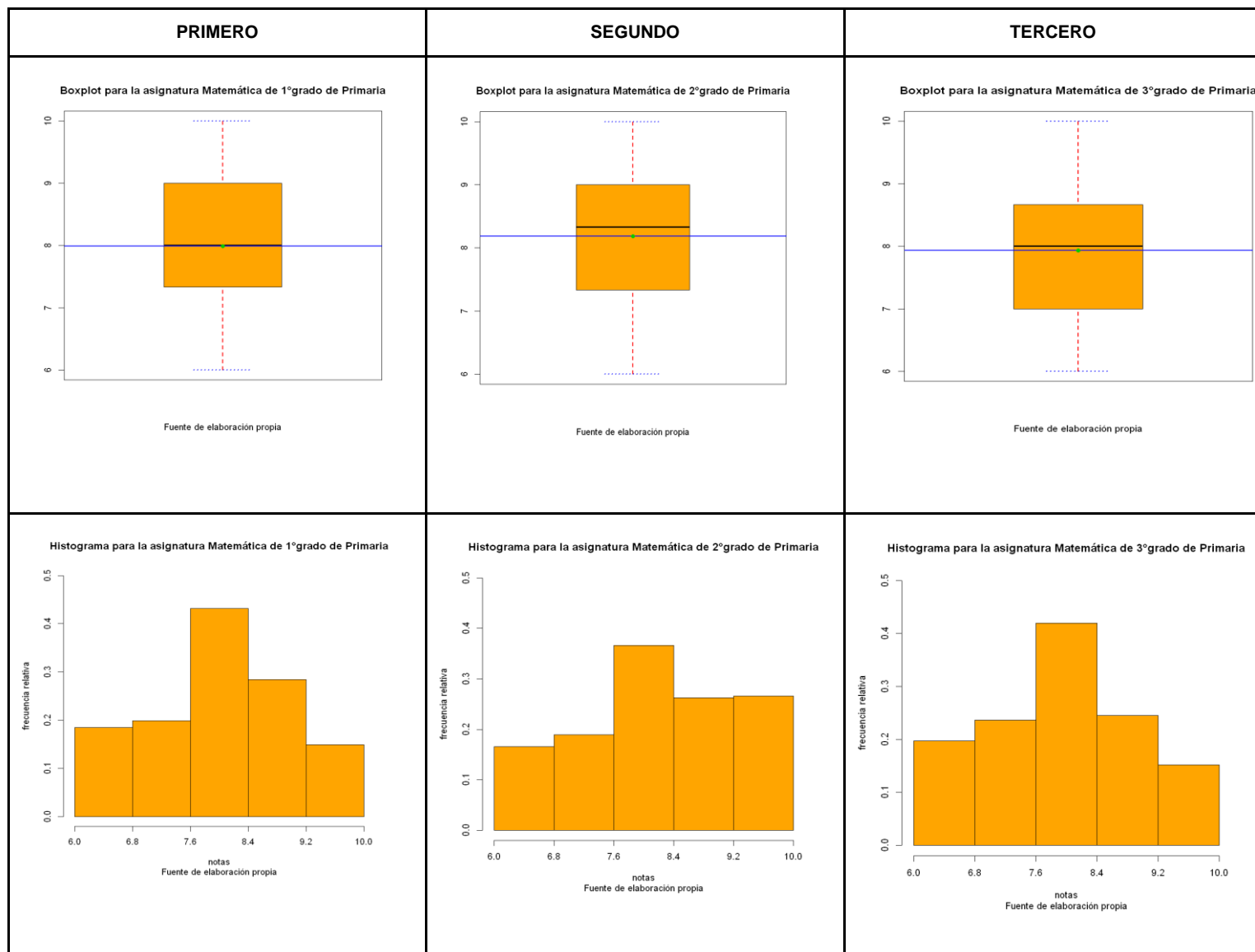
- Línea punteada roja: representa la dispersión de los datos junto a la línea punteada azul que representa el límite superior (25% de los datos) e inferior (25% de los datos) del bigote.
- Caja naranja: representa el rango intercuartil (IQR), que es la distancia entre el primer cuartil (Q1) y el tercer cuartil (Q3). La caja se extiende desde el primer cuartil hasta el tercer cuartil y abarca el 50% de los datos. La línea dentro del rectángulo representa la mediana, que divide los datos en dos partes iguales.
- Línea completa azul y punto verde: representa la media.

En cada boxplot podemos ver que el rango de los valores de la izquierda corresponde a los rangos de notas elegidos, particularmente podemos ver que la nota más alta para todo el ciclo es de 10 y la más baja es de 6, siendo la mediana de aproximadamente 8 para los casos de primero y tercero, mientras que en segundo no llega a superar el 8.5. También podemos ver cómo se distribuye la caja en cada grado, siendo que en primero y segundo el 50% de los datos está abarcado entre las notas de 7.5 a 9 aproximadamente y en tercero varía a los rangos de 7 a 8.7 aproximadamente. En los tres casos vemos que la distribución de los datos es simétrica, centrada en el valor 8, y que no contamos con valores atípicos.

Lo más interesante a analizar es cómo se comporta la media en cada caso, podemos ver que en primer grado la media y la mediana están superpuestas, mientras que en segundo y tercer grado la mediana es superior a la media, en este último caso la diferencia entre ambas es menor. Podemos concluir que la media es representativa de los datos.

Pasando a los histogramas, en el eje vertical encontramos las frecuencias relativas, es decir, las proporciones del total para cada rango de notas y en el eje horizontal los intervalos de nuestra variable de estudio. Vemos que los datos son más variados, en los tres casos son unimodales, con dicho valor en el rango de 7.6 a 8.4. En el caso de

primero, vemos que los datos en segundo lugar están más concentrados en el rango de 8.4 a 9.2, y los otros rangos tienen una distribución casi uniforme. En segundo grado podemos ver que los rangos de 8.4 a 9.2 y de 9.2 a 10, tienen casi la misma altura y por debajo de ellos encontramos a los rangos de 7.6 a 6. En el tercer grado ya vemos una distribución más simétrica de los datos, y el rango que posee menor cantidad de notas es el de 9.2 a 10. Dicho análisis está basado en los gráficos de la tabla número 6.



*Tabla 6- Boxplots e histogramas obtenidos para el primer ciclo de la asignatura de matemática de las escuelas de primarias*

Los valores estadísticos correspondientes a los gráficos del primer ciclo en la asignatura de Matemática en el año 2022 son los siguientes:

Estadístico	1er grado	2do grado	3er grado
Media	8	8.18	7.93
Mediana	8	8.33	8



Primer Cuartil	7.33	7.33	7
Tercer Cuartil	9	9	8.66
Mínimo / Máximo	6/10	6 / 10	6 /10

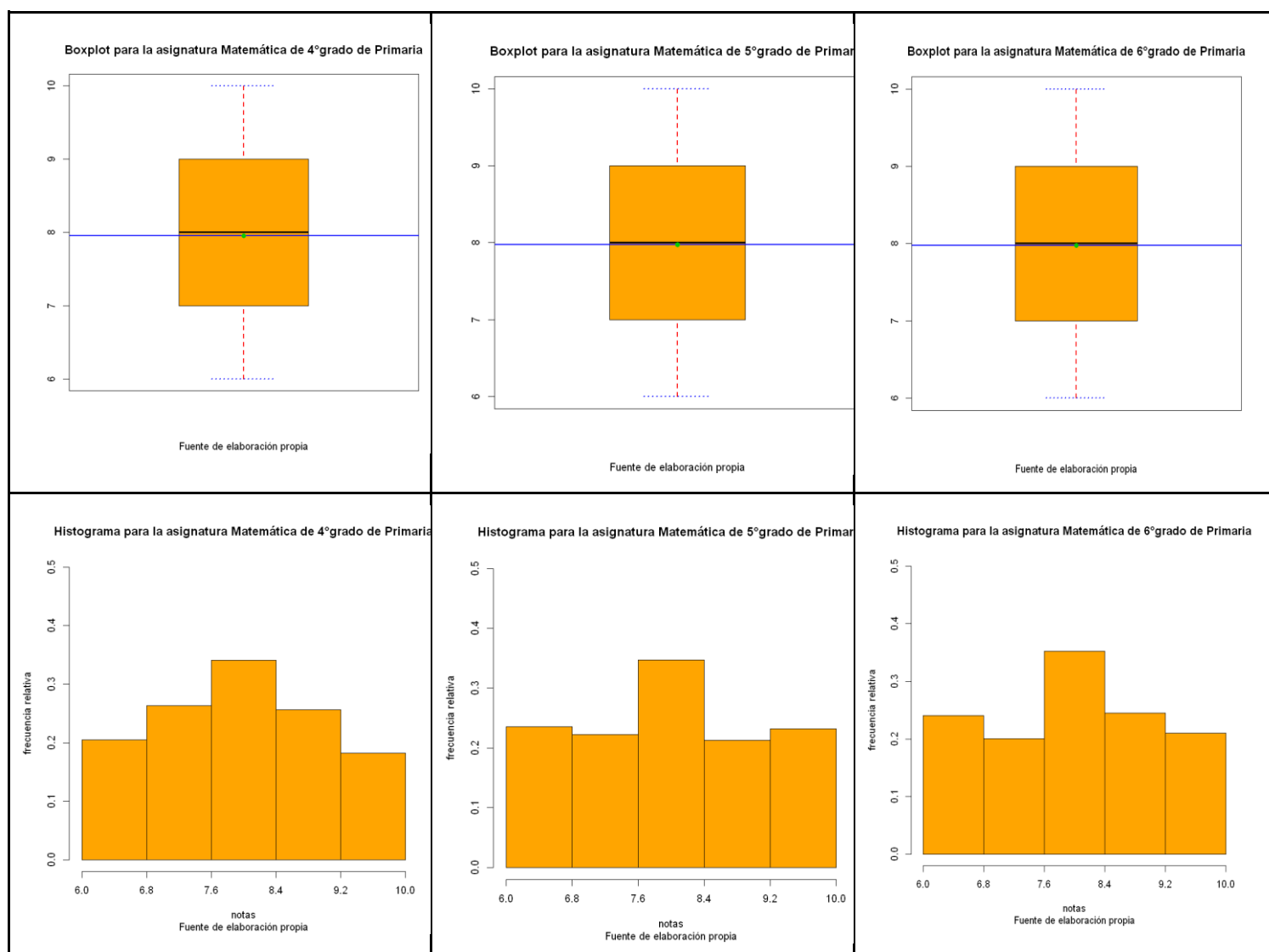
*Tabla 7- Resultados estadísticos obtenidos para el primer ciclo de la asignatura de matemática de las escuelas de primarias*

Para este primer análisis, podemos ver principalmente la poca variabilidad de los datos. Correspondiendo a lo antes descrito, el desempeño académico de los estudiantes dentro del primer ciclo varía muy poco, los rangos máximos y mínimos siempre son los mismos, esto se debe a que los estudiantes no pueden sacar notas menores a 6. Podemos diferenciar que el desempeño académico en el tercer año desciende un 3% con respecto a segundo año, mientras que en segundo año crece un 2,3% respecto del primer año. Podemos corroborar estos cambios reflejados en cómo se comporta la media contra la mediana en cada boxplot, en los cuales en primer año están alineadas y en segundo y tercer año la línea de la media es menor a la mediana.

Para el análisis del segundo ciclo de matemática de primaria en el año 2022 replicamos la misma tabla, encontrándonos con los gráficos de la tabla 8. En ellos podemos ver cómo cambia la distribución de los datos de un ciclo a otro, vemos en los boxplots que la mediana y la media casi siempre coinciden, podemos afirmar que la media es representativa de los datos, logrando capturar el 50% de los valores dentro de los rangos de 7 a 9 y que los límites superior e inferior son 10 y 6, respectivamente. En los tres casos vemos que la distribución de los datos es simétrica, centrada en el valor 8, y que no contamos con valores atípicos.

Siguiendo con el análisis de los histogramas, en los tres grados tenemos una sola moda, ubicada en el rango de 7.6 a 8.4. Particularmente en cuarto grado tenemos una distribución casi simétrica, que podemos visualmente decir que es normal, donde los dos rangos más completos son de 6.8 a 7.6 y de 8.4 a 9.2. En quinto grado vemos como se destaca el rango modal y el resto es casi uniforme, en cambio en sexto grado volvemos a ver un gráfico similar al de cuarto, en el que destaca el rango modal y los rangos más completos son de 6 a 6.8 y de 8.4 a 9.2, las dos barras restantes tienen una altura casi igual.

CUARTO	QUINTO	SEXTO
--------	--------	-------



*Tabla 8- Boxplots e histogramas obtenidos para el segundo ciclo de la asignatura de matemática de las escuelas de primarias*

Estadístico	4to grado	5to grado	6to grado
Media	7.95	7.97	7.97
Mediana	8	8	8
Primer Cuartil	7	7	7
Tercer Cuartil	9	9	9
Mínimo / Máximo	6/10	6 / 10	6 / 10

*Tabla 9- Valores estadísticos obtenidos para el segundo ciclo de la asignatura de matemática de las escuelas de primarias*

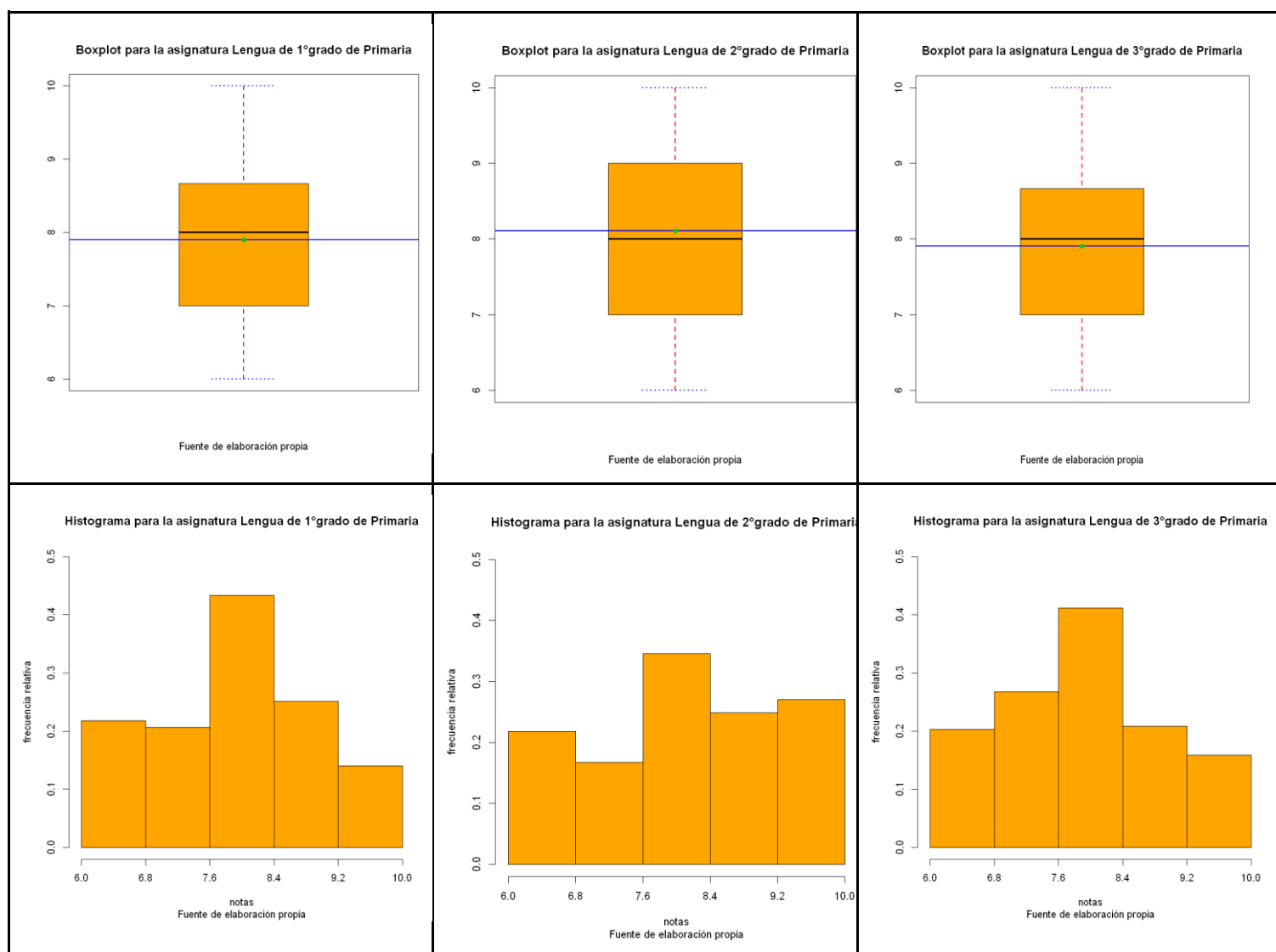
Para el segundo ciclo de Matemática, lo que más destaca es la casi nula variabilidad de los datos, el único dato que cambia con respecto a los anteriores es la media de cuarto año que desciende un 0,1% con respecto de los otros dos años, esto explica por qué

los boxplots son prácticamente idénticos. Es en este punto donde nos conviene analizar los gráficos para evaluar las frecuencias de notas de cada año. Podemos concluir que, al tener estadísticas tan similares, no podemos inferir cual año tuvo mejor desempeño. Luego de este exhaustivo análisis por los dos ciclos, podemos concluir que el rendimiento de la asignatura fue óptimo, obteniendo la mayoría de los datos dentro del rango de 7.6 a 8.4, con histogramas unimodales y boxplots representativos de la distribución de las notas.

Habiendo analizado el desempeño en la asignatura de matemática, es hora de analizar lo mismo, pero en la asignatura de Lengua para los estudiantes de primaria en el año 2022.

Comenzamos con la tabla correspondiente al primer ciclo, como en el análisis anterior, partiremos del boxplot y luego analizaremos los histogramas. En los tres gráficos podemos ver que el valor máximo es de 10 y el mínimo es de 6, que la mediana está alineada hacia el 8, en los tres casos la media está muy cerca de ella, podemos afirmar que es representativa, y no contamos con valores atípicos. Particularmente en primer grado podemos ver que hay mayor distribución de los datos entre el 7 y 8 que de 8 a 9, por ende, la caja no es simétrica. En segundo grado podemos ver que el 50% está comprendido entre el 7 y 9, lo que hace que la caja sea más simétrica, en este caso en particular la media es mayor a la mediana. Como tercer caso, podemos ver que el gráfico es muy similar al de primer grado, donde los datos están más concentrados en el intervalo de 7 a 8 que de 8 a 9, y eso hace que la caja se vea desplazada, por último, la mediana es mayor a la media, cosa que también se repite en el boxplot de primer grado. Ahora bien, si observamos los histogramas del primer ciclo, podemos ver que todos son unimodales, con la moda ubicada en el rango de 7.6 a 8.4, donde la nota más alta es de 10 y la más baja es de 6. Centrándonos en el histograma de primer grado, vemos que las barras comprendidas en los rangos de 6 a 7.4 tienen casi la misma altura, la barra del intervalo siguiente al modal es la segunda más alta y que la barra más baja le corresponde al intervalo de 9.2 a 10. En segundo grado podemos ver que los datos están distribuidos de forma más uniforme, se sigue destacando el intervalo modal pero las barras contiguas también tienen alturas similares, la barra más baja corresponde al intervalo de 6.8 a 7.4. Pasando a tercer grado, observamos que los datos ya no están tan uniformemente distribuidos, la barra más baja se encuentra en el rango de 9.2 a 10, y que las barras de los intervalos 6.8 a 7.4 y 8.4 a 9.2 tienen alturas muy similares. A pesar de las diferencias, los tres histogramas tienen distribución simétrica de los datos.

PRIMERO	SEGUNDO	TERCERO
---------	---------	---------



*Tabla 10- Boxplots e histogramas obtenidos para el primer ciclo de la asignatura de lengua de las escuelas de primarias*

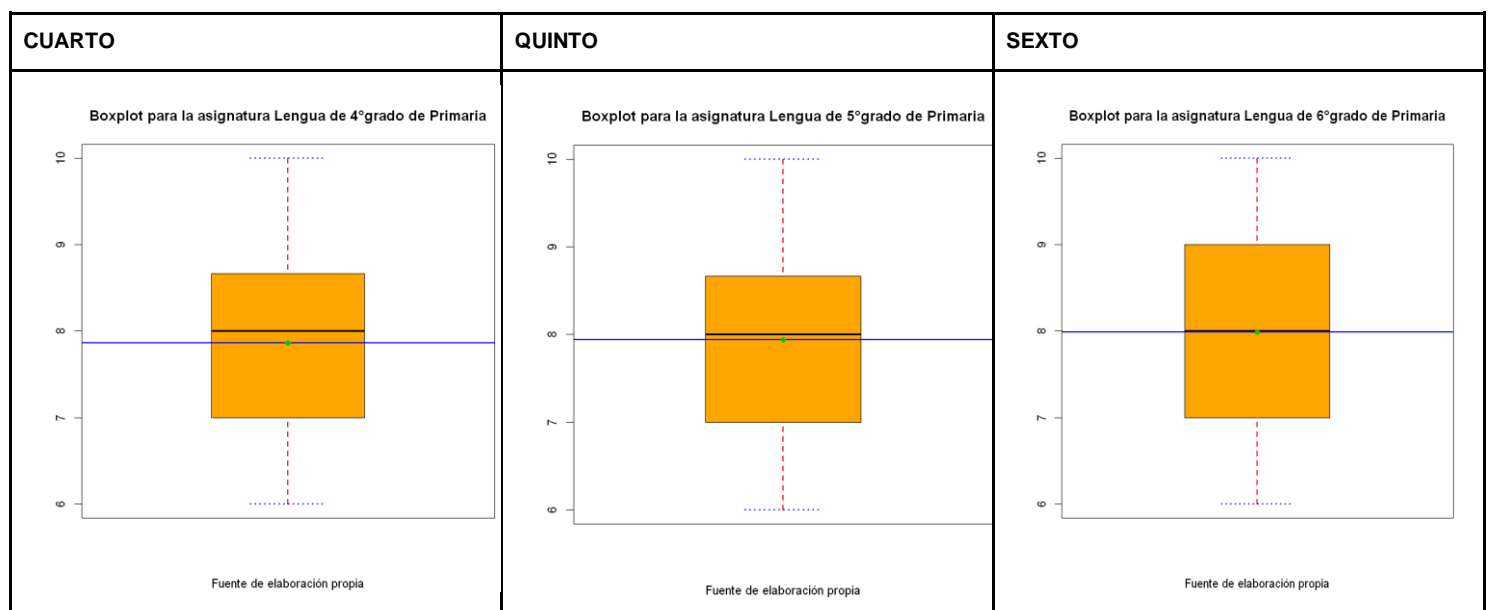
Estadístico	1er grado	2do grado	3er grado
Media	7.89	8.11	7.90
Mediana	8	8	8
Primer Cuartil	7	7	7
Tercer Cuartil	8.66	9	8.66
Mínimo / Máximo	6/10	6 / 10	6 /10

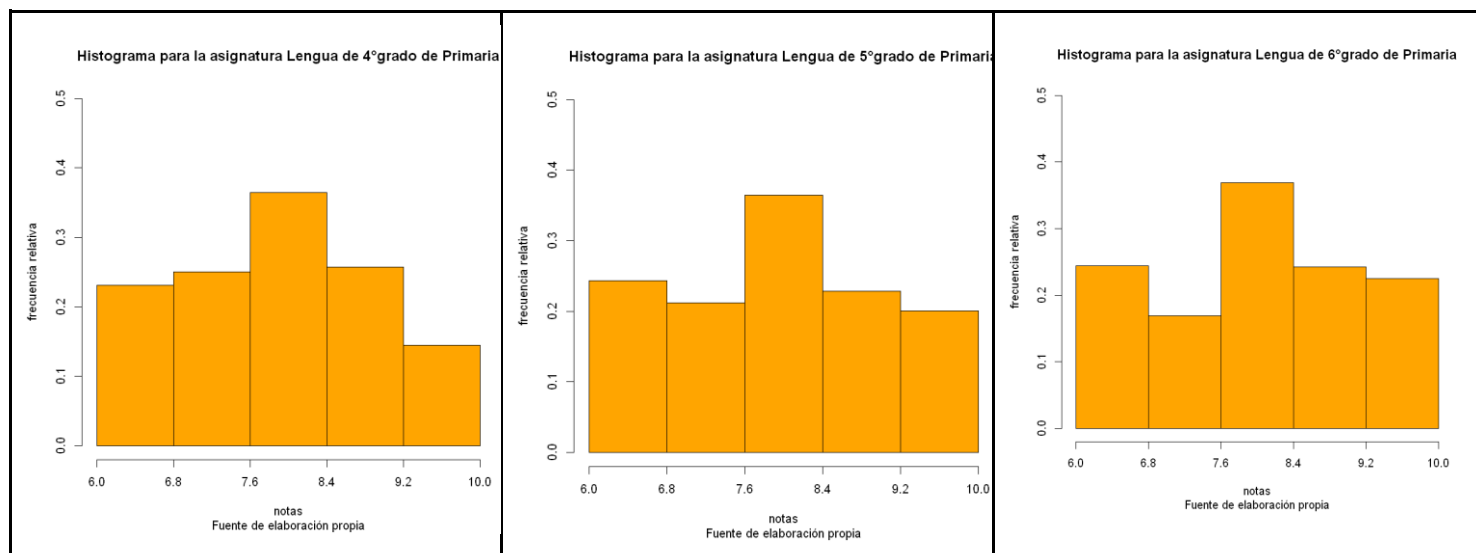
*Tabla 11- Valores estadísticos obtenidos para el primer ciclo de la asignatura de lengua de las escuelas de primarias*

Centrándonos en los estadísticos del primer ciclo de lengua para primaria del año 2022, vemos cómo se corresponden los boxplots con los valores obtenidos. En primer y tercer grado vemos que la media es menor a la mediana y que el tercer cuartil es el mismo para ambos casos. En segundo año es apreciable una leve diferencia entre la media y

la mediana, lo que hace que la primera sea superior. Como en el caso anterior, al tener poca variabilidad de los datos, es difícil concluir qué año tuvo mejor desempeño en este ciclo.

Finalizando el análisis de primaria, falta el segundo ciclo de lengua para el 2022. Comenzando con los boxplots, podemos ver que en los tres casos la máxima es de 10, mínima de 6 y que la mediana ronda el valor de 8, también podemos ver que la media es muy cercana a la mediana, por ende, podemos decir que es representativa de la misma. No hay valores atípicos. Comenzamos por el análisis de cuarto grado, vemos que el 50% de las notas las encontramos entre 7 y 8.5 aproximadamente, los datos están más centrados entre los valores 7 y 8, lo que hace que la caja se vea desplazada y que la mediana es mayor a la media. Para quinto grado el análisis de la caja es casi el mismo salvo con la diferencia de que la media está casi por encima de la mediana. En sexto grado ya podemos ver una caja distribuida más uniformemente, donde el 50% de las notas se encuentra entre 7 y 9, y la media está casi superpuesta a la mediana. Al igual que en el análisis anterior de los histogramas, observamos que los 3 son unimodal, con dicho valor en el rango de 7.6 a 8.4. En cuarto y quinto grado vemos que los gráficos son muy similares, salvo que en cuarto el rango en el cual encontramos menor frecuencia es de 9.2 a 10 y en quinto podemos decir que tenemos dos barras con alturas similares que corresponden a las notas menos frecuentes, estos rangos son 6.8 a 7.6 y de 9.2 a 10. En cambio, en sexto grado podemos observar que la barra más baja corresponde al rango de notas de 6.8 a 7.6, y que el resto de las barras tiene una altura muy similar. Resumiendo, podemos afirmar que la distribución es uniforme y que su media es representativa, además de que los datos están centrados en el rango de 7.6 a 8.4, los datos no están dispersos y que no hay presencia de valores atípicos. El desempeño de los alumnos para la asignatura fue óptimo.





*Tabla 12- Boxplots e histogramas obtenidos para el segundo ciclo de la asignatura de lengua de las escuelas de primarias*

Estadístico	4to grado	5to grado	6to grado
Media	7.86	7.94	7.99
Mediana	8	8	8
Primer Cuartil	7	7	7
Tercer Cuartil	8.66	8.66	9
Mínimo / Máximo	6/10	6 / 10	6 / 10

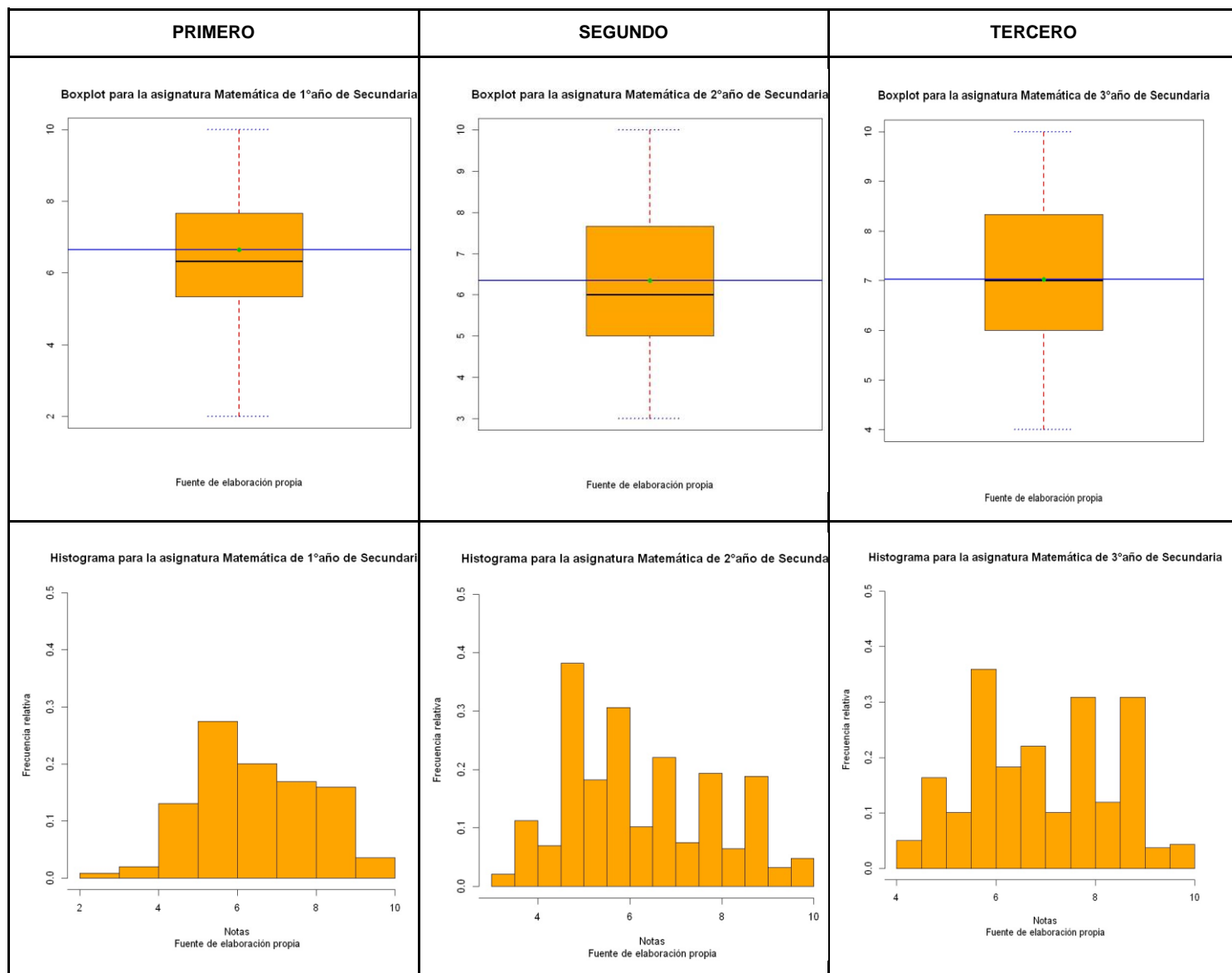
*Tabla 13- Valores estadísticos obtenidos para el segundo ciclo de la asignatura de lengua de las escuelas de primarias*

Los últimos estadísticos para analizar de primaria corresponden a lengua, en estos podemos ver algo que caracteriza al conjunto de estadísticos de ambas asignaturas, la poca variabilidad de los datos. En este caso en particular, la media es menor a la mediana, además podemos ver que como dijimos previamente, al ser el valor del tercer cuartil menor en cuarto y quinto año, la caja de los boxplots no es simétrica respecto a la de sexto año.

Analizando el desempeño tanto de matemática como de lengua en todos los años, concluimos que este fue óptimo ya que ninguno desaprobó, pero si nos enfocamos en los valores de las notas, en ambos casos la media ronda el valor de 8, por ende, con este valor si podemos afirmar que el desempeño fue bueno.

Habiendo completado el análisis para primaria, es hora que nos enfoquemos en el desempeño de los estudiantes secundarios para las asignaturas de matemática y lengua del 2022. Al igual que en primaria, dividiremos en primer ciclo y segundo ciclo.

A diferencia de la primaria, en secundaria si está permitido desaprobado a los estudiantes, por lo que la variabilidad de las notas o promedios será mucho mayor. Dicho esto, comenzaremos con el análisis para la asignatura de Matemática:



*Tabla 14- Boxplots e histogramas obtenidos para el primer ciclo de la asignatura de matemática de las escuelas de secundaria*

En la siguiente tabla podemos ver los valores de algunos estadísticos que nos ayudarán a analizar mejor sus gráficos correspondientes.

Estadístico	1er Año	2do Año	3er Año
Media	6.64	6	7
Mediana	6.33	6.34	7
Primer Cuartil	5.33	5	6
Tercer Cuartil	7.66	7.66	8.33
Mínimo / Máximo	2 / 10	3 / 10	4 / 10

*Tabla 15- Valores estadísticos obtenidos para el primer ciclo de la asignatura de matemática de las escuelas de secundarias*

Para este primer ciclo, en los dos primeros años el 50% de las notas apenas logra alcanzar un aprobado, estando por debajo de un promedio de 6, mientras que el 50% central ronda los valores entre 5 y 7. En el tercer año, el rendimiento académico mejora un poco, con el 50% de los datos no incluyendo desaprobados. Si observamos los histogramas, en el segundo año existen más cantidad de registros en promedios más bajos, cercanos a 5, mientras que, en tercero, si bien el intervalo con mayor frecuencia es menor a 6, los dos más repetidos siguientes, compensan con notas cercanas a 8. Como conclusión para los primeros tres años de secundaria podemos decir que el rendimiento en la asignatura de matemática fue regular y mejorable, si bien existen casos de muy buen rendimiento, la mayoría de registros rondan el 6 y muchos están por debajo. Saltando a los tres años siguientes, los gráficos correspondientes son:

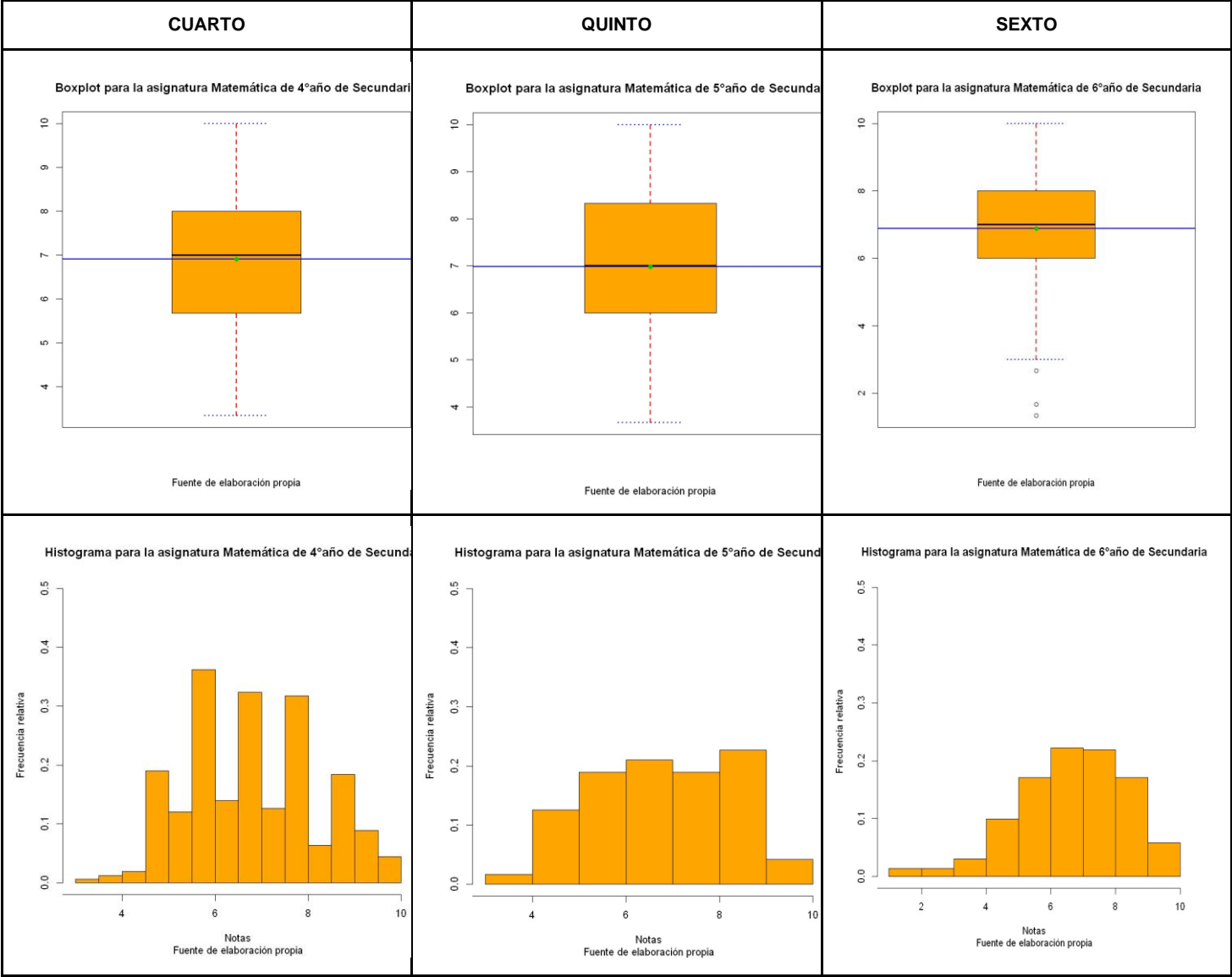


Tabla 16- Boxplots e histogramas obtenidos para el segundo ciclo de la asignatura de matemática de las escuelas de secundaria

Mientras que sus estadísticos:

Estadístico	4to Año	5to Año	6to Año
-------------	---------	---------	---------

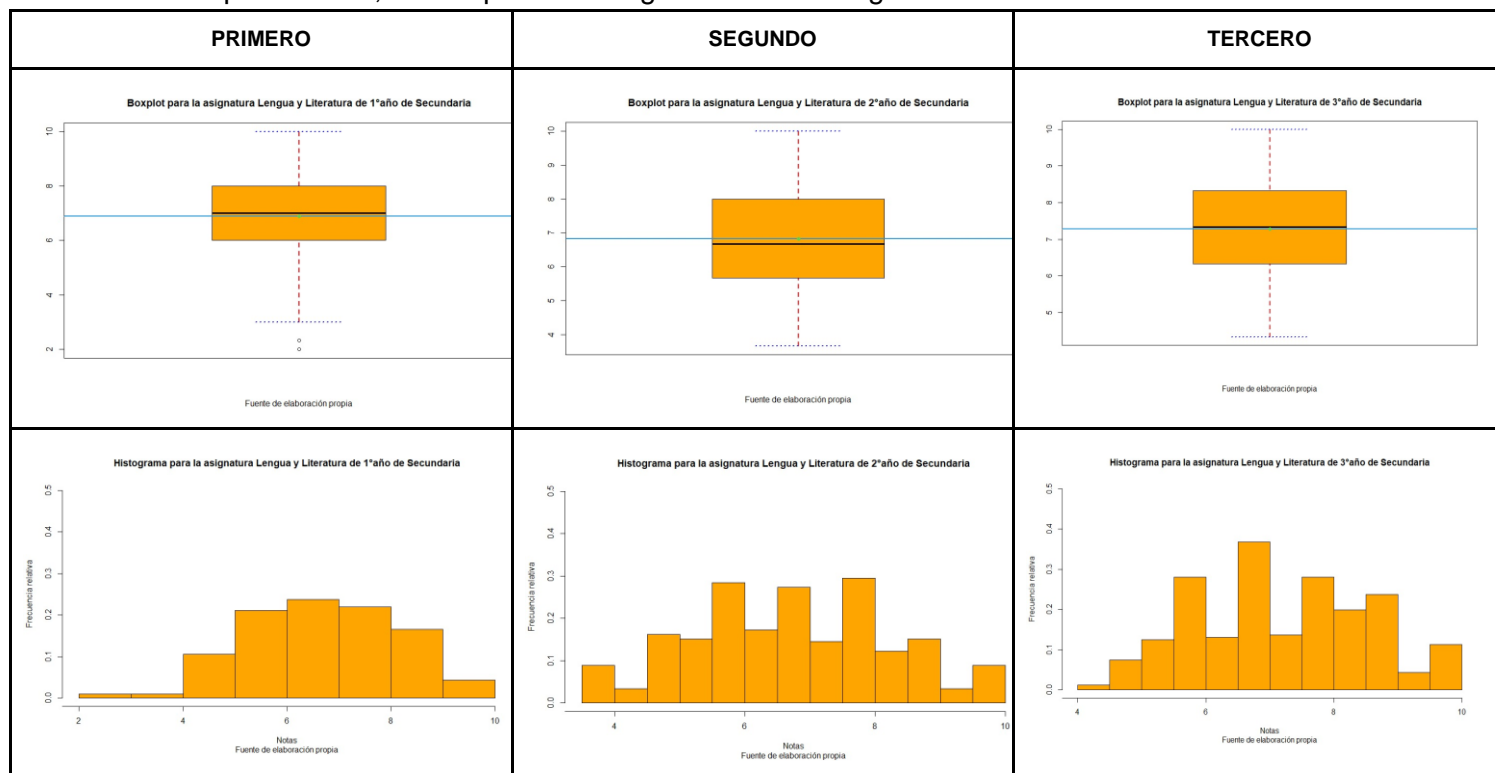


Media	6.9	7	6.88
Mediana	7	7	7
Primer Cuartil	5.6	6	6
Tercer Cuartil	8	8.33	8
Mínimo / Máximo	3.33 / 10	1.33 / 10	1.33 / 10

*Tabla 17- Valores estadísticos obtenidos para el segundo ciclo de la asignatura de matemática de las escuelas de secundarias*

Primeramente, ya podemos destacar una mejoría en cuanto al rendimiento comparándolo con los años del primer ciclo. Analizando los boxplots podemos ver que los de cuarto y quinto se parecen bastante, mientras que la caja del gráfico de sexto año es bastante más achatada, por lo que la variabilidad de los promedios es bastante menor, también se aprecian outliers en las notas más bajas, podríamos decir que es el año con mejor rendimiento de los tres, aunque tampoco es ampliamente superior, en todos los casos el 75% superior de los datos está por arriba de la nota mínima para aprobar. Si contemplamos los histogramas, podemos apreciar que en el de cuarto año existen más intervalos y los de mayor frecuencia rondan entre 6 y 7. Mientras que para quinto y sexto, el número de intervalos disminuye y ninguno resalta en frecuencia. Como conclusión podemos decir que el rendimiento académico para esta asignatura fue bueno, mejorando en comparación a lo analizado para los primeros tres años de la escuela secundaria analizados anteriormente.

Continuando con el análisis de la asignatura de lengua para secundaria en el 2022, para el primer ciclo, los boxplots e histogramas son los siguientes:



*Tabla 18- Boxplots e histogramas obtenidos para el primer ciclo de la asignatura de lengua de las escuelas de secundarias*

Sus estadísticos asociados son

Estadístico	1er Año	2do Año	3er Año
Media	6.9	6.8	7.28
Mediana	7	6.67	7.33
Primer Cuartil	6	5.66	6.33
Tercer Cuartil	8	8	8.33
Mínimo / Máximo	2 / 10	3.66 /10	4.33 /10

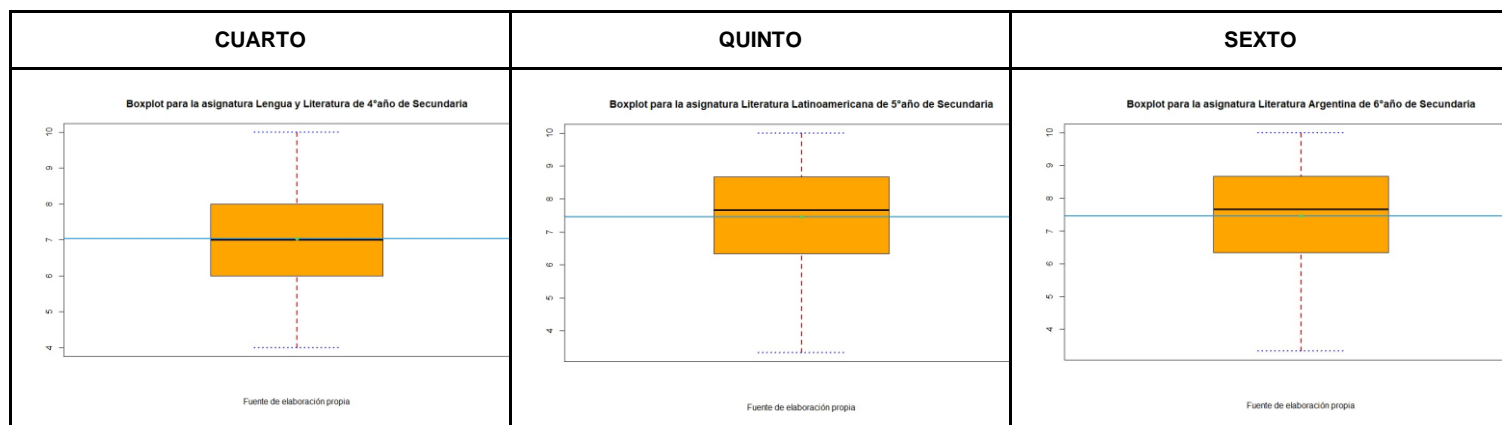
*Tabla 19- Valores estadísticos obtenidos para el primer ciclo de la asignatura de lengua de las escuelas de secundarias*

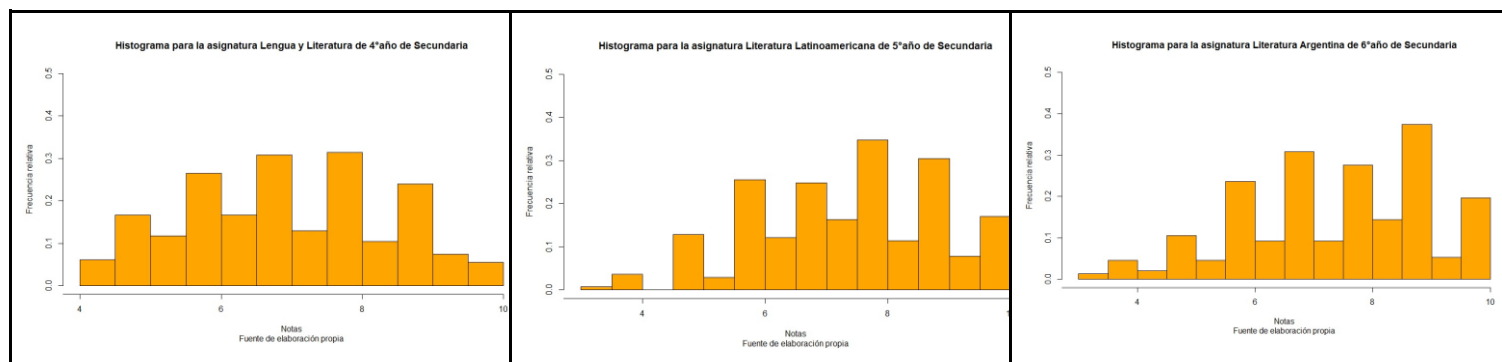
A diferencia de matemática, para estos primeros tres años podemos notar que los gráficos difieren entre sí, la caja de primer año es bastante más chata que las otras dos y cuenta con dos outliers en las notas más bajas (menos variabilidad). De los tres casos, podría decirse que tercer año fue quien obtuvo mejores resultados, con una mejor media y una caja (50% central de los datos) acomodada más arriba en términos de notas. En síntesis, podemos concluir que el rendimiento para esta asignatura fue bueno, si bien existen casos de peor desempeño, no destacan demasiado y podemos decir que los tres años obtuvieron buenos resultados.

Antes de realizar el análisis para el segundo ciclo, debemos realizar algunas aclaraciones. Cuando exploramos nuestro dataset vimos que para los últimos dos años no existía como tal la materia “Lengua y Literatura”, así que decidimos optar por analizar los resultados obtenidos en las dos materias que más se asemejan dentro de las posibilidades:

- Literatura Latinoamericana para quinto año
- Literatura Argentina para sexto año

Dicho esto, los gráficos obtenidos son los siguientes:





*Tabla 20- Boxplots e histogramas obtenidos para el segundo ciclo de la asignatura de lengua de las escuelas de secundarias*

La tabla con los estadísticos correspondientes:

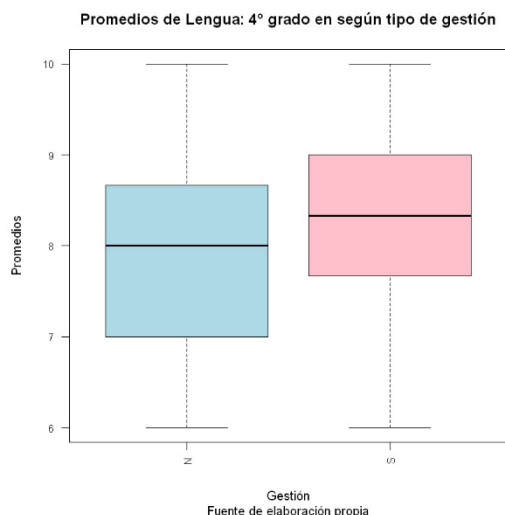
Estadístico	4to Año	5to Año	6to Año
Media	7.03	7.5	7.4
Mediana	7	7.67	7.67
Primer Cuartil	6	6.33	6.33
Tercer Cuartil	8	8.7	8.66
Mínimo / Máximo	4 / 10	3.33 / 10	3.33 / 10

*Tabla 21- Análisis estadístico obtenidos para el primer ciclo de la asignatura de lengua de las escuelas de secundarias*

Como se explicó previamente, los resultados obtenidos no son tan comparables entre sí porque no estamos hablando de la misma materia, pero podemos resaltar el buen rendimiento dentro de esta área. Analizando los boxplots, los tres poseen un 50% que está por arriba del mínimo para aprobar. Si bien, como en todos los casos existen situaciones de promedios que están muy por debajo de lo pedido, el desempeño de los tres años dentro de esta área de literatura es bueno, destacando principalmente el de quinto y sexto.

Una vez finalizada la primera etapa de análisis univariado, procederemos con el análisis bivariado.

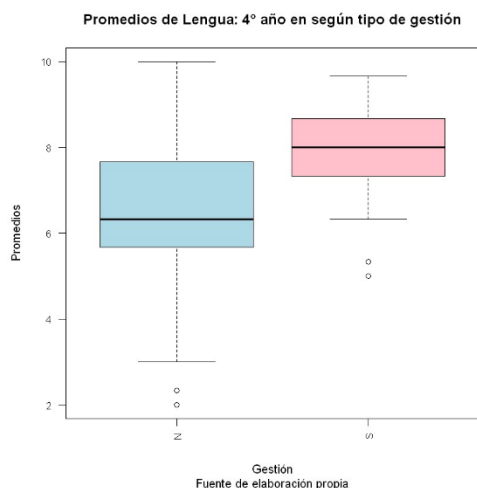
Primeramente, compararemos para 4to grado de primaria, diferenciando en la gestión de la escuela, privada o pública y veremos si existe alguna diferencia notoria entre los promedios.



*Ilustración 10- Análisis bivariado primaria*

Puede verse una diferencia sobre todo en la posición y tamaño de las cajas, la caja rosa, que corresponde a las escuelas privadas, se encuentra ubicada más arriba, por lo que podemos concluir que tuvieron un mejor rendimiento en Lengua que los alumnos de escuela pública.

Por otra parte, saltando a la educación secundaria, tomamos un criterio similar al anterior, donde analizamos el rendimiento de 4to año según la gestión de las escuelas, el resultado obtenido fue el siguiente:



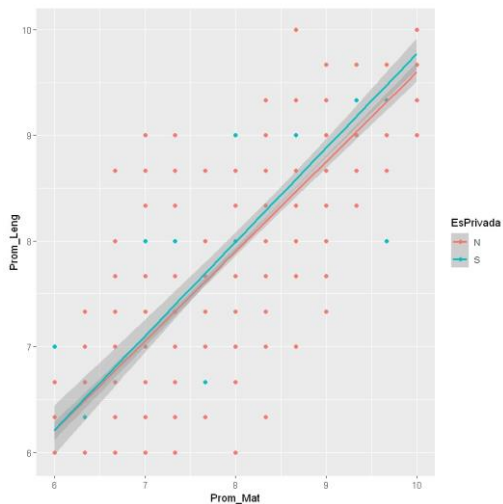
*Ilustración 11- Análisis bivariado secundaria*

En esta comparación los resultados obtenidos son similares a los de primaria, el boxplot de color rosa, perteneciente a la gestión privada, presenta mucho menor variabilidad con respecto a los promedios y está mejor ubicado en altura, dejándonos en evidencia un mejor rendimiento en comparación a la gestión pública.

En la segunda parte, nos centraremos en analizar si existe una correlación entre los promedios de las materias analizadas en este informe. De la misma manera en la que venimos trabajando, dividiremos el análisis

para cada tipo de educación.

Para primaria elegimos nuevamente a los 4to grado e intentaremos deducir si los promedios de matemática y lengua están correlacionados. Primeramente, realizamos una exploración gráfica a través de un scatterplot.



*Ilustración 12- Scatterplot para correlación por tipo de gestión y promedios de lengua y matemática en primaria*

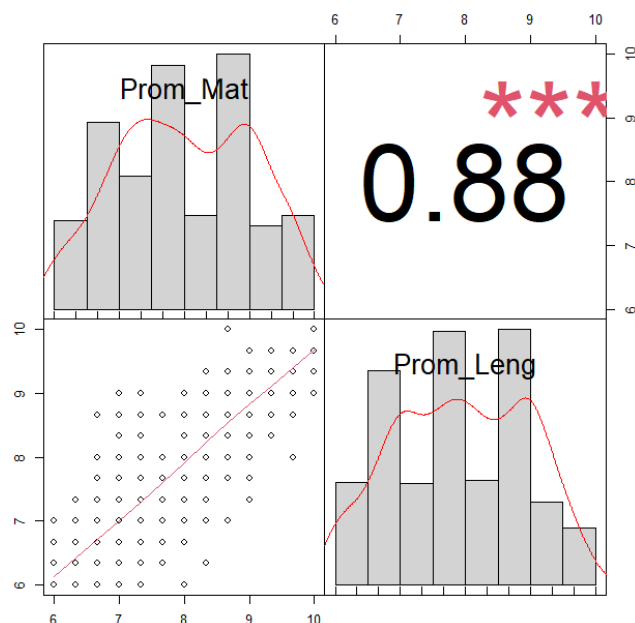
En un principio podemos ver que existe una relación lineal y creciente, aunque esto no es suficiente para establecer una conclusión.

El siguiente paso es pasar por un test de correlación propiamente dicho, para elegir qué tipo de test realizaremos, debemos revisar si las variables estudiadas presentan una distribución normal, al tratarse de una muestra de más de 700 variables, utilizamos la prueba de Kolmogorov-Smirnov con la variación de Lilliefors, obteniendo los siguientes resultados:

Variable	Valor D	p-valor	¿Es aproximadamente normal?
Promedios Mat	0.1027	$2.2 \times 10^{-16}$	No
Promedios Lengua	0.102	$2.2 \times 10^{-16}$	No

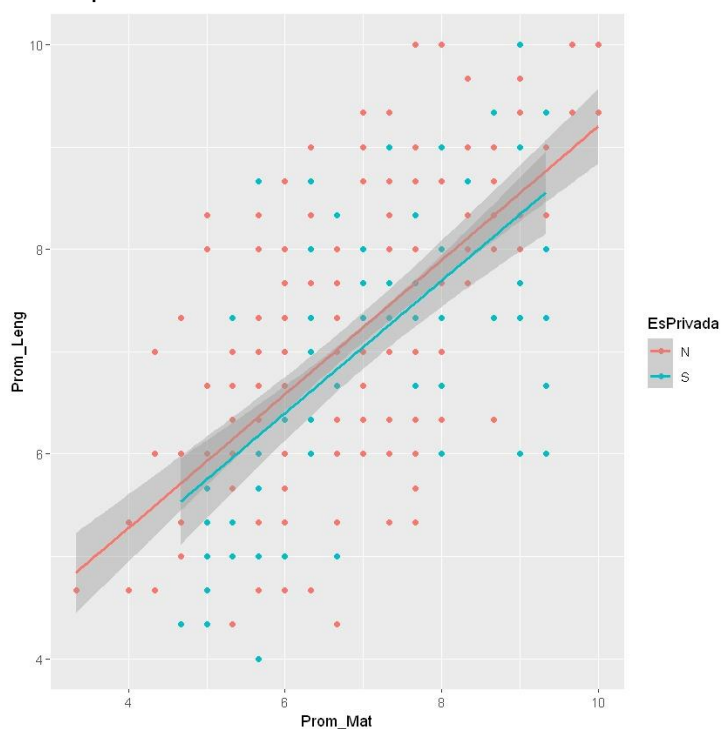
*Tabla 22- Análisis estadístico obtenidos para los promedios de matemática y lengua de primaria*

Como el p-valor nos resultó mucho menor que el  $\alpha = 0.05$  debemos rechazar la  $H_0$  que plantea la normalidad de la distribución. Sabiendo esto el siguiente paso es realizar el test de correlación con la variante de Spearman, dicha prueba arroja de coeficiente igual a 0.877 por lo que podemos concluir que existe una correlación lineal muy buena entre las variables estudiadas, como puede verse en el siguiente gráfico:



*Ilustración 13- Correlación entre lengua y matemática de primaria para cuarto grado*

Pasando al análisis de secundaria, de manera similar, tomamos los promedios de matemática y lengua de 4to año para ver si existe una correlación entre ellos. Siguiendo la misma metodología del caso anterior, realizamos un análisis gráfico a través de un scatterplot:



*Ilustración 14- Scatterplot para correlación por tipo de gestión y promedios de lengua y matemática en secundaria*

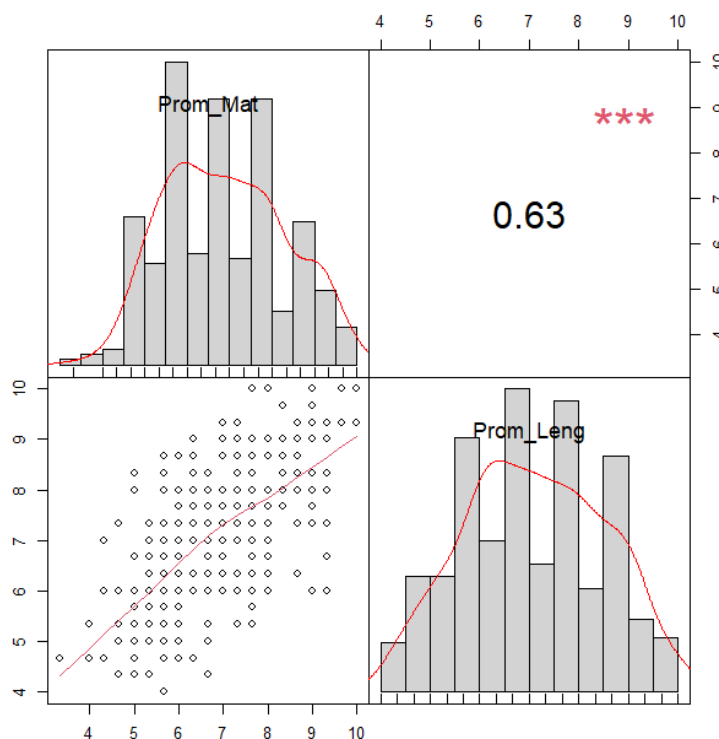
En un principio puede verse que si existe una tendencia de relación directamente proporcional. Como esto no es suficiente, pasamos a realizar los test correspondientes, primero analizando la normalidad de las variables utilizando la prueba de Kolmogorov-

Smirnov-Lilliefors (las muestras superan los 300 valores), obtuvimos los siguientes resultados:

Variable	Valor D	p-valor	¿Es aproximadamente normal?
Promedios Mat	0.096	$2.95 \times 10^{-7}$	No
Promedios Lengua	0.076	0.00015	No

*Tabla 23- Análisis estadístico obtenidos para los promedios de matemática y lengua de secundaria*

Dado los p-valores obtenidos podemos descartar la normalidad de las variables estudiadas por lo que para poder determinar si existe una correlación entre ellas, debemos optar por realizar un test con la variable de Spearman, la misma nos devolvió el siguiente valor del coeficiente: 0.634, podría decirse que sí, existe una relación lineal buena entre los promedios de matemática y lengua para 4to año de secundaria. Es un valor menor al que obtuvimos para la correlación de primaria, esto puede ser esperado ya que durante los estudios secundarios existen más materias y puede que exista una tercera variable que esté influyendo, ya que las materias estudiadas en este análisis pertenecen a áreas totalmente opuestas. Los resultados obtenidos pueden verse reflejados mejor en el siguiente gráfico:



*Ilustración 15- Correlación entre lengua y matemática de primaria para cuarto año*

## Conclusiones

Luego del extenso análisis realizado sobre la educación primaria y secundaria en el departamento de Diamante durante el año 2022 pudimos obtener las siguientes conclusiones.

La educación primaria cuenta con la mayor cantidad de alumnos en comparación a la secundaria, por otro lado, la mayoría de los estudiantes eligen la educación pública por sobre la gestión privada, prefiriendo los turnos mañana y tarde. Con respecto al análisis univariado de los rendimientos académicos, en primaria los resultados obtenidos fueron certeros, tanto para matemática como para lengua, los promedios se encuentran dentro de lo esperado. Por otro lado, al existir mucha más variabilidad en notas, existen múltiples situaciones para los alumnos, en términos generales, el rendimiento para matemática fue regular y mejorable, mientras en el área de lengua y literatura los promedios estuvieron mejor que en la otra materia.

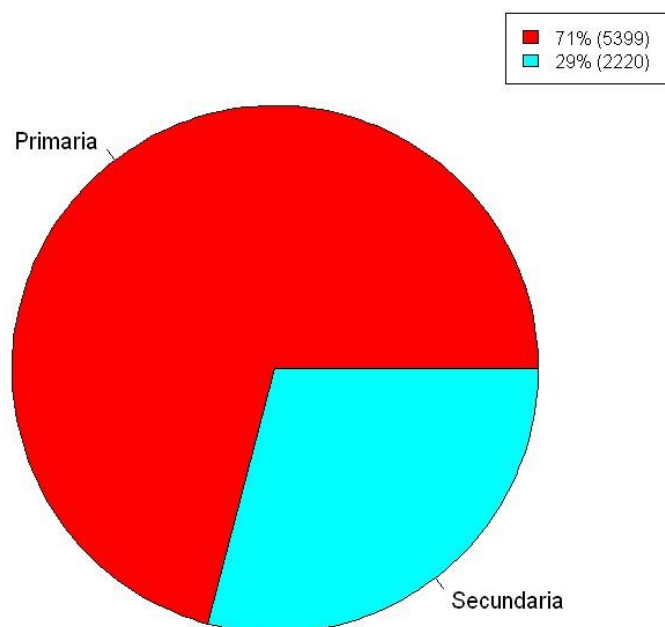
En contraposición, en nuestro análisis bivariado obtuvimos que los estudiantes de escuelas privadas tienden a tener mejor rendimiento. Por último, con la intención de encontrar si existía alguna correlación entre las materias estudiadas, encontramos que sí, existe una correlación positiva entre los promedios para ambos niveles de educación, siendo mayor en primaria.



## Anexo

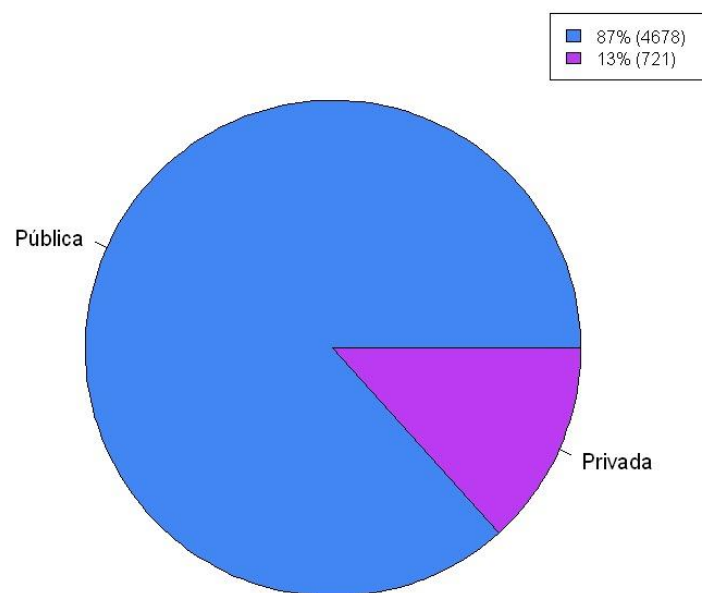
Imágenes:

**Distribución de los alumnos por Nivel de Enseñanza**



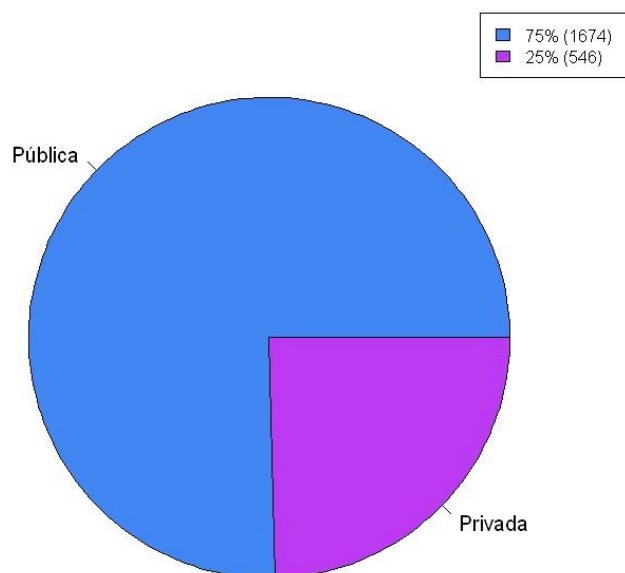
Fuente de elaboración propia

### Alumnos por tipo de gestión en escuelas primarias



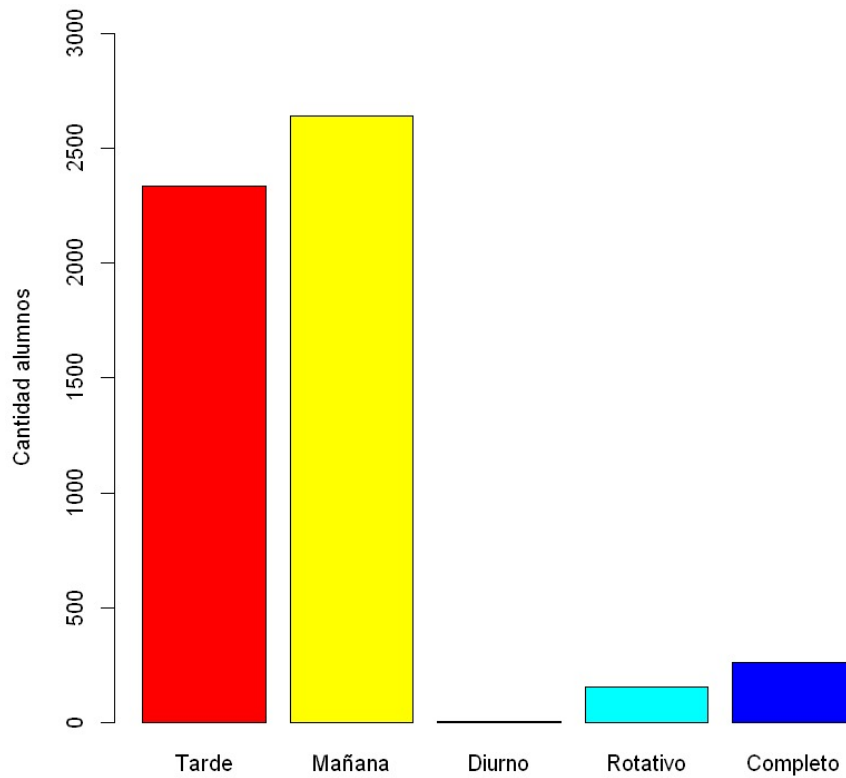
Fuente de elaboración propia

### Alumnos por tipo de gestión en escuelas secundarias



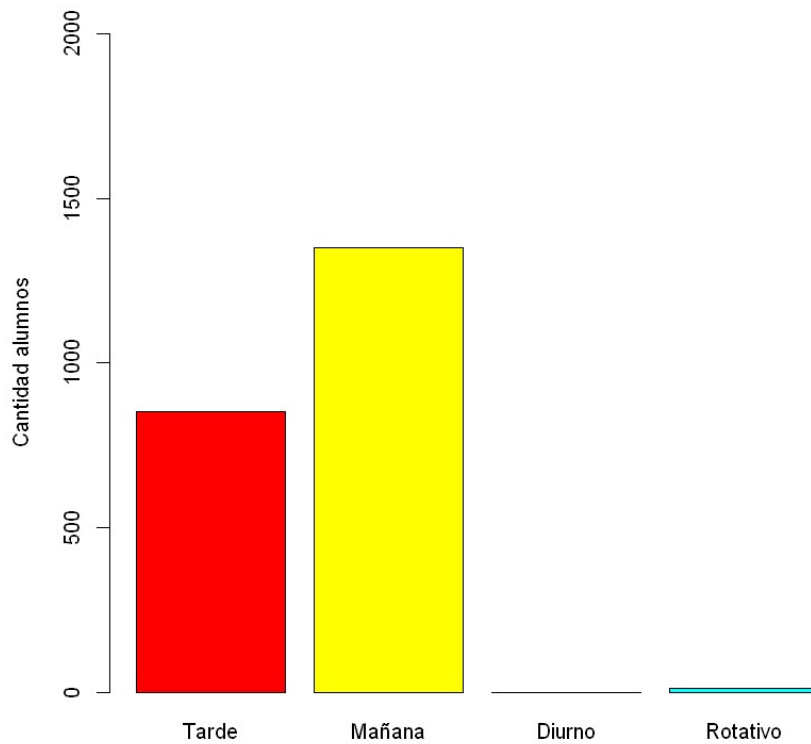
Fuente de elaboración propia

**Frecuencia absoluta de alumnos por turno en escuelas primarias**



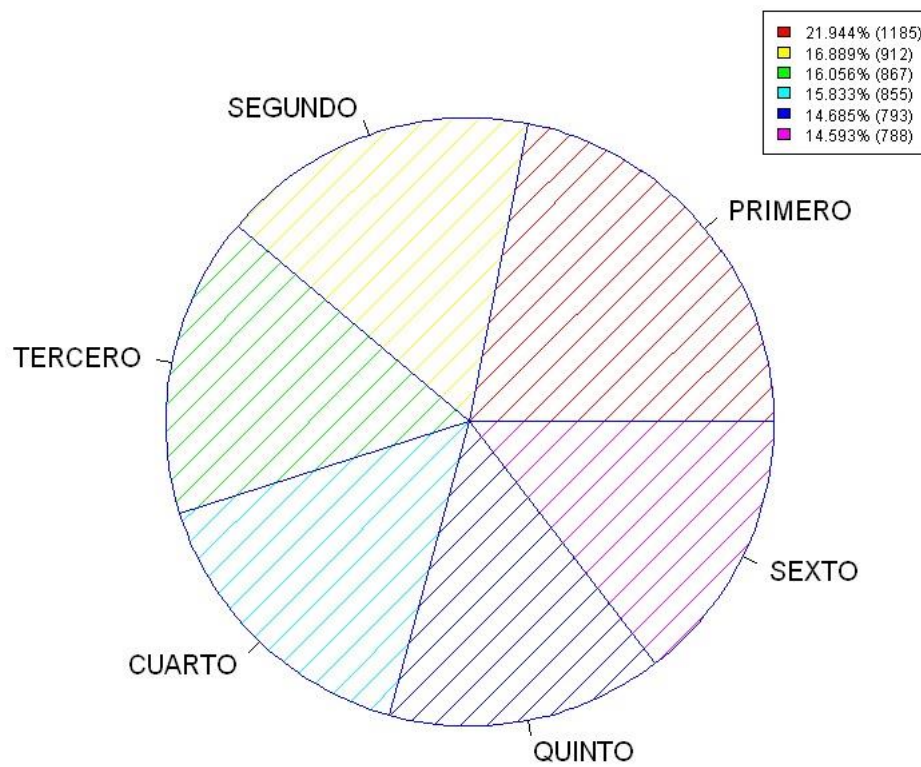
Fuente de elaboración propia

**Frecuencia absoluta de alumnos por turno en escuelas secundarias**



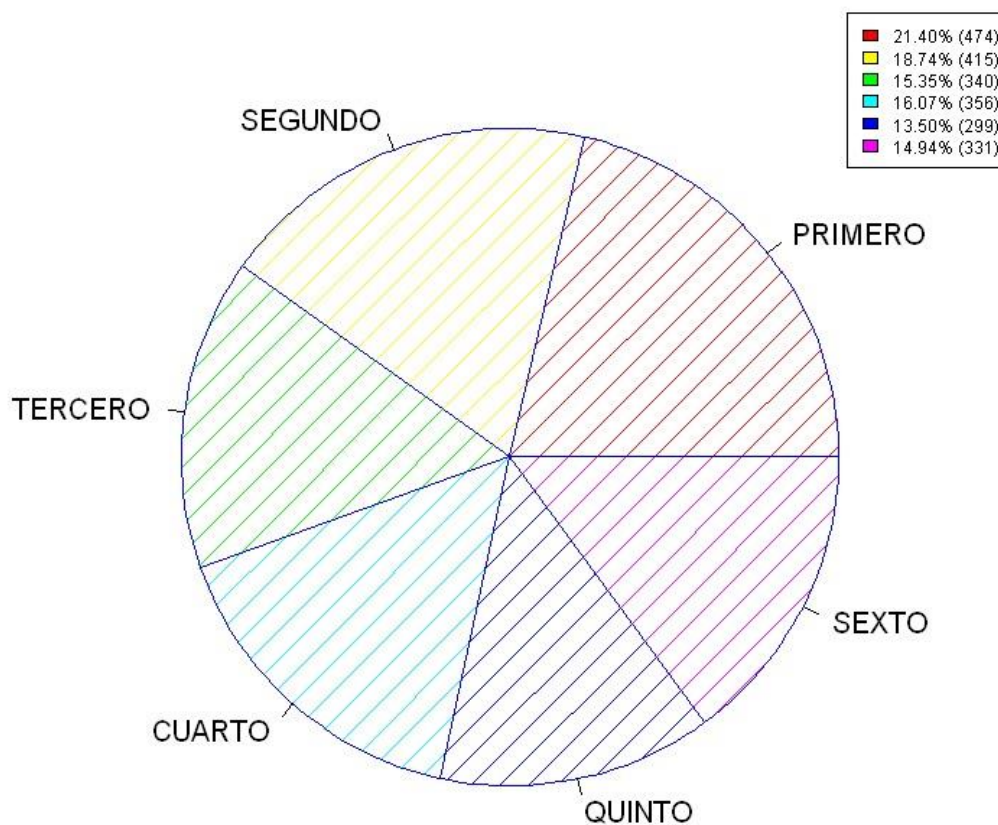
Fuente de elaboración propia

## Distribución de los alumnos por curso en escuelas primarias



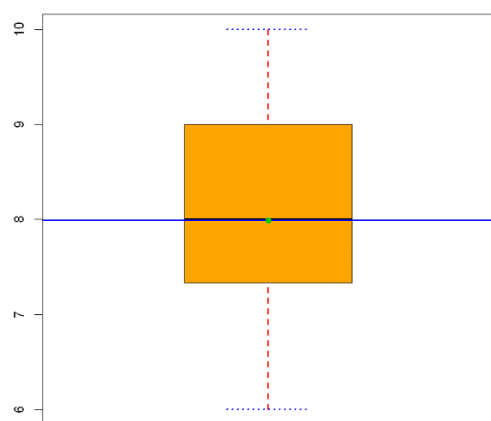
Fuente de elaboración propia

## Distribución de los alumnos por curso en escuelas secundarias



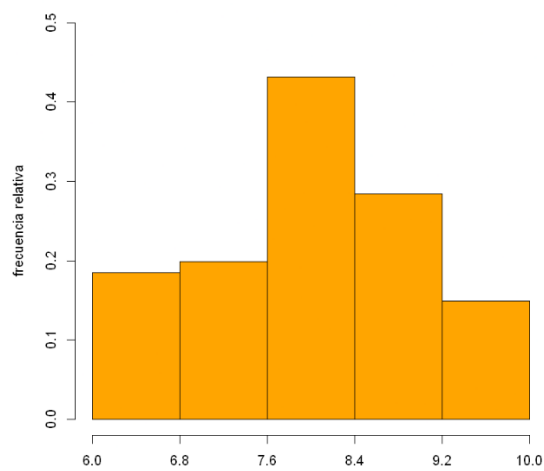
## Fuente de elaboración propia

Boxplot para la asignatura Matemática de 1º grado de Primaria



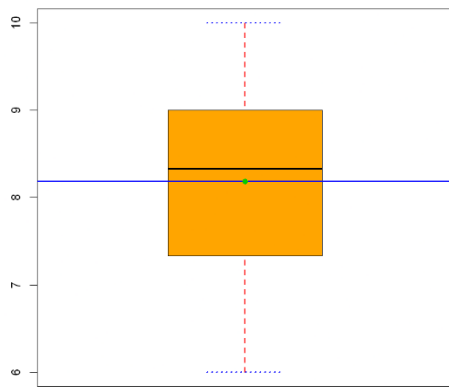
Fuente de elaboración propia

Histograma para la asignatura Matemática de 1º grado de Primaria



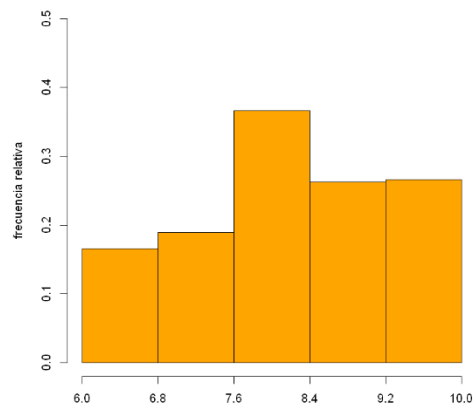
notas  
Fuente de elaboración propia

Boxplot para la asignatura Matemática de 2º grado de Primaria



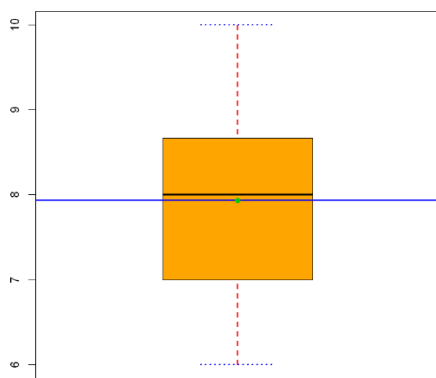
Fuente de elaboración propia

Histograma para la asignatura Matemática de 2º grado de Primaria



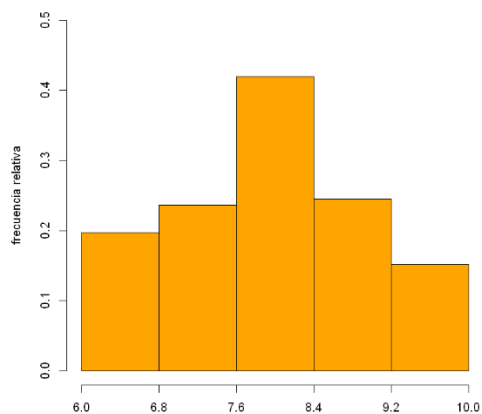
Fuente de elaboración propia

Boxplot para la asignatura Matemática de 3º grado de Primaria



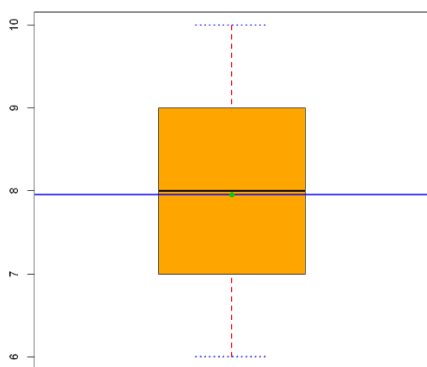
Fuente de elaboración propia

Histograma para la asignatura Matemática de 3º grado de Primaria



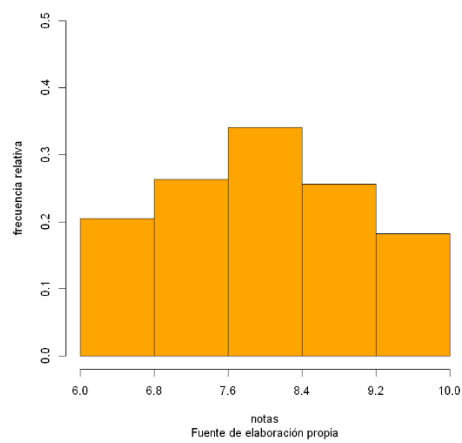
Fuente de elaboración propia

Boxplot para la asignatura Matemática de 4º grado de Primaria



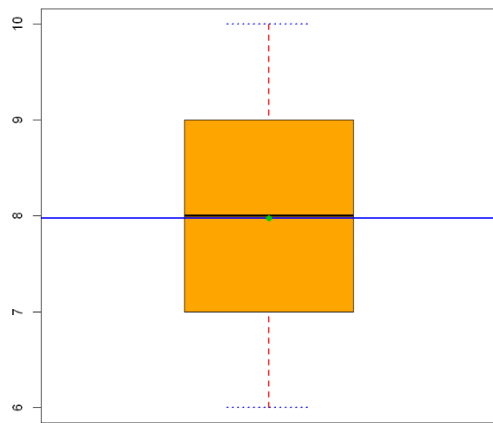
Fuente de elaboración propia

Histograma para la asignatura Matemática de 4º grado de Primaria



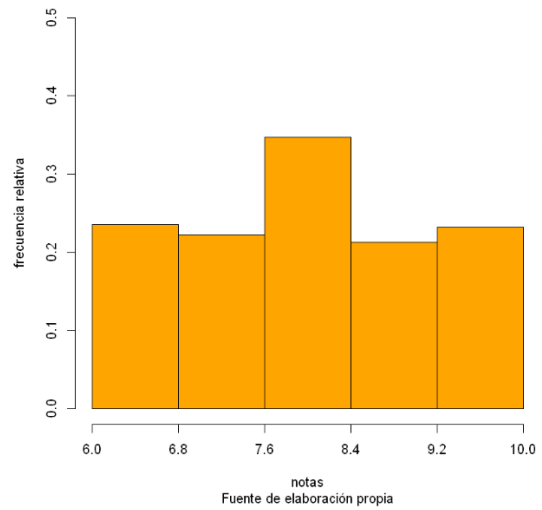
Fuente de elaboración propia

Boxplot para la asignatura Matemática de 5º grado de Primaria

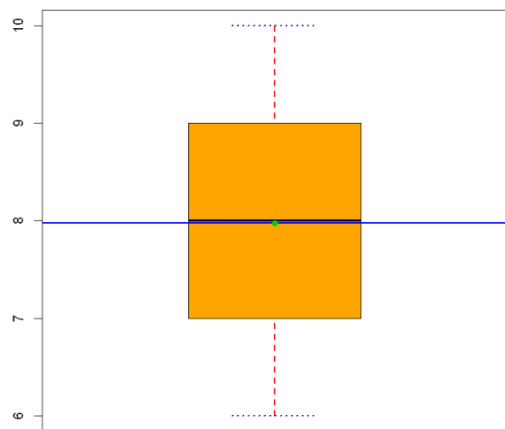


Fuente de elaboración propia

Histograma para la asignatura Matemática de 5º grado de Primaria

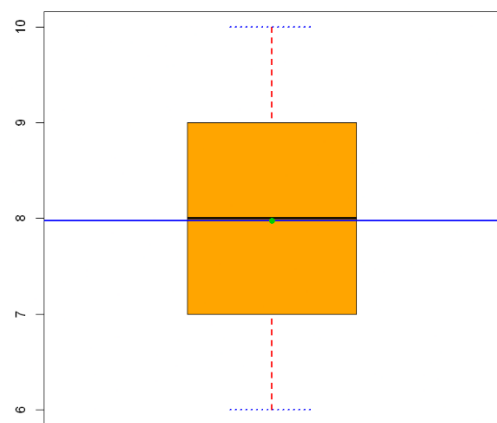


Boxplot para la asignatura Matemática de 6º grado de Primaria



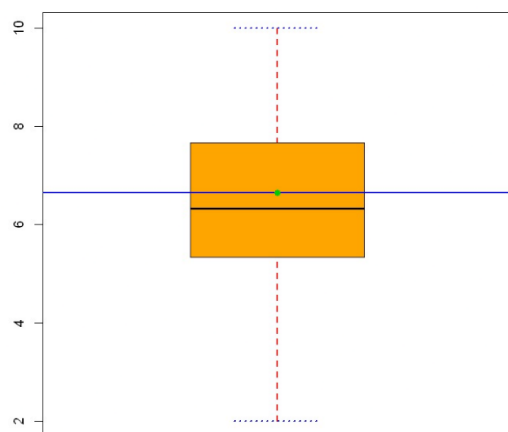
Fuente de elaboración propia

Boxplot para la asignatura Matemática de 6º grado de Primaria



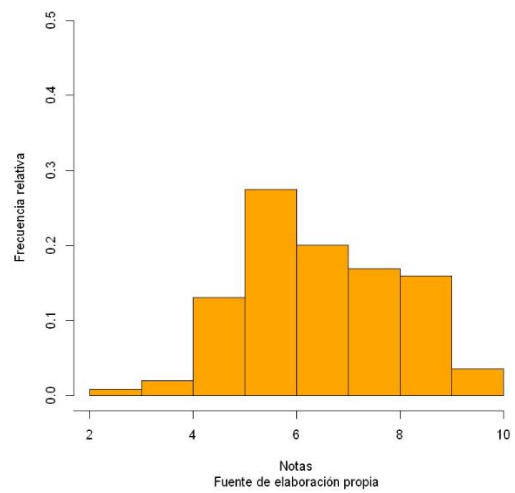
Fuente de elaboración propia

Boxplot para la asignatura Matemática de 1º año de Secundaria

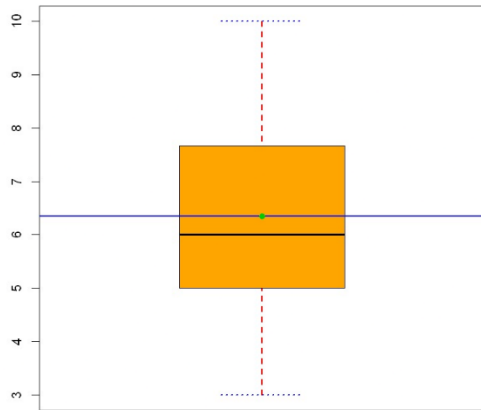


Fuente de elaboración propia

Histograma para la asignatura Matemática de 1º año de Secundaria

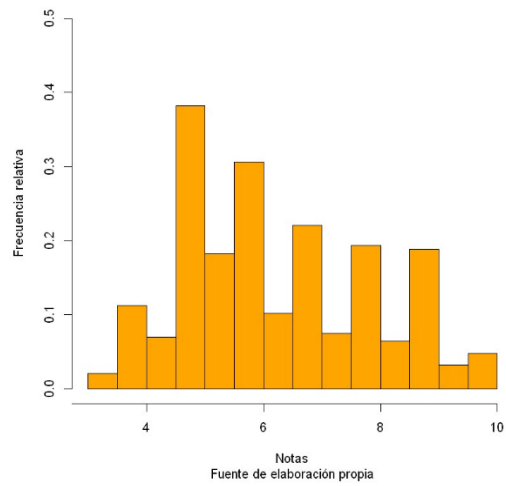


Boxplot para la asignatura Matemática de 2º año de Secundaria

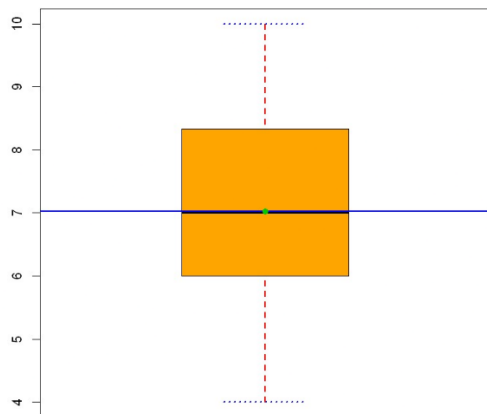


Fuente de elaboración propia

Histograma para la asignatura Matemática de 2º año de Secundaria

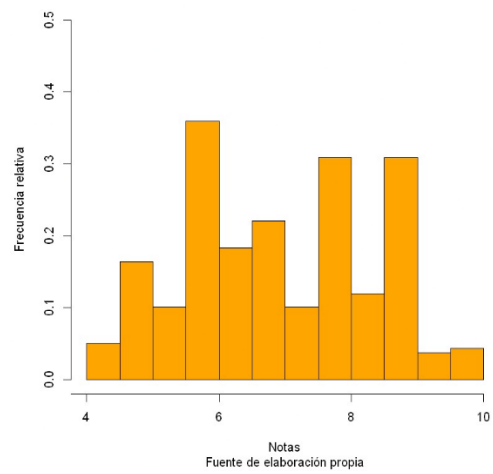


Boxplot para la asignatura Matemática de 3º año de Secundaria

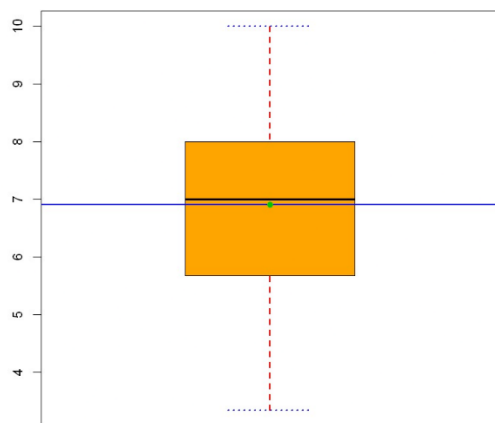


Fuente de elaboración propia

Histograma para la asignatura Matemática de 3º año de Secundaria

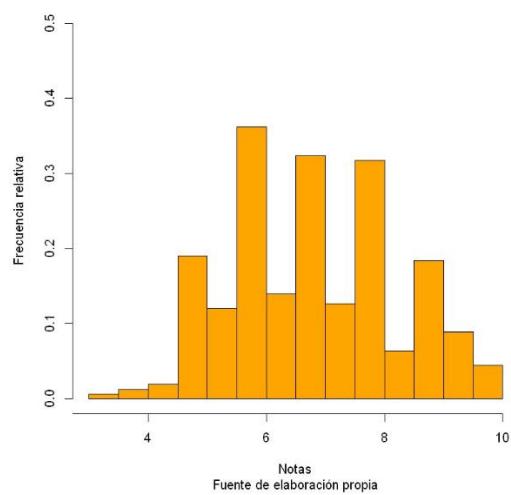


Boxplot para la asignatura Matemática de 4º año de Secundaria



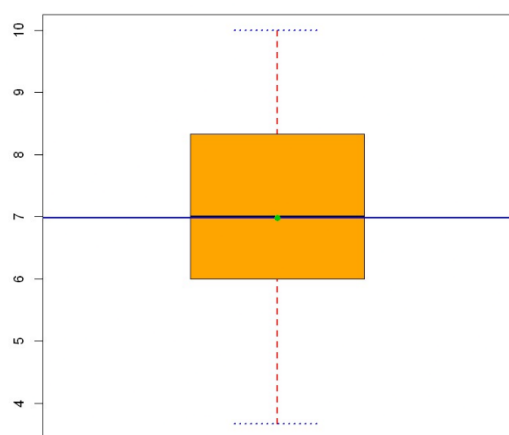
Fuente de elaboración propia

Histograma para la asignatura Matemática de 4º año de Secundaria



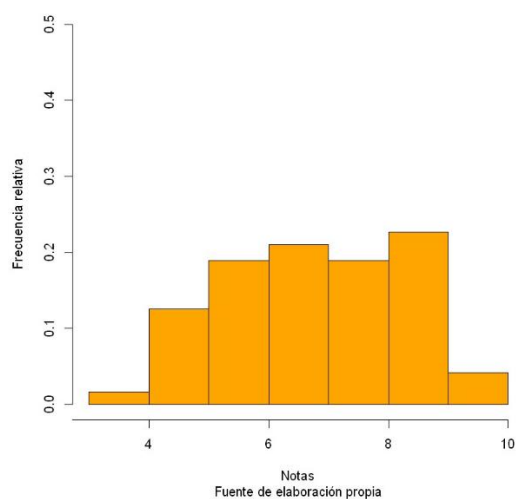


Boxplot para la asignatura Matemática de 5º año de Secundaria

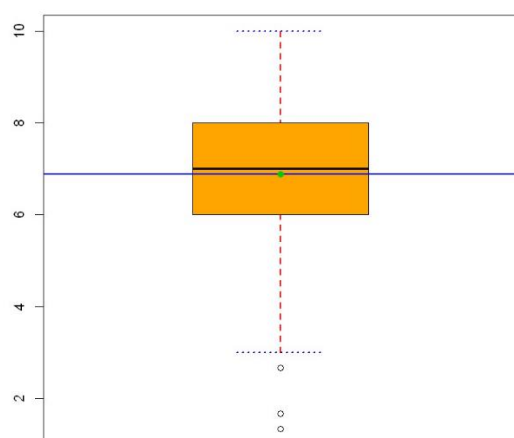


Fuente de elaboración propia

Histograma para la asignatura Matemática de 5º año de Secundaria

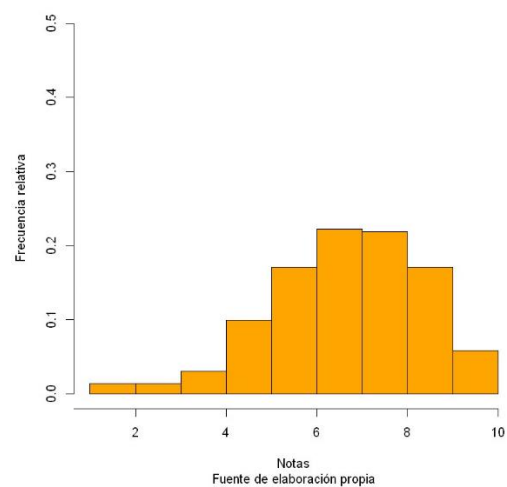


Boxplot para la asignatura Matemática de 6º año de Secundaria

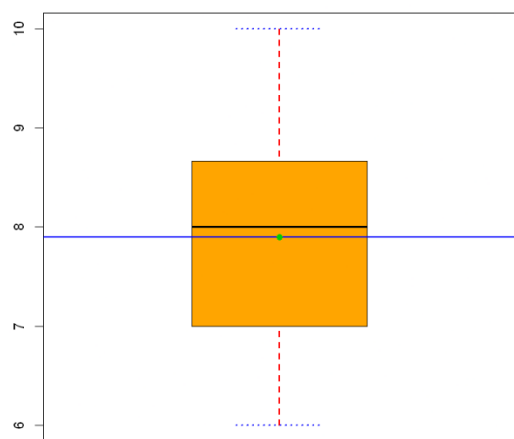


Fuente de elaboración propia

Histograma para la asignatura Matemática de 6º año de Secundaria

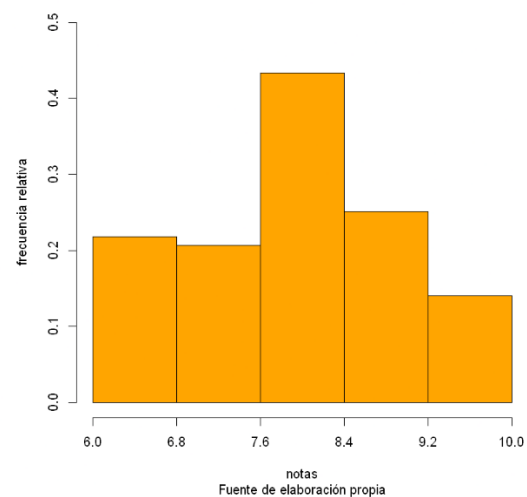


Boxplot para la asignatura Lengua de 1º grado de Primaria

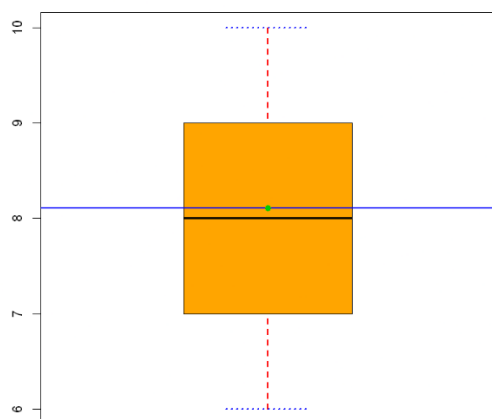


Fuente de elaboración propia

Histograma para la asignatura Lengua de 1º grado de Primaria

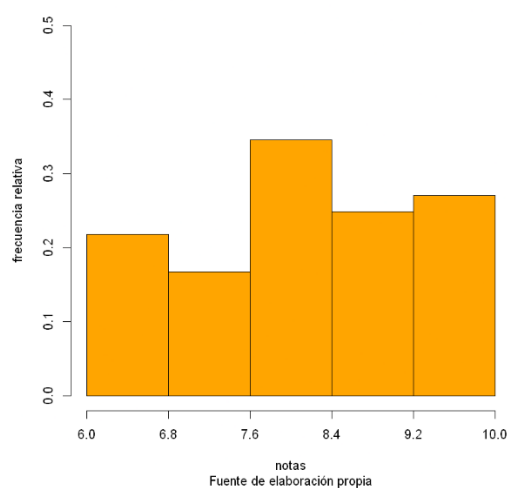


Boxplot para la asignatura Lengua de 2º grado de Primaria

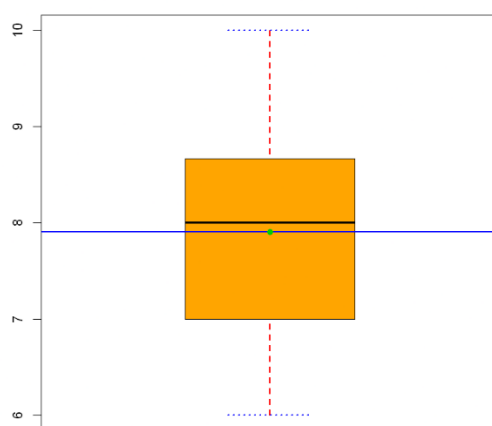


Fuente de elaboración propia

Histograma para la asignatura Lengua de 2º grado de Primaria

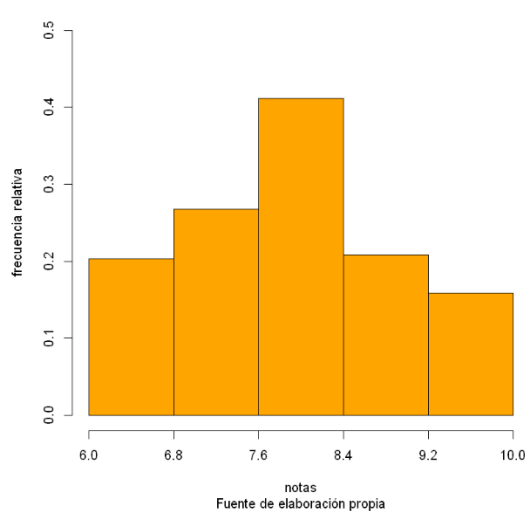


Boxplot para la asignatura Lengua de 3º grado de Primaria

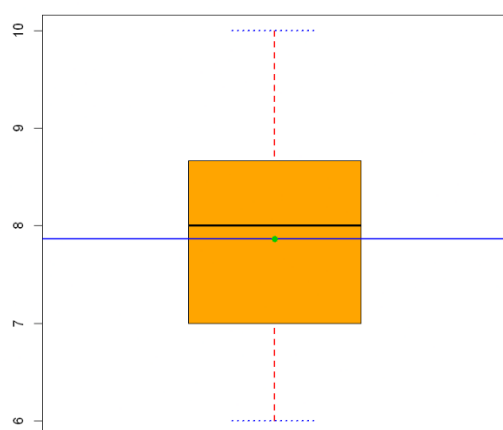


Fuente de elaboración propia

Histograma para la asignatura Lengua de 3º grado de Primaria

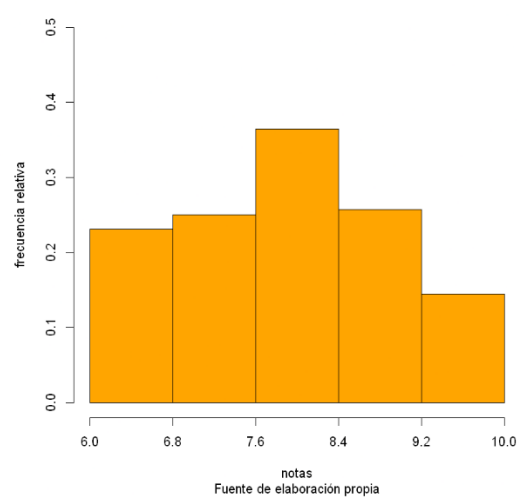


Boxplot para la asignatura Lengua de 4º grado de Primaria

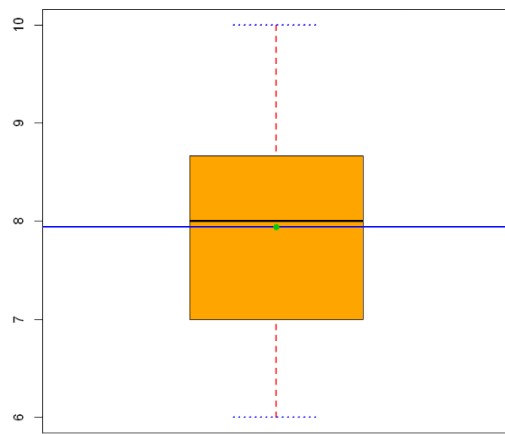


Fuente de elaboración propia

Histograma para la asignatura Lengua de 4º grado de Primaria

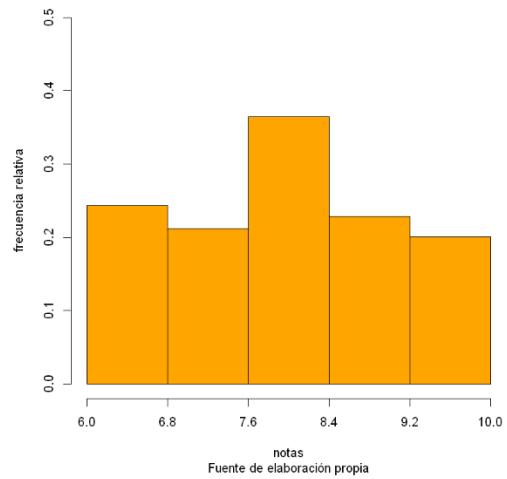


Boxplot para la asignatura Lengua de 5º grado de Primaria



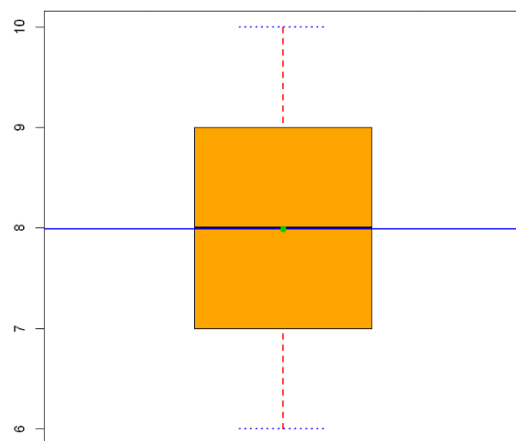
Fuente de elaboración propia

Histograma para la asignatura Lengua de 5º grado de Primaria



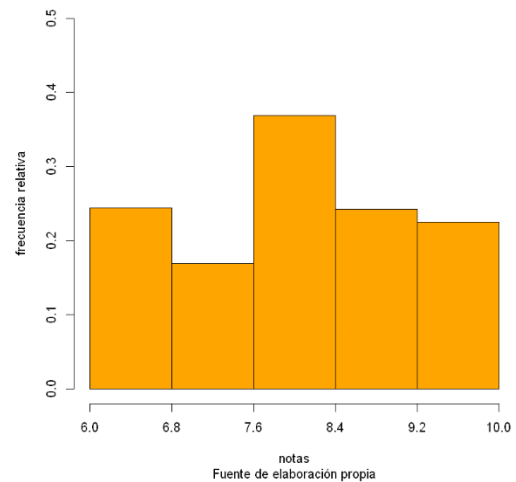
Fuente de elaboración propia

Boxplot para la asignatura Lengua de 6º grado de Primaria



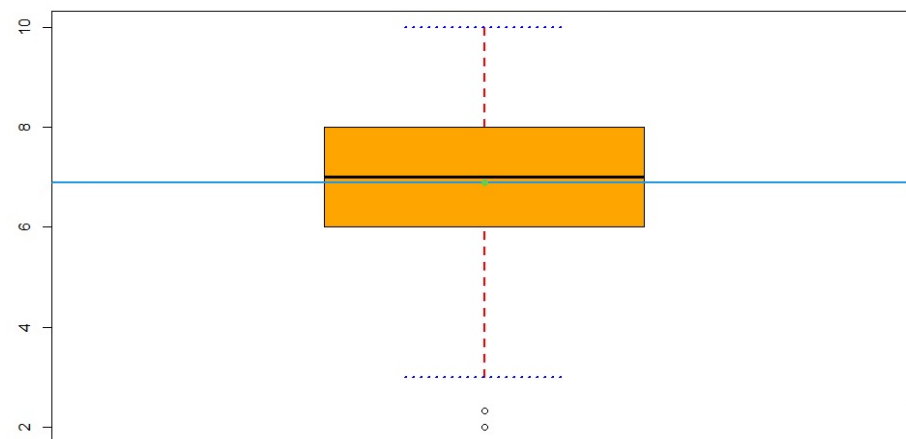
Fuente de elaboración propia

Histograma para la asignatura Lengua de 6º grado de Primaria



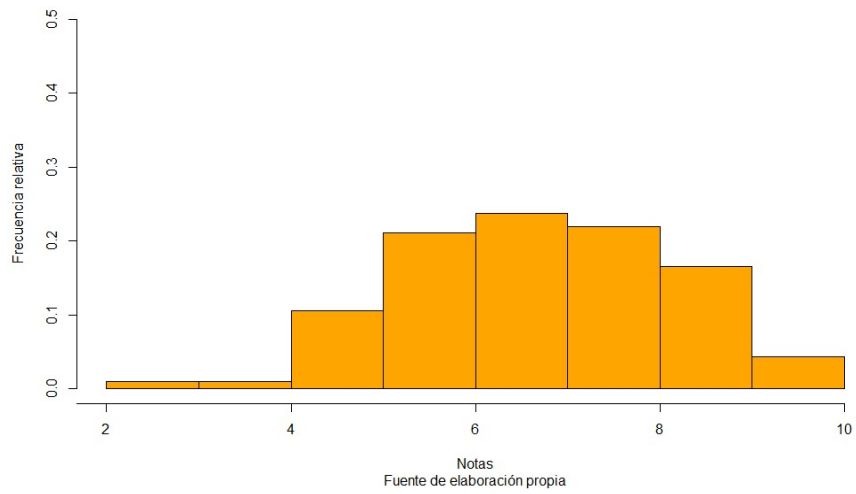
Fuente de elaboración propia

Boxplot para la asignatura Lengua y Literatura de 1º año de Secundaria

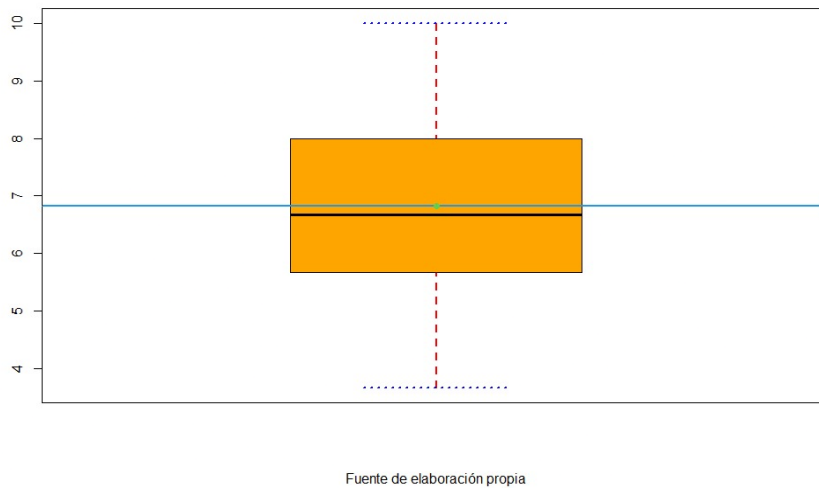


Fuente de elaboración propia

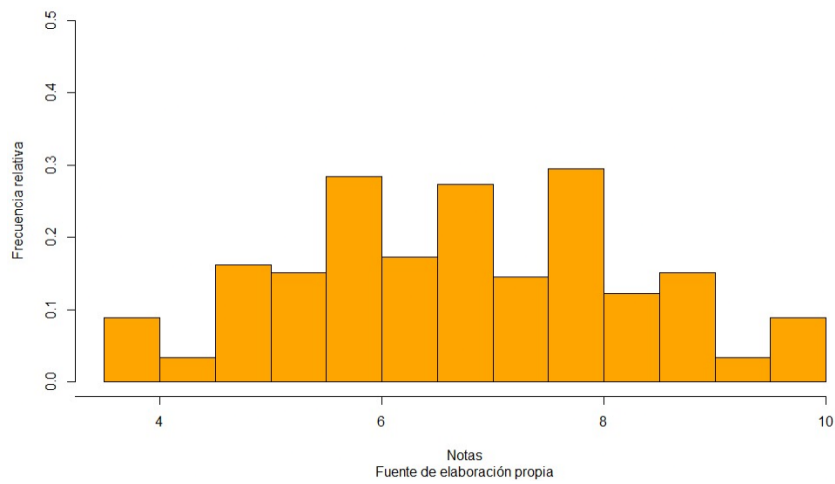
**Histograma para la asignatura Lengua y Literatura de 1º año de Secundaria**



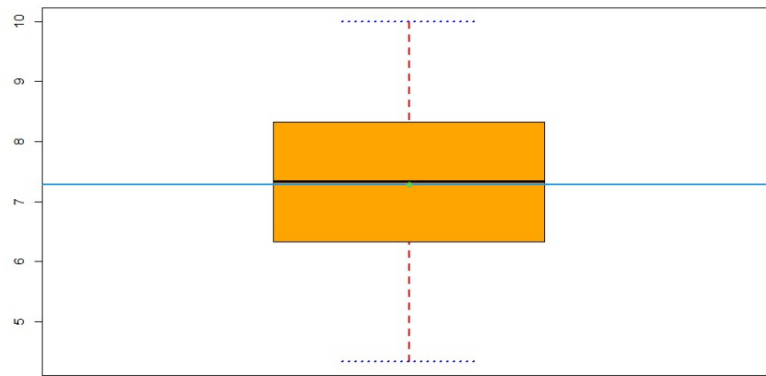
**Boxplot para la asignatura Lengua y Literatura de 2º año de Secundaria**



**Histograma para la asignatura Lengua y Literatura de 2º año de Secundaria**

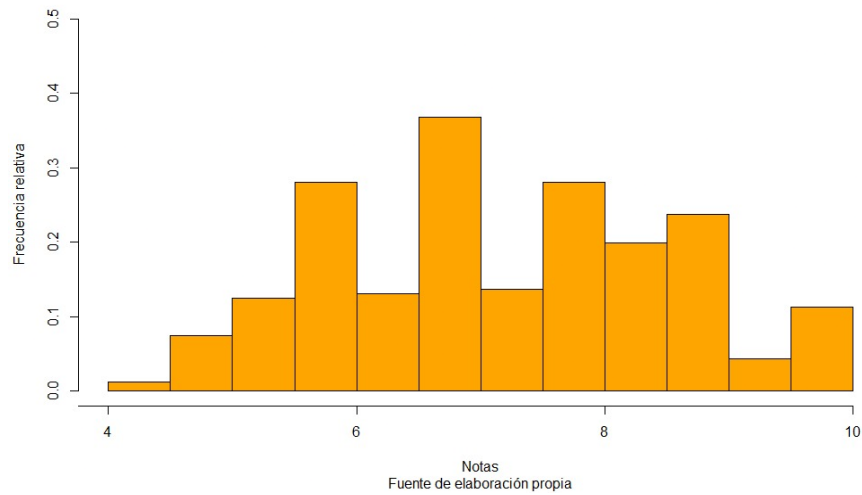


**Boxplot para la asignatura Lengua y Literatura de 3º año de Secundaria**



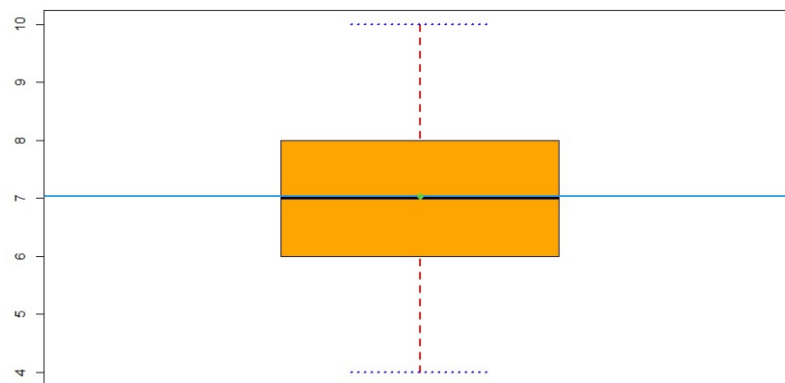
Fuente de elaboración propia

**Histograma para la asignatura Lengua y Literatura de 3º año de Secundaria**



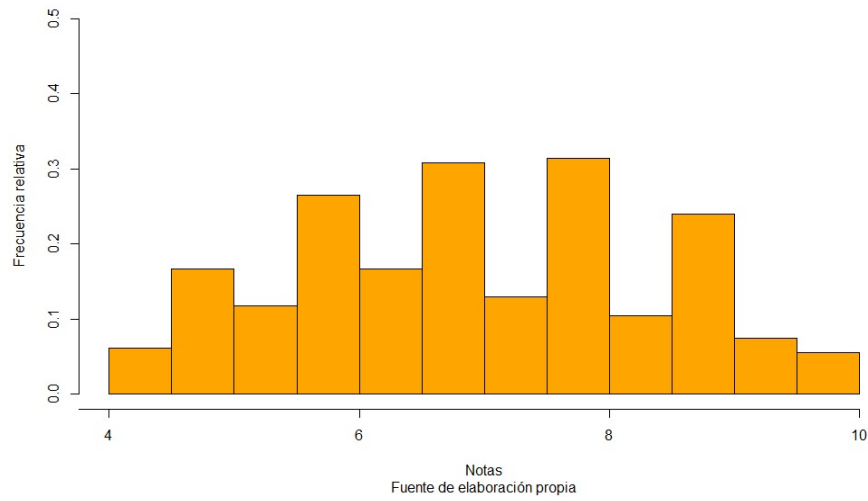
Fuente de elaboración propia

**Boxplot para la asignatura Lengua y Literatura de 4º año de Secundaria**

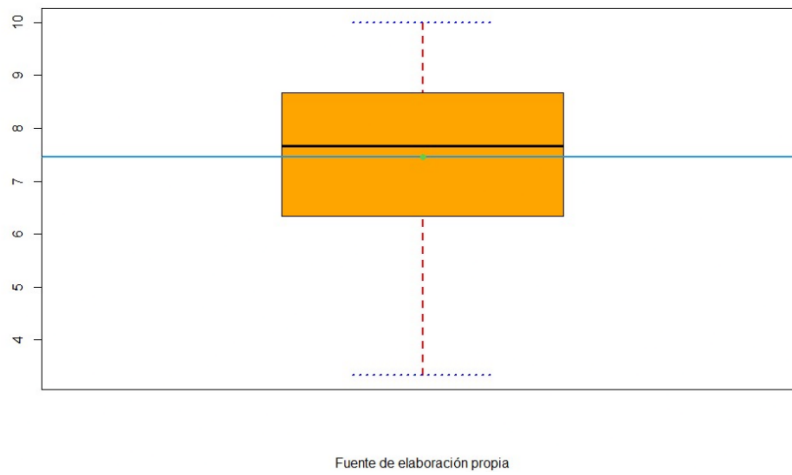


Fuente de elaboración propia

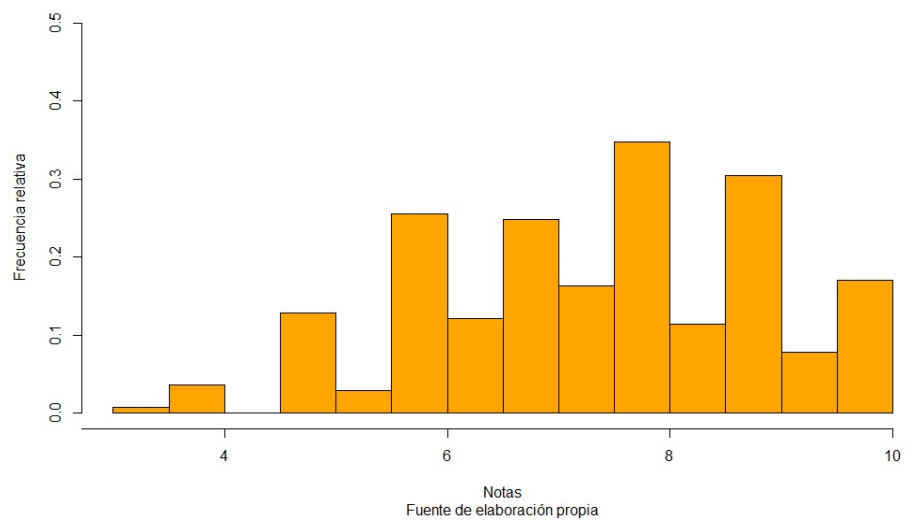
**Histograma para la asignatura Lengua y Literatura de 4° año de Secundaria**



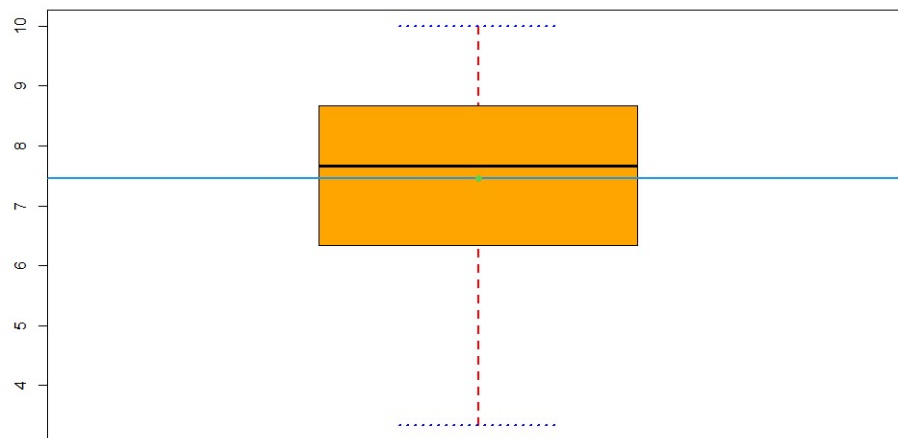
**Boxplot para la asignatura Literatura Latinoamericana de 5° año de Secundaria**



**Histograma para la asignatura Literatura Latinoamericana de 5° año de Secundaria**

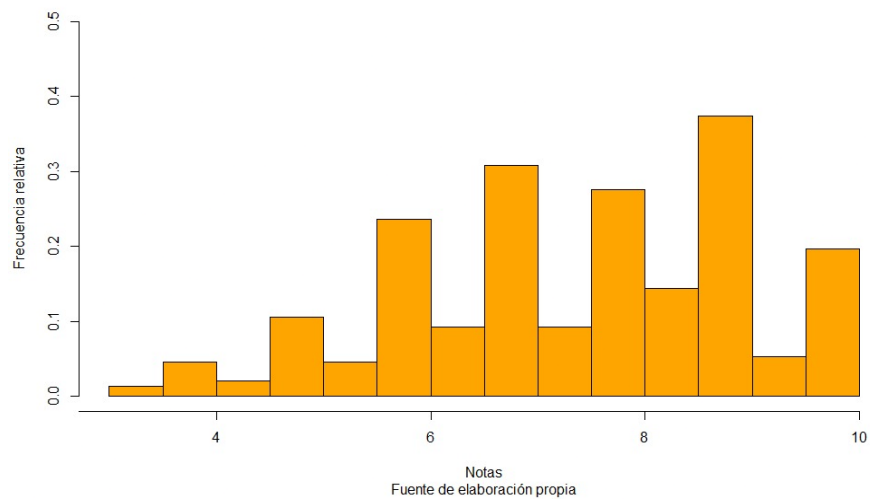


**Boxplot para la asignatura Literatura Argentina de 6° año de Secundaria**



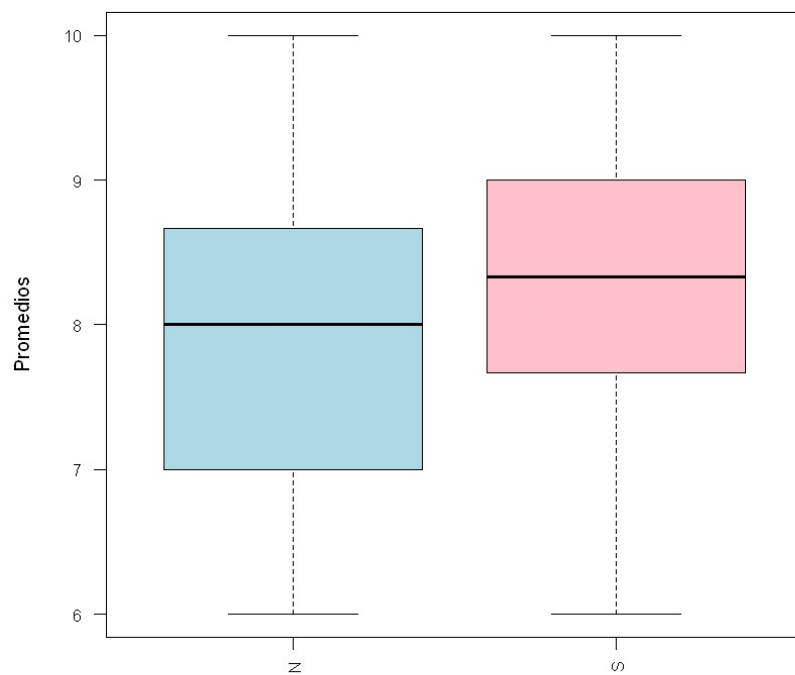
Fuente de elaboración propia

**Histograma para la asignatura Literatura Argentina de 6° año de Secundaria**



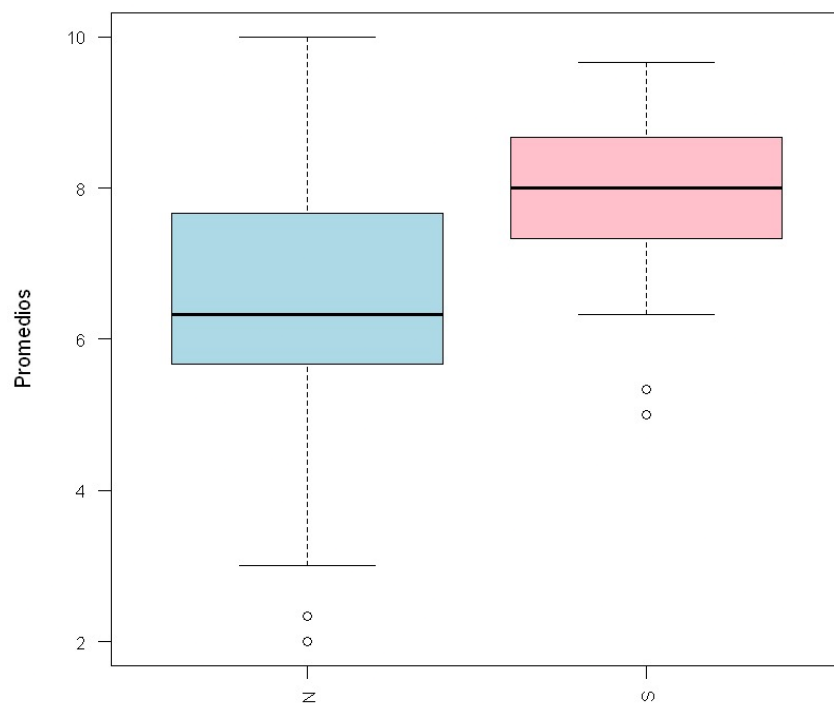
Fuente de elaboración propia

**Promedios de Lengua: 4º grado en según tipo de gestión**



Gestión  
Fuente de elaboración propia

**Promedios de Lengua: 4º año en según tipo de gestión**



Gestión  
Fuente de elaboración propia



