



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 Práctico parte 2

**Tema: Primer informe de las notas de las escuelas del
departamento La Paz**

Fecha de Entrega: 14/06/2023

Profesores: Esp. Prof. Melisa Fernández

Alumnos: Venturini, Angelo

Ruiz Diaz, Enzo

Informe de las notas de las escuelas del departamento La Paz

ÍNDICE:

Introducción	3
Desarrollo	4
Exploración	4
Reorganización y limpieza	9
Análisis de las características generales de la educación	10
Rendimiento académico de matemática y lengua segmentado por año académico.	16
Primaria: matemática	17
Primaria: Lengua	18
Secundaria: matemática	19
Secundaria: Lengua y literatura	19
Correlaciones	22
Comparación de Medias	26
Análisis Bivariado Cualitativo	27
Análisis Multivariado Cuantitativo	28
Conclusiones	29
Anexo	31

Introducción

En el presente informe encontrarán un análisis de una base de datos que contiene las escuelas primarias y secundarias de educación común y técnica del departamento La Paz de la provincia de Entre Ríos. El análisis tiene el fin de diagnosticar el estado general de la educación en dicho departamento, y en particular analizar el rendimiento de los estudiantes en las materias básicas de secundaria (Matemática, Lengua, Geografía, Historia, Biología, Física, Química, Inglés, Educación Física) y primaria cursadas en el año 2022, con la idea de asesorar en la implementación de programas complementarios.

Se realizó el análisis con la herramienta Jupyter Notebook y el lenguaje de programación R. Además para analizar la representatividad se utilizaron dos bases de datos que contienen la totalidad de escuelas primarias y secundarias de la provincia. Dichas bases de datos fueron proporcionadas por la cátedra.

Desarrollo

Exploración

En una primera instancia el grupo realizó una exploración del dataset donde se reconoció la dimensión del mismo, se analizaron los nombres y tipo de las variables, como también se analizaron los tipos de los datos cargados. Por otra parte se verificó la existencia de datos ausentes, datos duplicados y datos mal cargados. En otra etapa de la exploración, analizamos si había variables redundantes o que no aporten información al problema. En la tabla 1 se muestra la información obtenida del proceso de exploración.

Tabla 1

Análisis exploratorio

Departamento:	La Paz
Tamaño	12127150
N.º de filas	485086
N.º de columnas	25
Variables Cuantitativas	idalumno, documento, anioLectivo, nota, idSuborganizacion, idDivision, idorganizaciones,
Variables Cualitativas	CUE, orden, nivel, EsPrivada, turno, AñoCursado, periodoEvaluatorio, asignatura, NivelEnsenanza, ddivision, esMultianio, Modalidad, ModEnsenaza
Variables a modificar nombre	anioLectivo, NivelEnsenaza, ModEnsenaza
Variables mal declaradas	curso, turno, Modalidad, nota, EsPrivada, ModEnsenansa
Variables redundantes	idalumno
Variables que no aportan información	idSuborganizacion, idDivision, idorganizaciones, esMultianio, motivo_ausente
Variables con NA	motivo_ausente, observaciones
Variables con datos sucios	AñoCursado, asignatura
Variables importantes	documento, nota, periodoEvaluatorio, asignatura

Además, se realizó un conteo de escuelas y de matrículas donde obtuvimos para escuelas primarias, 49 escuelas y 5151 matrículas luego estudiamos la representatividad de la muestra, donde utilizando una calculadora de tamaño de muestra (SurveyMonkey) para una población de 56 escuelas primarias, información obtenida del dataset común proporcionado por la cátedra, con un nivel de confianza del 95% se necesita un tamaño de muestra de 49 y para una población de 5987 matrículas con un nivel de confianza del 95% se necesita un tamaño de muestra de 362. Contrastando esta información con nuestra muestra concluimos que es representativa.

Para escuelas secundarias encontramos 22 escuelas y 4564 matrículas.

Tabla 2

Estudio de la representatividad de la muestra de escuelas secundarias

SECUNDARIA							
Cantidad de escuelas: 22				Cantidad de escuelas cargadas: 22			
	1°	2°	3°	4°	5°	6°	7°
Matrícula total	1067	1078	1000	1036	885	819	101
Matrículas cargadas	892	844	849	763	598	561	57
Tamaño mínimo de muestra	283	284	278	281	269	262	81

Como podemos ver en la tabla 2, la muestra para séptimo año no es representativa por lo que no tendremos en cuenta este año para los posteriores análisis.

También realizamos un resumen de los estadísticos para todos los cursos de primaria de las materias matemática y lengua.

1°	conteo	media	varianza	desvío	min	max	IQR	rango	mediana
Lengua	985	8,36	1,29	1,14	6	10	1,67	4	8,67
Matemática	1010	8,47	1,25	1,12	6	10	1,67	4	8,67

2°	conteo	media	varianza	desvío	min	max	IQR	rango	mediana
Lengua	862	8,5	1,30	1,14	6	10	1,67	4	8,67
Matemática	830	8,49	1,35	1,16	6	10	1,67	4	8,67

3°	conteo	media	varianza	desvío	min	max	IQR	rango	mediana
Lengua	867	8,53	1,15	1,07	6	10	1,67	4	8,67
Matemática	866	8,56	1,29	1,14	6	10	2	4	8,67

4°	conteo	media	varianza	desvío	min	max	IQR	rango	mediana
Lengua	856	8,27	1,11	1,05	6	10	1,33	4	8,33
Matemática	851	8,31	0,994	0,997	6	10	1,33	4	8,33

5°	conteo	media	varianza	desvío	min	max	IQR	rango	mediana
Lengua	824	8,16	1,13	1,06	6	10	1,67	4	8,33
Matemática	820	8,16	1,23	1,11	6	10	1,67	4	8,33

6°	conteo	media	varianza	desvío	min	max	IQR	rango	mediana
Lengua	838	8,09	1,14	1,07	6	10	1,67	4	8
Matemática	835	8,12	1,24	1,11	6	10	1,67	4	8

Ahora lo realizamos para los años de las escuelas secundarias

1°	conteo	media	varianza	desvío	min	max	IQR	rango	mediana
Lengua y Literatura	850	6,95	2,59	1,61	2,33	10	2,33	7,67	7
Matemática	887	6,8	2,98	1,73	1	10	2	9	6,67

2°	conteo	media	varianza	desvío	min	max	IQR	rango	mediana
Lengua y Literatura	756	6,34	3,93	1,98	1	10	2,67	9	6,5
Matemática	843	6,32	3,69	1,92	1,67	10	3	8,33	6,33

3°	conteo	media	varianza	desvío	min	max	IQR	rango	mediana
Lengua y Literatura	918	6,51	3,40	1,84	1	10	2,5	9	6,33
Matemática	936	6,23	3,93	1,98	1	10	2,67	9	6,33

4°	conteo	media	varianza	desvío	min	max	IQR	rango	mediana
Lengua y Literatura	1377	6,04	3,45	1,86	1	10	2,67	9	6
Matemática	1414	6,14	4,10	2,03	1	10	3	9	6,33

5°	conteo	media	varianza	desvío	min	max	IQR	rango	mediana
Lengua y Literatura	131	7,52	2,14	1,46	2,33	10	2,33	7,67	7,67
Matemática	1548	6,63	3,10	1,76	1	10	2,67	8,33	6,67

6°	conteo	media	varianza	desvío	min	max	IQR	rango	mediana
Lengua y Literatura	9	8,30	0,457	0,676	7	9	1	2	8
Matemática	1451	6,91	3,30	1,82	1	10	2,67	9	7

Reorganización y limpieza

Luego de la exploración de la base de datos se concluye que el estado general de la misma es mala y no se podría realizar el análisis encargado sin antes realizar ciertas modificaciones. Estas incluyen renombrar variables para que se entienda la información que estas contienen, cambiar el formato de los valores para poder procesar dichos datos, normalizar ciertos datos de tipo string o char para que haya homogeneidad y agregar variables nuevas que ayuden al análisis como es el caso de la variable promedio y curso.

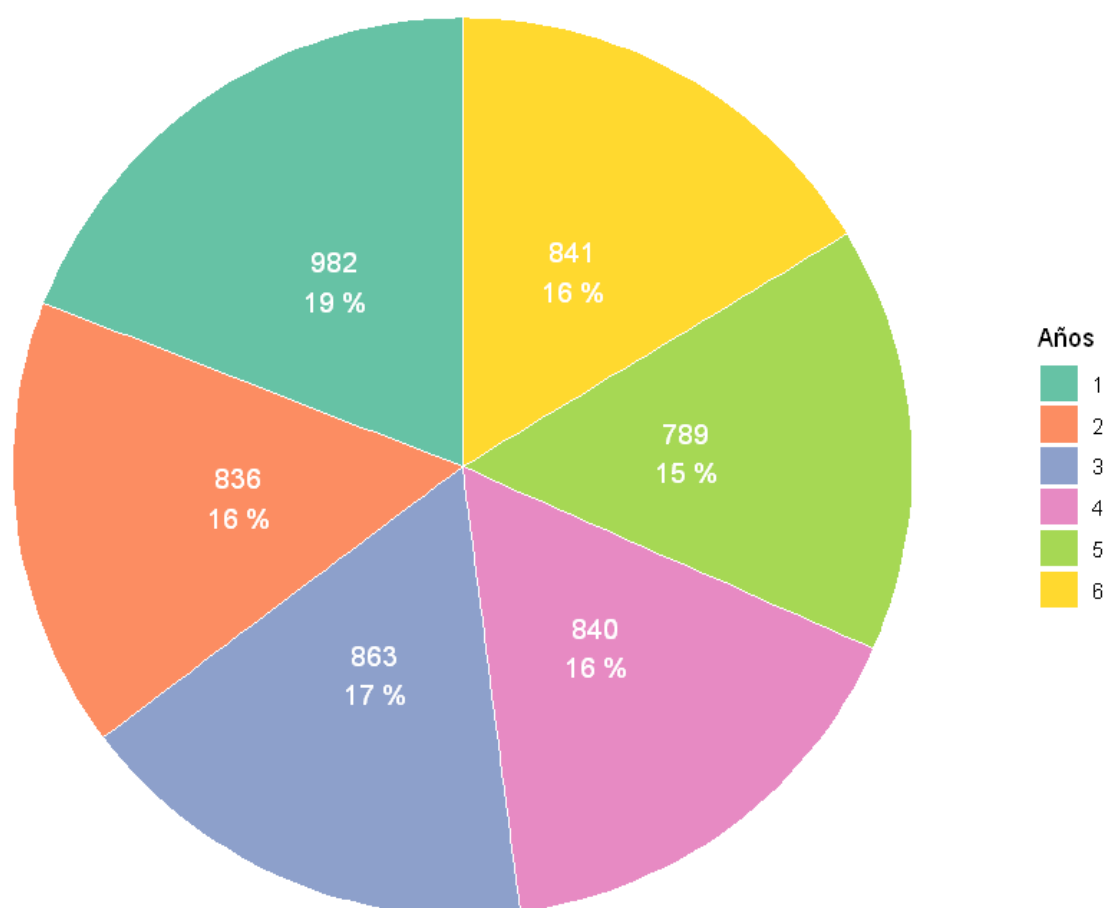
Encontramos que la columna nivel enseñanza está relacionada con la columna nivel, con la diferencia de que esta última nos brinda menos información por lo que decidimos eliminarla. También, eliminamos la columna de año ya que todos nuestros datos son del año 2022 y sería redundante dejarla. Adicionalmente, solo se dejaron las filas cuyos promedios pudieron ser calculados con las principales materias troncales de cada nivel de enseñanza, siendo para primaria ciencias naturales, ciencias sociales, matemática, lengua, educación física e inglés. Para secundaria, matemática, lengua y literatura, educación física, geografía, historia, biología, físico-química, química, física e inglés.

Análisis de las características generales de la educación

Primeramente analizamos la distribución de la matrícula, segmentada por año, de las escuelas primarias. Como se observa en el gráfico 1, el mayor número de matriculados pertenecen al segmento de primer año con 19 %, seguido de tercer grado con 17%,

Gráfico 1

Distribucion Porcentual de la Matricula de Primaria de 2022 en el Departamento La Paz por Año de Cursado

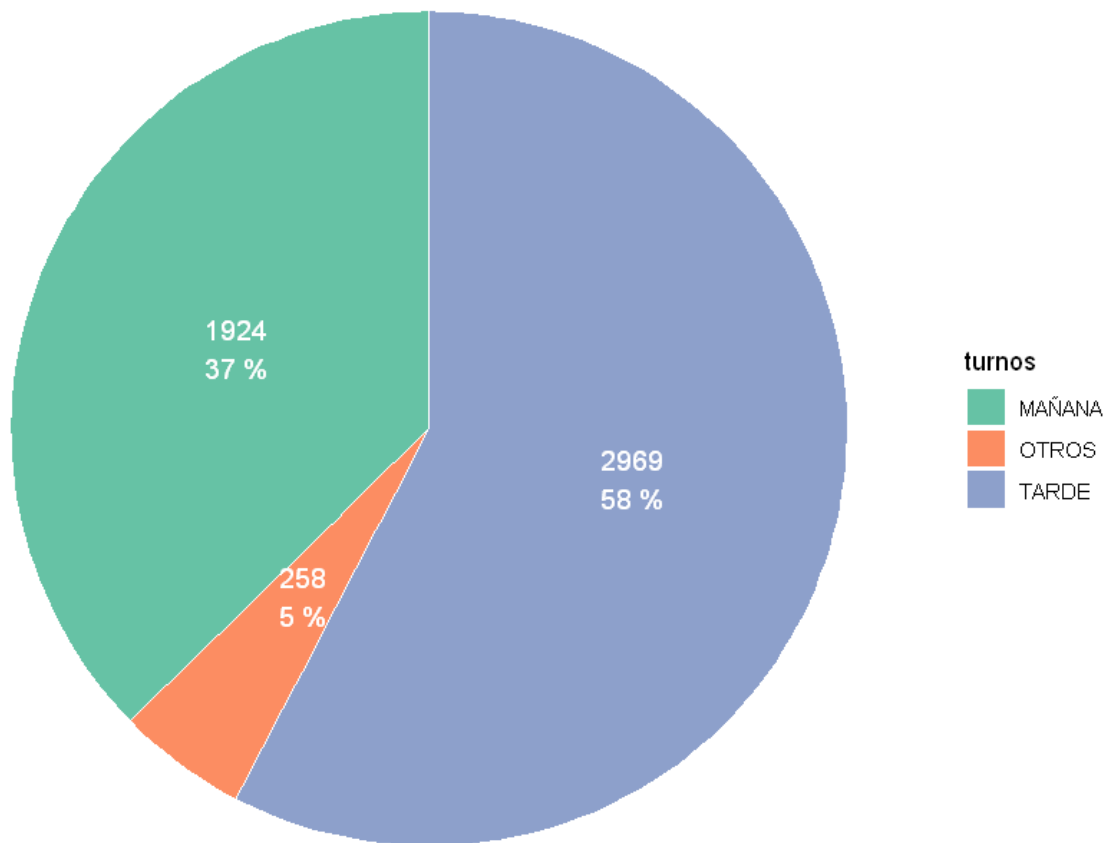


Fuente: autoría propia

Además analizamos su distribución segmentada por turnos, donde agrupamos los turnos “Intermedio” (frecuencia 10, porcentaje 0%), “Completo” (frecuencia 631, porcentaje 2%) y “Rotativo” (frecuencia 785, porcentaje 3%) en el segmento “Otros” para una mejor lectura del gráfico 2, ya que los valores porcentuales son muy pequeños y se superponen entre sí. En este podemos observar que la mayoría de estudiantes pertenecen al turno tarde con 57%, seguido del turno mañana con 38% y por último otros con 5%.

Gráfico 2

Distribucion Porcentual de la Matricula de Primaria de 2022 en el Departamento La Paz por Turnos

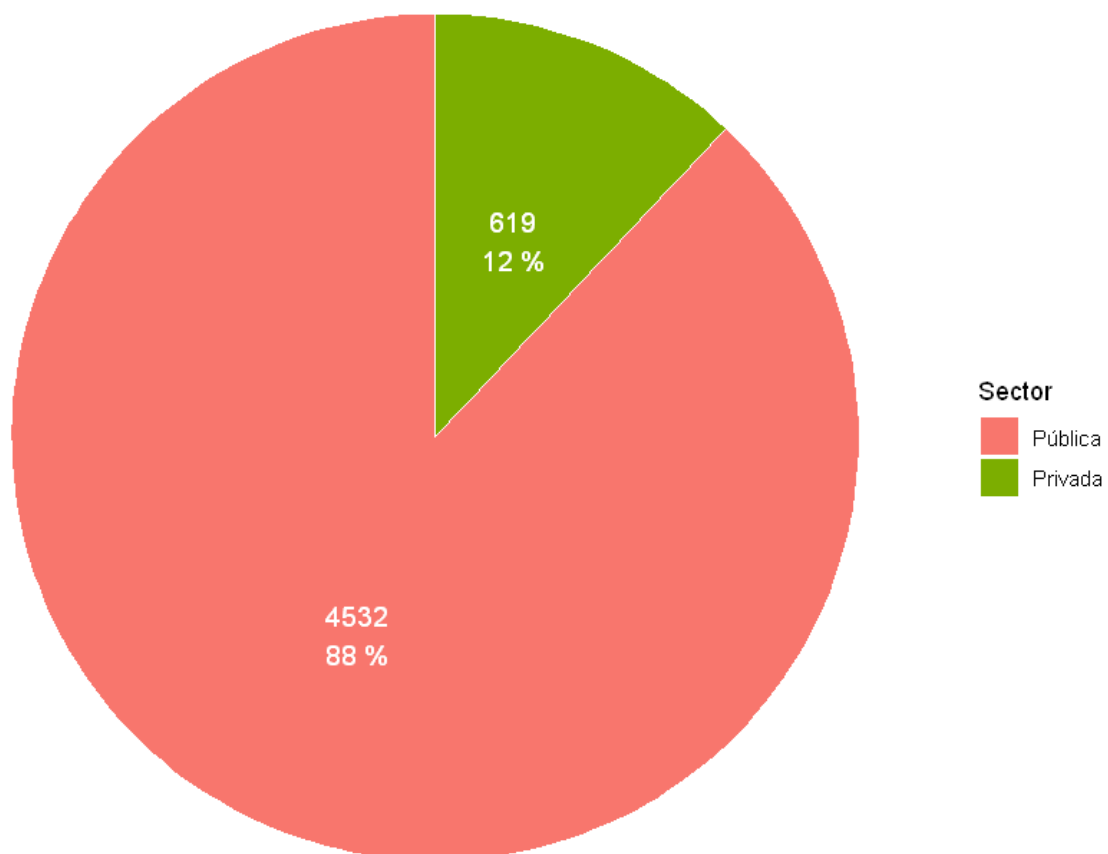


Fuente: autoría propia

Por último, analizamos la distribución segmentada por escuelas públicas y privadas de las notas cargadas, donde encontramos que el 88% pertenecen al sector público.

Gráfico 3

Distribucion Porcentual de Estudiantes de Primaria Segmentado por Sector



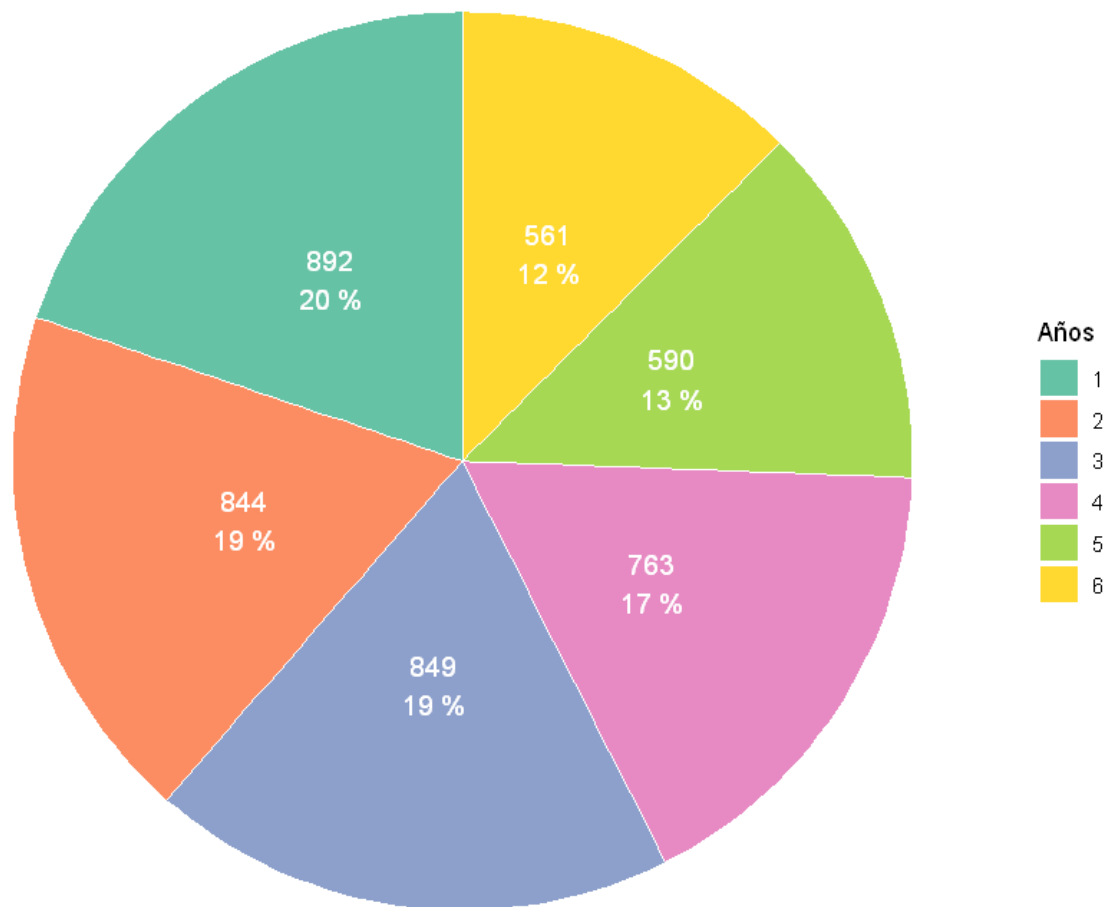
Fuente: autoría propia

Ahora procedemos con el mismo análisis para las escuelas secundarias.

Podemos ver en el gráfico 4 que el mayor porcentaje se encuentra en cuarto año con 22% seguido por quinto año con 20% , tercer y sexto año con 15 % y por último segundo y primer año con 14 %. También podemos notar la nula injerencia porcentual de los alumnos de séptimo año de escuelas técnicas debido a su muy baja cantidad de matrículas

Gráfico 4

Distribucion Porcentual de la Matricula de secundaria de 2022 en departamento La Paz por Año de Cursado

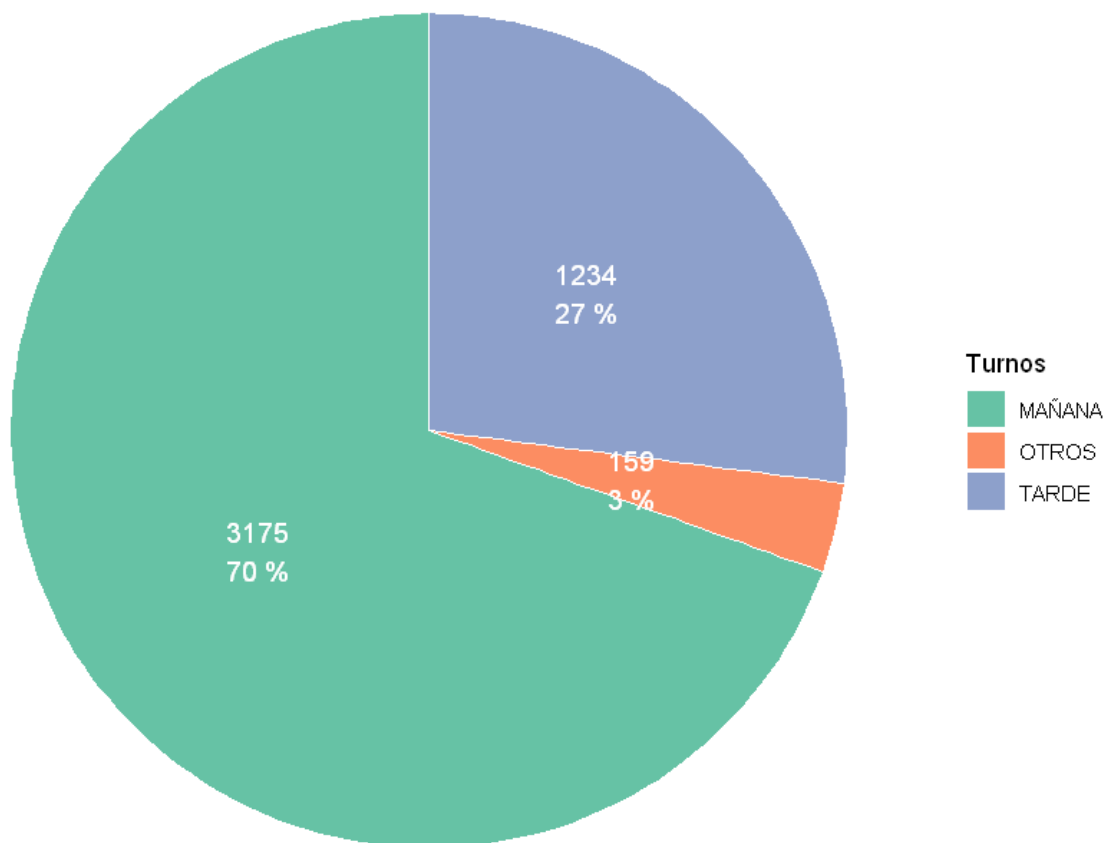


Fuente: autoría propia

Para la distribución por turnos de escuelas secundarias, para una mejor legibilidad del gráfico 5 agrupamos los turnos diurno, noche, rotativo y vespertino en la categoría otros. Podemos observar en el gráfico 5, al contrario de lo que sucedía con las escuelas primarias, el turno predominante es el turno mañana con 77 %.

Gráfico 5

Distribucion Porcentual de la Matricula de Secundaria de 2022 en el Departamento La Paz por Turnos



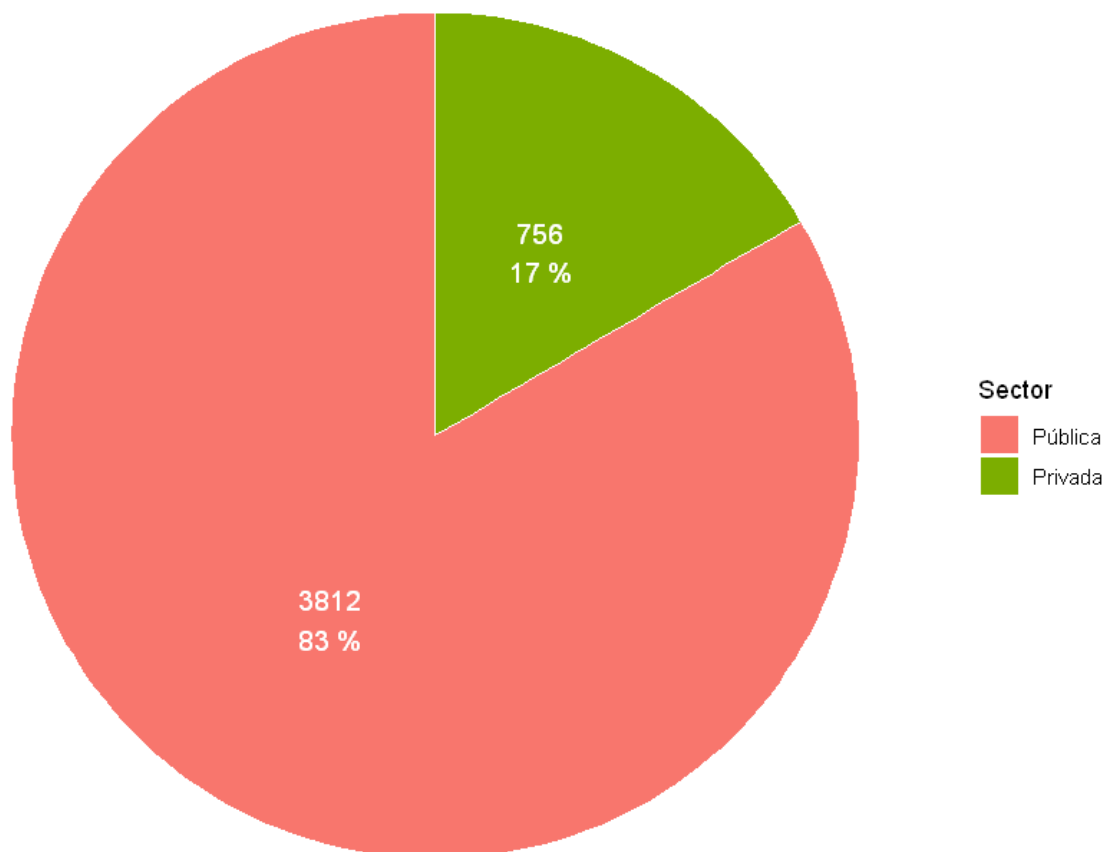
Fuente: Autoría propia

Fu

En el gráfico 6 podemos observar la distribución de las notas cargadas para escuelas secundarias segmentada por sector público y privado. Nótese que la distribución es parecida a la distribución de escuelas primarias.

Gráfico 6

Distribución Porcentual de Estudiantes de Secundaria Segmentado por Sector



Fuente: autoría propia

Rendimiento académico de matemática y lengua segmentado por año académico.

En el apartado anexo se encuentran los gráficos 7 y 8 (1°), 9 y 10 (2°), 11 y 12 (3°), 13 y 14 (4°), 15 y 16 (5°), 17 y 18 (6°) correspondientes a los histogramas segmentados por año académico para el promedio de matemática y lengua de las escuelas primarias.

Pasando a las escuelas secundarias en el apartado anexo se encuentran los gráficos 19 y 20 (1 año), 21 y 22 (2 año), 23 y 24 (3 año), 25 y 26 (4 año), 27 y 28 (5 año), 29 y 30 (6 año) correspondientes a los histogramas segmentados por año académico para el promedio de matemática y lengua y literatura de las escuelas secundarias.

Primaria: matemática

Tabla 3

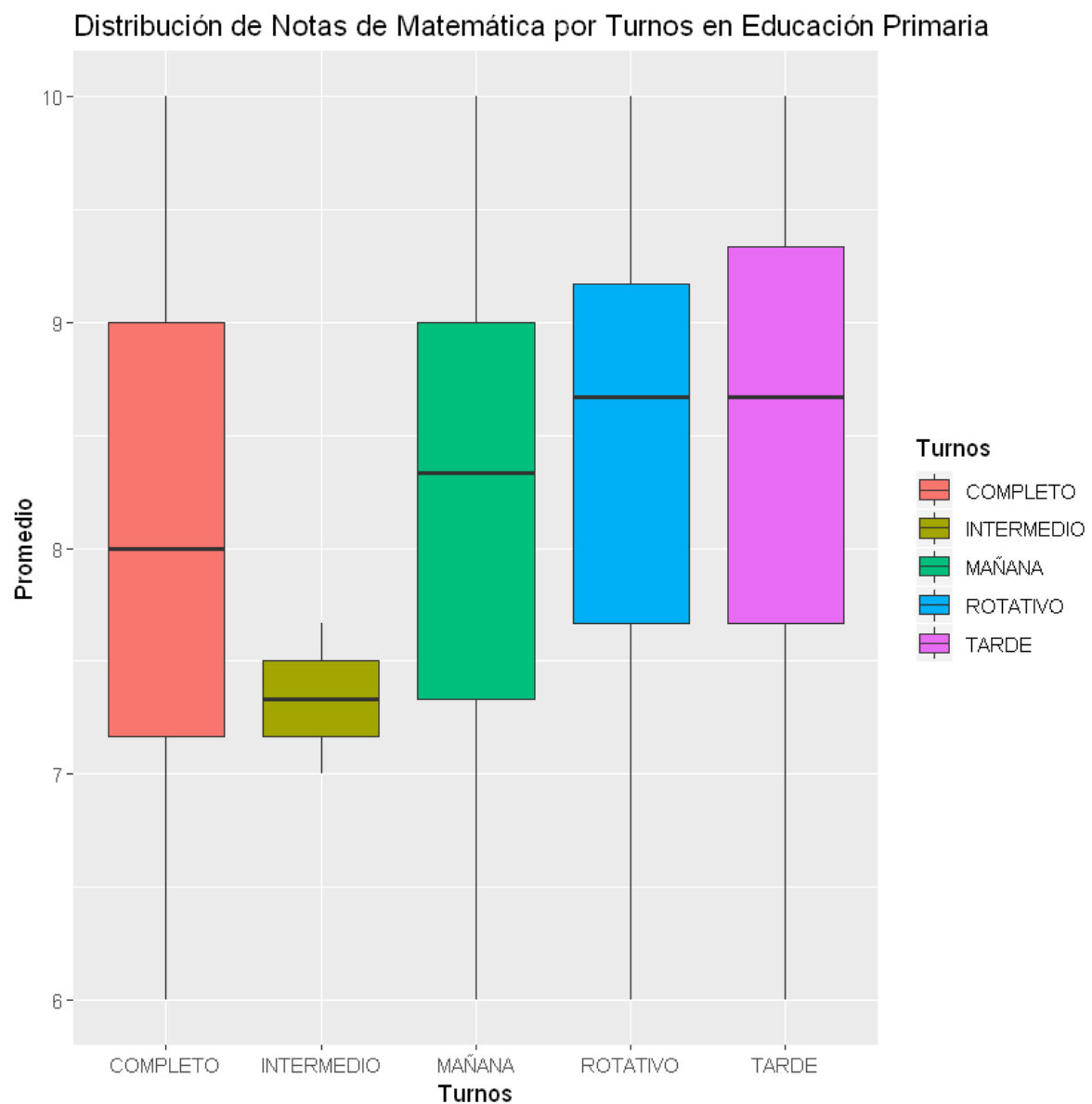
Tabla de frecuencia de promedios segmentado en turnos para la materia de matemáticas

COMPLETO	INTERMEDIO	MAÑANA	ROTATIVO	TARDE
118	2	1936	157	2999

En el gráfico 31 podemos observar que el rango, exceptuando el turno intermedio, se encuentra entre 6 y 10. También se observa que la mediana se encuentra entre [8,9) para todos los turnos exceptuando intermedio.

Gráfico 31

Distribución de notas de Matemáticas por Turnos para todos los años de Primaria



Primaria: Lengua

Tabla 4

Tabla de frecuencia de promedios segmentado en turnos para la materia de lengua

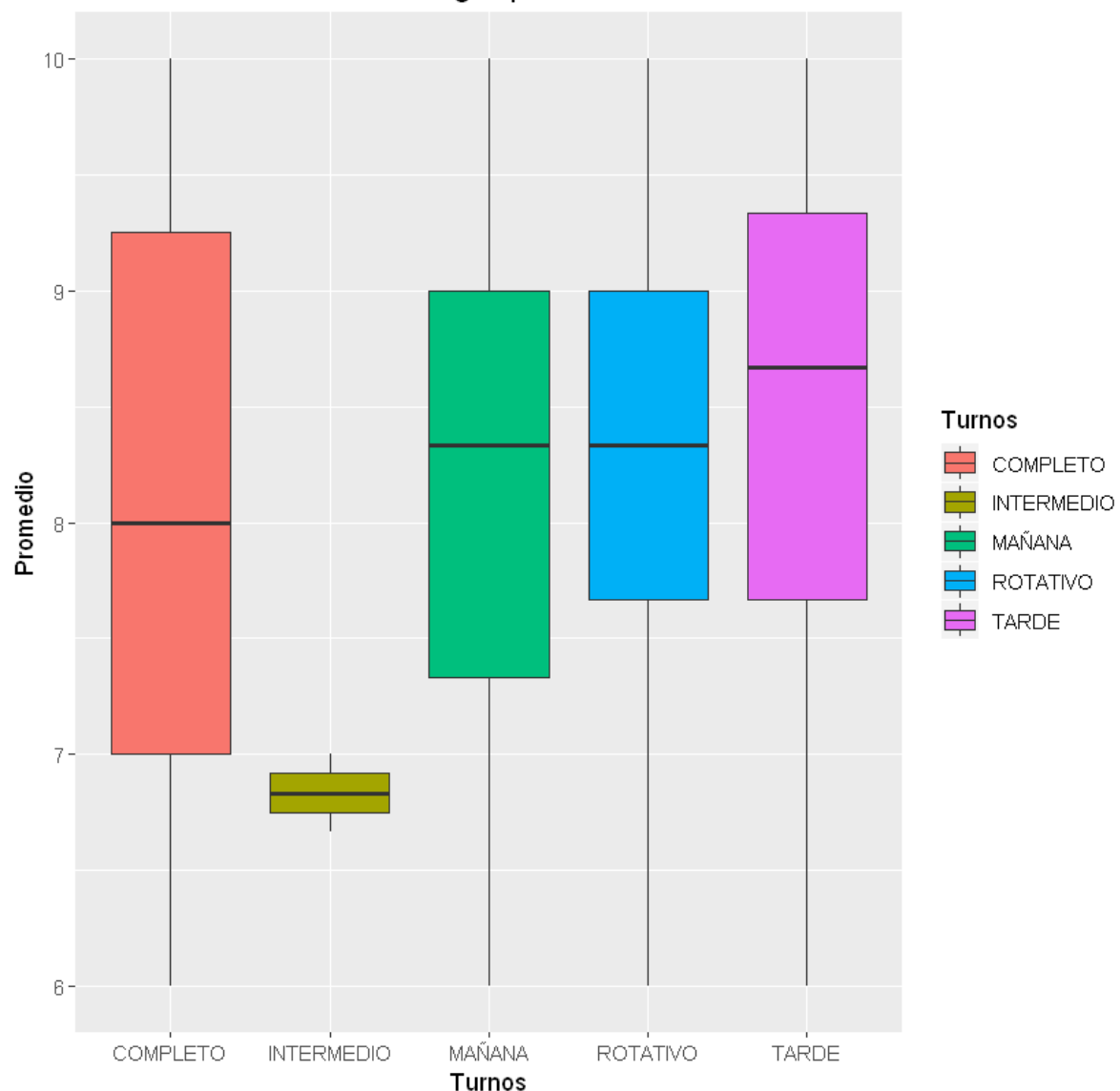
COMPLETO	INTERMEDIO	MAÑANA	ROTATIVO	TARDE
118	2	1964	158	2990

En el gráfico 32 vemos que, al igual que sucedía con los promedios de matemática, la mediana para todos los turnos se encuentra entre [8,9), exceptuando el turno intermedio.

Gráfico 32

Distribución de notas de Lengua por Turnos para todos los años de Primaria

Distribución de Notas de Lengua por Turnos en Educación Primaria



Podemos observar en la tabla 3 como en la tabla 4, que la cantidad de notas cargadas para los turnos “Completo”, “Intermedio” y “Rotativo” no es significativa a comparación del turno

“Mañana” y “Tarde”, por lo que la información de ellos en estos gráficos puede ser imprecisa.

Secundaria: matemática

Tabla 5

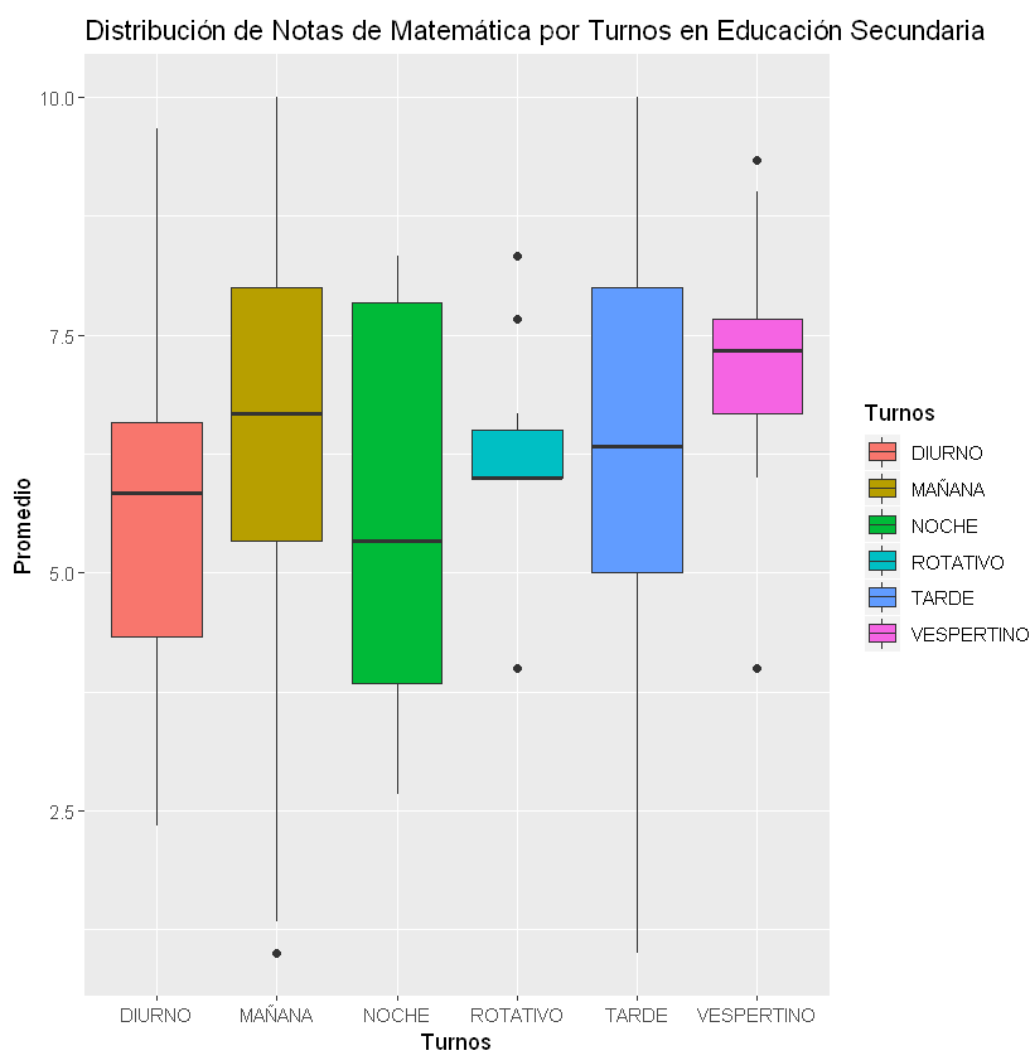
Tabla de frecuencia de promedios segmentado en turnos para la materia de matemática

DIURNO	MAÑANA	NOCHE	ROTATIVO	TARDE	VESPERTINO
91	5727	16	11	1259	21

El gráfico 33 muestra la distribución de los promedios de matemática segmentado por turnos donde podemos notar que el turno mañana y tarde hay más variabilidad. En el turno mañana el 50% tiene un promedio de 6,5 aproximadamente, mientras que el turno de la tarde tiene un promedio de 6 aproximadamente.

Gráfico 33

Distribución de notas de Matemática por Turnos de todos los años de Secundaria



Secundaria: Lengua y literatura

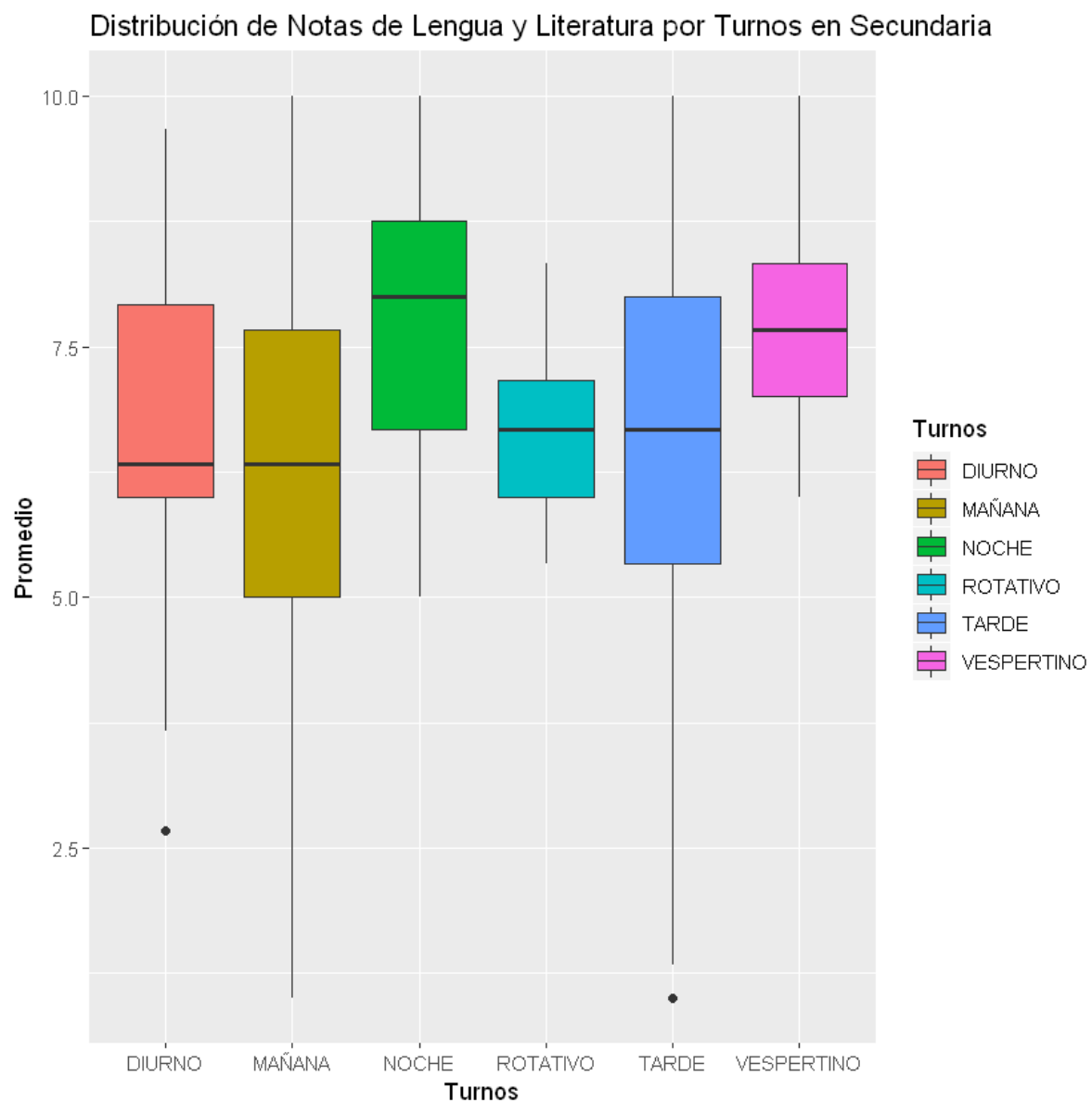
Tabla 6

Tabla de frecuencia de promedios segmentado en turnos para la materia de lengua y literatura

DIURNO	MAÑANA	NOCHE	ROTATIVO	TARDE	VESPERTINO
86	2829	52	11	1055	21

Gráfico 34

Distribución de notas de Lengua y Literatura por Turnos de todos los años de Secundaria



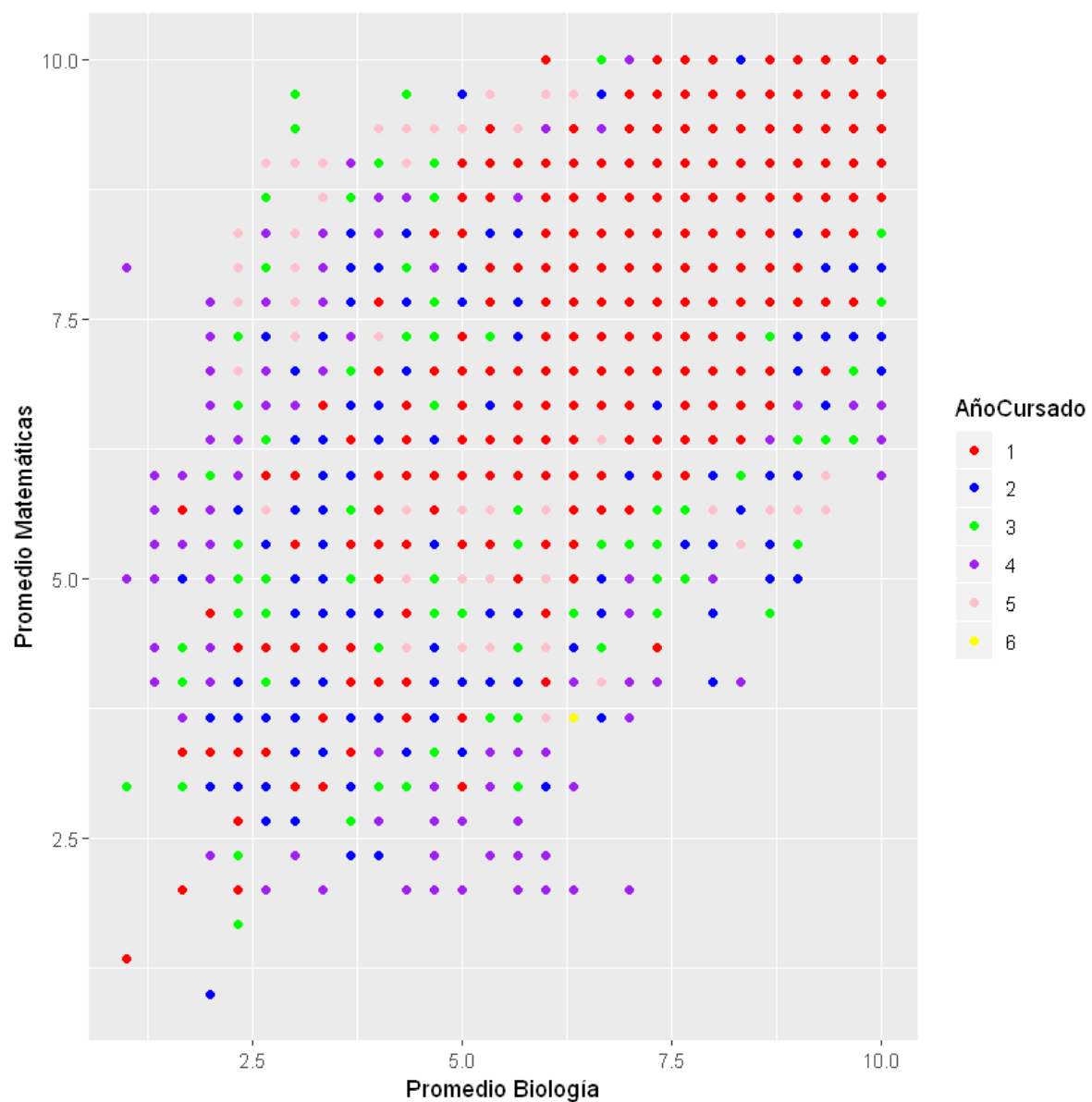
Podemos observar en la tabla 5 como en la tabla 6, que la cantidad de notas cargadas para los turnos “Diurno”, “Noche”, “Rotativo” y “Vespertino” no es significativa a comparación del

turno “Mañana” y “Tarde”, por lo que la información del gráfico 34 para estos turnos puede ser imprecisa. De todas maneras, podemos afirmar que el 25% de las notas del turno tarde se encuentra por debajo del 6, mientras que el 25% del turno mañana presenta notas menores a 5.

Correlaciones

Se plantea como hipótesis que si el rendimiento de un alumno de un mismo año en la materia de Matemáticas es bueno, será de igual manera para la materia de Biología. Por consiguiente, realizamos un diagrama de dispersión entre los promedios de estas materias diferenciado por año para ver el comportamiento entre las dos variables, donde tomaremos solamente las escuelas secundarias ya que es el dataset que presenta más variación en los datos.

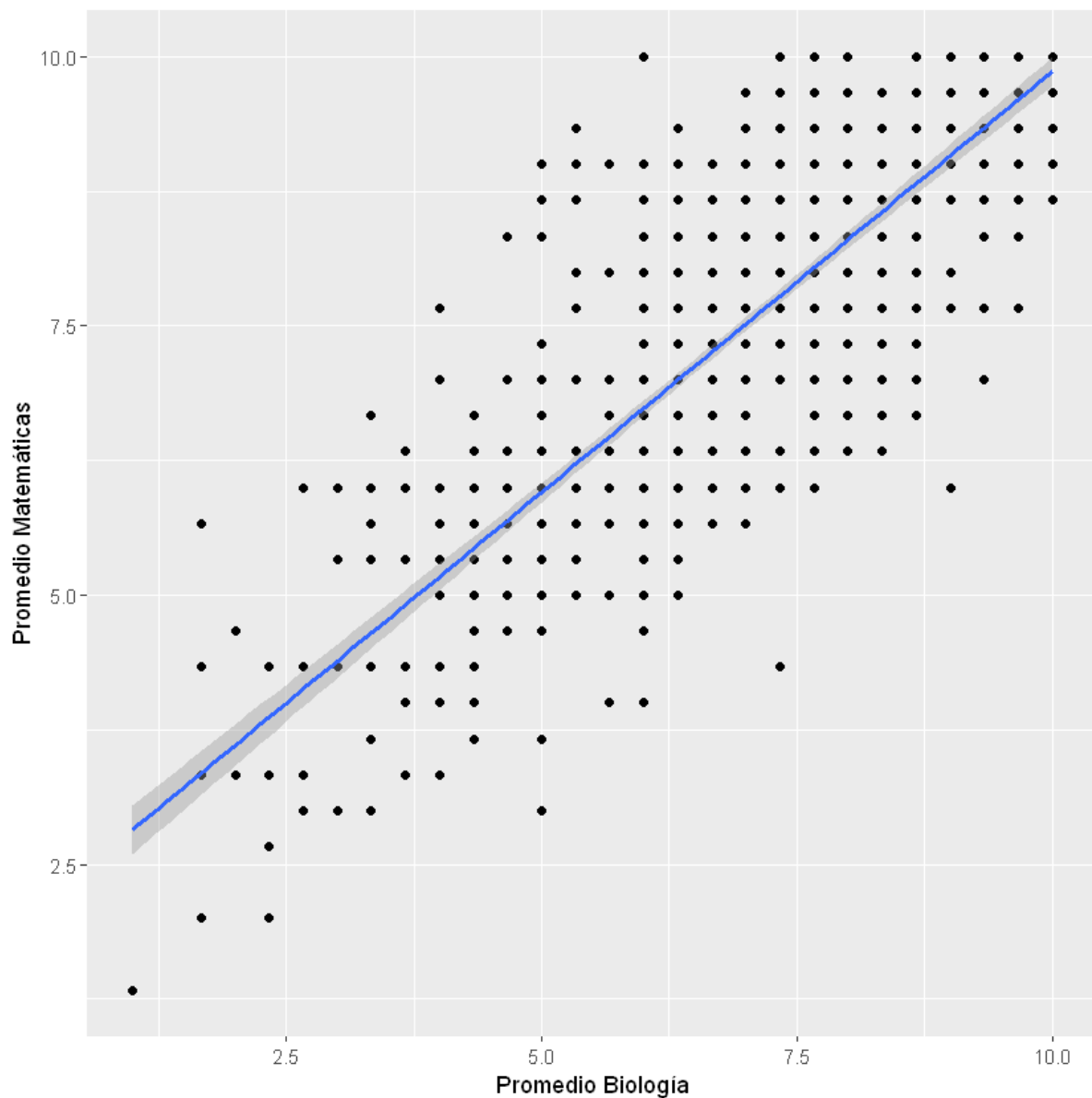
Gráfico 35



Como podemos observar en el Gráfico 35, el primer año de cursado es el que aparentemente presenta una relación lineal entre ambas variables, por lo que a partir de

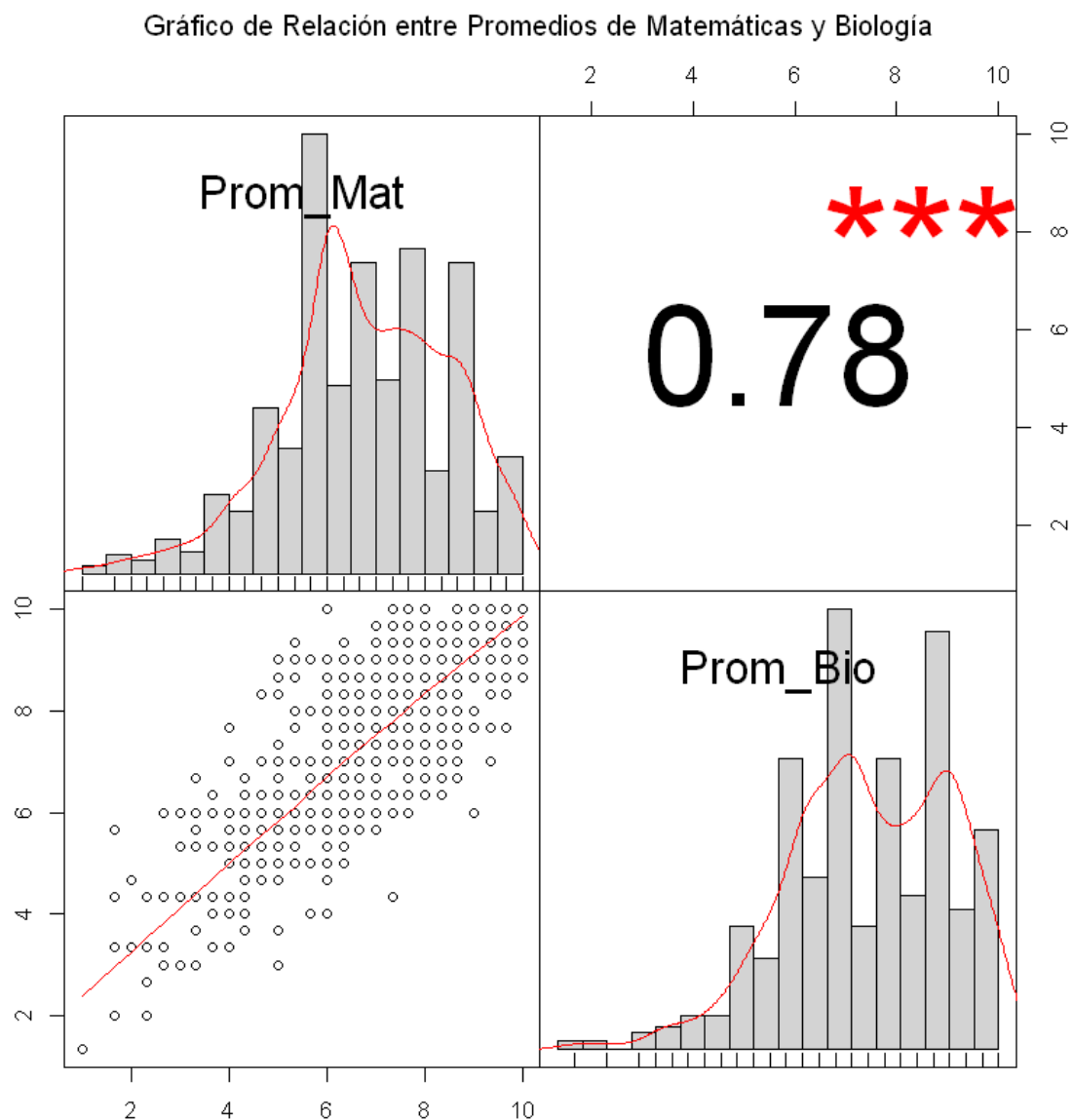
este momento, centraremos el estudio solamente en este año. Volvemos a graficar, pero esta vez solo con los datos del primer año (Gráfico 36).

Gráfico 36



Para verificar que efectivamente las variables estén correlacionadas, debemos realizar un test de normalidad para saber qué tipo de test utilizar para verificar la correlación, y como nuestra muestra tiene más de 50 elementos, utilizaremos el test de Lilliefors (Kolmogorov-Smirnov) para normalidad para cada una de las variables con una confianza del 95%. Al hacerlo, el test para ambas variables da como resultado un p-value menor a 0.05, por lo que podemos afirmar que las variables no siguen una distribución normal. Como consecuencia, utilizaremos el método de Spearman para comprobar la correlación entre las variables, y, para ello, lo haremos de forma gráfica.

Gráfico 37



Observando el Gráfico 37, el nivel de significancia es alto (***) y el ρ de Spearman da 0.78, por lo que se puede afirmar que tenemos una buena relación.

Ahora bien, procederemos a agregar una nueva variable a nuestro estudio, siendo esta Lengua y Literatura, para comprobar que la correlación encontrada no sea espuria. Para ello hallamos la matriz de correlaciones parciales utilizando los promedios de Matemáticas, Biología y Lengua y Literatura.

Gráfico 38

\$estimate

	Prom_Mat	Prom_Bio	Prom_Len
Prom_Mat	1.0000000	0.3266264	0.4784281
Prom_Bio	0.3266264	1.0000000	0.5280509
Prom_Len	0.4784281	0.5280509	1.0000000

\$p.value

	Prom_Mat	Prom_Bio	Prom_Len
Prom_Mat	0.000000e+00	6.389809e-30	1.159734e-66
Prom_Bio	6.389809e-30	0.000000e+00	2.358793e-83
Prom_Len	1.159734e-66	2.358793e-83	0.000000e+00

\$statistic

	Prom_Mat	Prom_Bio	Prom_Len
Prom_Mat	0.00000	11.69369	18.43581
Prom_Bio	11.69369	0.00000	21.04079
Prom_Len	18.43581	21.04079	0.00000

\$n

1148

\$gp

1

\$method

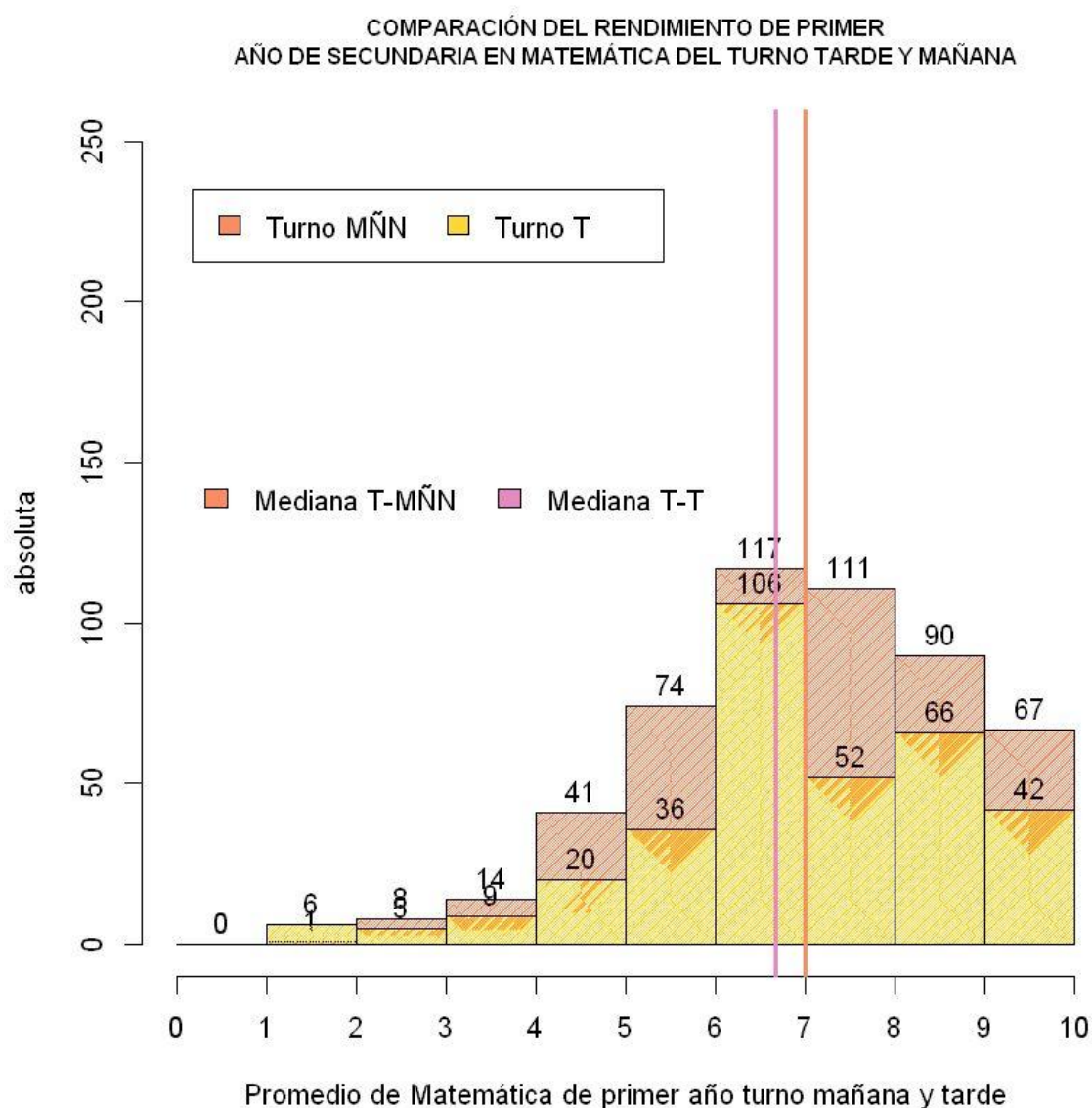
'spearman'

El resultado de esto puede verse en el Gráfico 38, donde podemos ver que al agregar Lengua y Literatura, la buena correlación que existía entre Matemáticas y Biología ahora es baja, por lo que podemos decir que la correlación estaba enmascarada con la relación de Lengua y Literatura. Queda por ver si la correlación entre Lengua y Biología, o entre Lengua y Matemáticas son significativas analizándolas por sí mismas.

Comparación de Medias

En esta sección utilizaremos comparaciones de medias y tipificación de valores con el fin de analizar y comprender las diferencias de las medias de variables entre los diferentes grupos, así como también evaluar la normalidad y la distribución de los datos de estos grupos. Esto nos permite detectar diferencias significativas y tomar decisiones informadas basadas en la evidencia estadística.

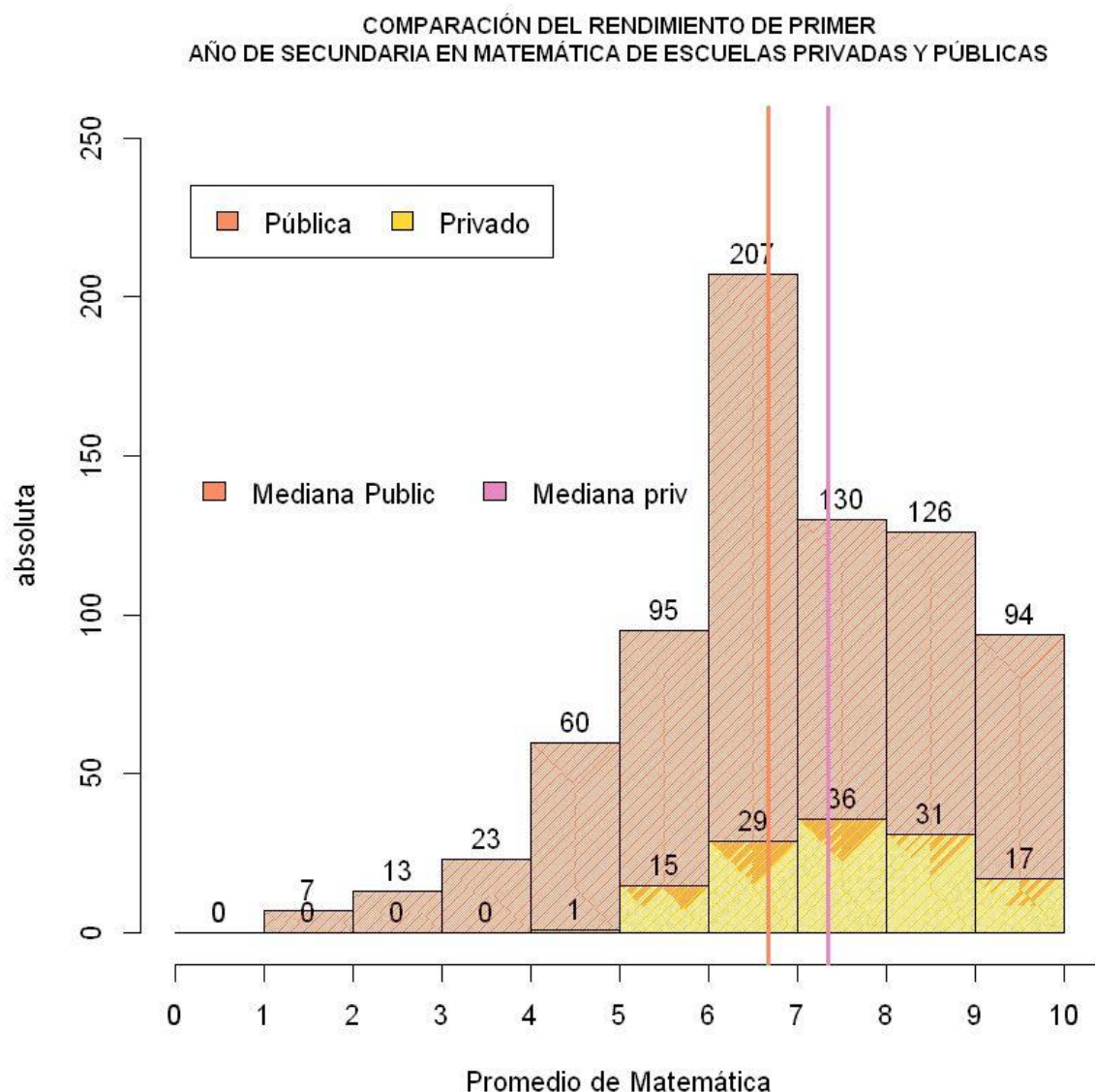
Gráfico 39



Se realizó una comparación de medias de los promedios de matemática de primer año respecto a los turnos, con la hipótesis de que generalmente el turno mañana obtiene mejor rendimiento que el turno tarde debido a una creencia popular, intentando corroborar si existen diferencias significativas. Dado que los datos no son normales, se realizó la prueba

no paramétrica de Wilcoxon, obteniendo un p valor de 0.95, por lo cual concluimos que no hay diferencias significativas entre las medidas centrales, por lo que no hay diferencia entre las notas de matemática del turno tarde con el de mañana de primer año. Esto puede verificarse en el gráfico 39, notando una muy pequeña diferencia entre las medianas graficadas y los histogramas que difieren casi únicamente en las frecuencias, pero de forma similar.

Gráfico 40



Se realizó una comparación de medias de los promedios de matemática de primer año respecto a la gestión, intentando corroborar si existen diferencias significativas. Dado que los datos no son normales, se realizó la prueba no paramétrica de Wilcoxon, obteniendo un p valor de 0.0001486, por lo cual concluimos que hay diferencias significativas entre las

medidas centrales, siendo las notas de matemática superiores en las escuelas de gestión privada sobre las públicas, que puede verse en el gráfico 40.

Luego, se realizó un análisis comparativo del desempeño en el área de Lengua y Matemática de los estudiantes de primaria, considerando los diferentes turnos de estudio. El objetivo principal es identificar si existen diferencias significativas en el rendimiento académico entre los distintos años y turnos en las asignaturas de Matemática y Lengua. Para este estudio se descartaron los turnos “Rotativo” e “Intermedio” por falta de frecuencias y por no estar en todos los años del estudio, dejando únicamente los turnos “Completo”, “Mañana” y “Tarde”.

Tabla 7

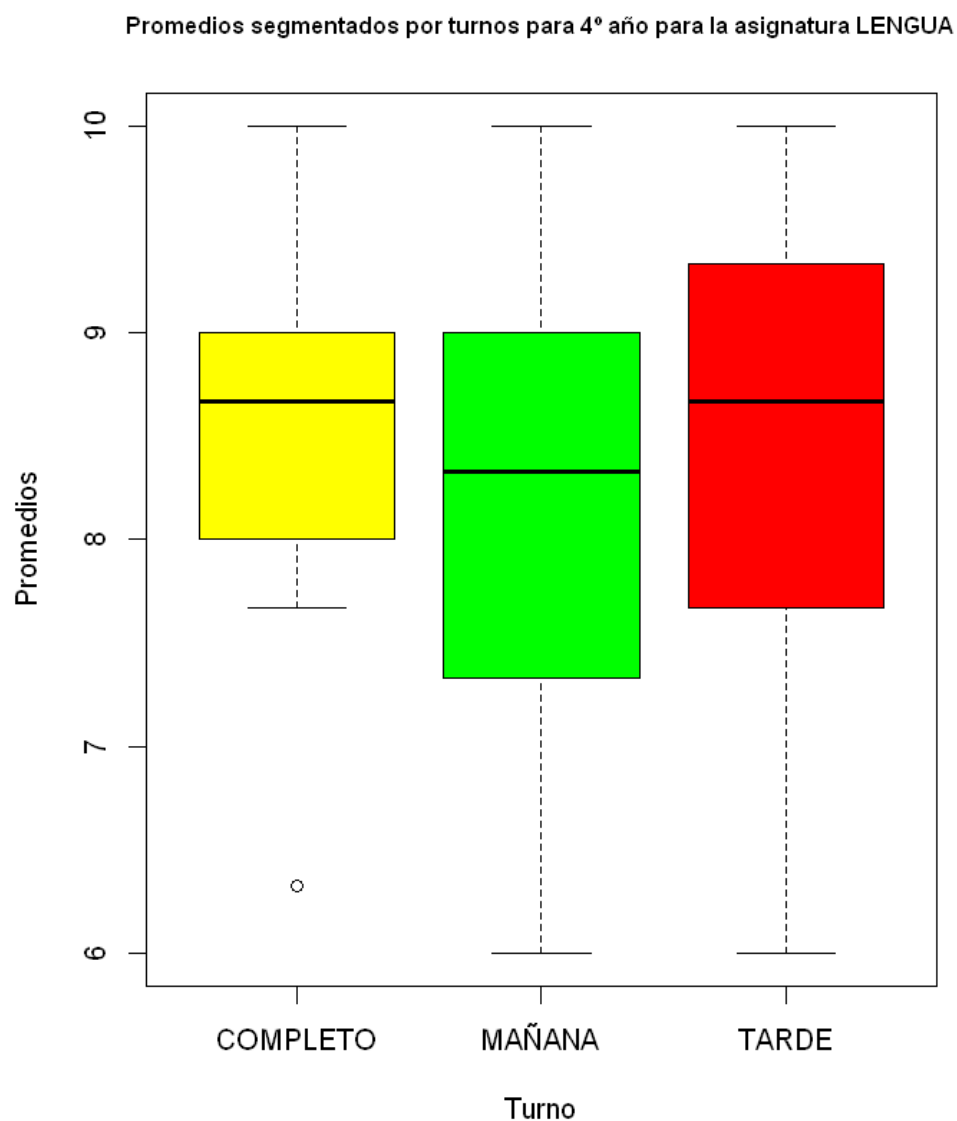
Descriptivos y frecuencias de cuarto, quinto y sexto año segmentados por turnos para la asignatura Lengua.

Año	Turno	Media	Mediana	N
4°	Completo	8,5	8,7	13
	Mañana	8,2	8,3	377
	Tarde	8,4	8,7	274
5°	Completo	7,6	7,7	16
	Mañana	8,1	8	439
	Tarde	8,3	8,3	257
6°	Completo	7,7	7,7	25
	Mañana	7,9	8	411
	Tarde	8,3	8,3	289

Solo se tomarán en cuenta en este estudio las notas de los cursos de cuarto, quinto y sexto año para la asignatura Lengua segmentado por turnos, ya que solamente en estos hay diferencias significativas. Inicialmente, en el análisis exploratorio de los datos nos encontramos con las frecuencias y los descriptivos dados en la tabla 7, los cuales parecen indicar una diferencia entre los estadísticos. Es curioso que las medianas entre los turnos de quinto y sexto sean iguales en cada turno, sin embargo, no pasa lo mismo con las medias.

Gráfico 41

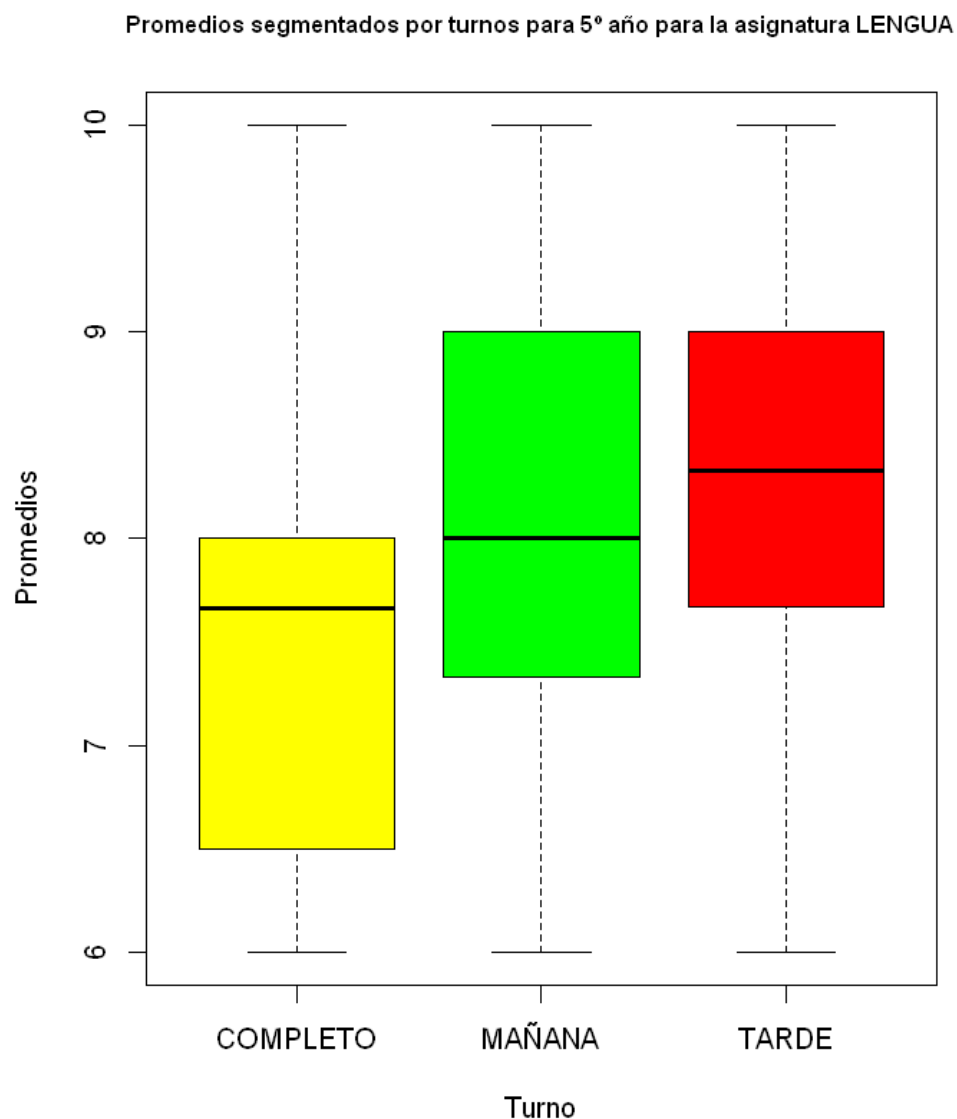
Boxplot exploratorio de los promedios segmentados por turno para cuarto año para la asignatura de Lengua.



En el gráfico 41, podemos ver una mayor variación de los datos para los turnos Tarde y Mañana a comparación del Completo, y observamos que el turno Mañana tiene la mediana más baja de los 3 turnos. Gráficamente podemos observar una pequeña diferencia entre el turno Mañana con los demás turnos.

Gráfico 42

Boxplot exploratorio de los promedios segmentados por turno para quinto año para la asignatura de Lengua.

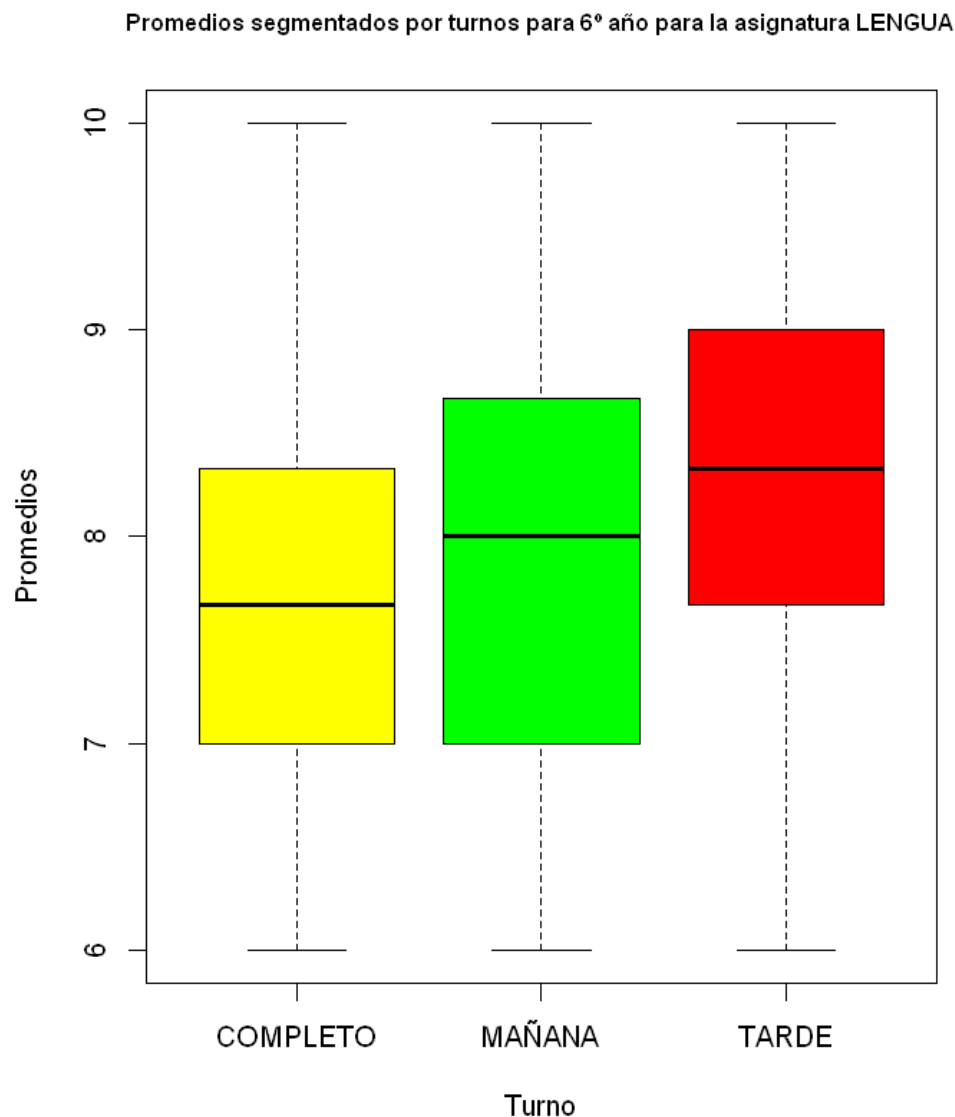


En el gráfico 42 podemos observar que la mediana del turno completo es la menor entre los tres turnos y el 75 % de los datos está por debajo de la mediana de los otros turnos. En este

caso, gráficamente podemos observar que hay una diferencia entre el turno Completo con los demás.

Gráfico 43

Boxplot exploratorio de los promedios segmentados por turno para sexto año para la asignatura de Lengua.



En el gráfico 43, observamos nuevamente que la mediana del turno completo es la menor de los 3 turnos y gráficamente observamos una diferencia entre los turnos completo y tarde.

Para el análisis de estos datos, se siguieron exactamente los pasos que encontramos en el ejercicio pertinente, donde primero realizamos el contraste ANOVA, planteamos las hipótesis, analizamos el resultado del test, en caso de encontrar diferencias usamos el test

de Tukey, acto seguido verificamos el cumplimiento de las condiciones del contraste ANOVA y, en caso de no cumplirse, utilizamos una prueba no paramétrica en su lugar.

Las hipótesis de los contrastes anovas son las siguientes:

- H0: No existe diferencia significativa entre las medias de los turnos de la asignatura indicada.
- H1: Existe diferencia significativa entre las medias de los turnos de la asignatura indicada.

Gráfico 44

Contraste de Anova y Test de Tukey para la asignatura de Lengua para cuarto año.

```
[1] "Contraste de Anova para la materia LENGUA para 4º año"
      Df Sum Sq Mean Sq F value    Pr(>F)
dataset_p_año$turnos  2   11.0    5.508     5.01 0.00693 **
Residuals           661  726.7    1.099
---

Tukey multiple comparisons of means
 95% family-wise confidence level

Fit: aov(formula = dataset_p_año$promedios ~ dataset_p_año$turnos)

$`dataset_p_año$turnos`
      diff      lwr      upr      p adj
MAÑANA-COMPLETO -0.34424403 -1.03904888  0.3505608 0.4752924
TARDE-COMPLETO  -0.08946659 -0.78861099  0.6096778 0.9514175
TARDE-MAÑANA     0.25477744  0.05924563  0.4503093 0.0064904
```

Podemos observar en el gráfico 44 que el p-value del contraste de Anova da menor a 0.05, por lo que podemos rechazar H0, por lo que decimos que existe diferencia significativa, y el Test de Tukey nos indica que la diferencia se encuentra entre el turno Tarde y Mañana.

Gráfico 45

Contraste de Anova y Test de Tukey para la asignatura de Lengua para quinto año.

```
[1] "Contraste de Anova para la materia LENGUA para 5º año"
      Df Sum Sq Mean Sq F value Pr(>F)
dataset_p_año$turnos  2   13.1    6.527   5.837 0.00306 **
Residuals            709   792.8    1.118
---

Tukey multiple comparisons of means
 95% family-wise confidence level

Fit: aov(formula = dataset_p_año$promedios ~ dataset_p_año$turnos)

$`dataset_p_año$turnos`
      diff      lwr      upr    p adj
MAÑANA-COMPLETO 0.4403431 -0.19177175 1.0724580 0.2311384
TARDE-COMPLETO  0.6723517  0.03241451 1.3122888 0.0368065
TARDE-MAÑANA    0.2320085  0.03693950 0.4270776 0.0147926
```

Para el caso de quinto año para la asignatura de Lengua, observamos en el gráfico 45 un p-value menor a 0.05, por lo que nuevamente rechazamos H_0 y decimos que hay una diferencia significativa en las notas de los turnos. Con el test de Tukey podemos observar que la diferencia se encuentra entre los turnos Tarde y Completo, junto con Tarde y Mañana.

Gráfico 46

Contraste de Anova y Test de Tukey para la asignatura de Lengua para sexto año.

```
[1] "Contraste de Anova para la materia LENGUA para 6º año"
      Df Sum Sq Mean Sq F value    Pr(>F)
dataset_p_año$turnos  2   26.5   13.258   12.02 7.32e-06 ***
Residuals            722   796.4    1.103
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
[1] "-----"
```

```
Tukey multiple comparisons of means
 95% family-wise confidence level

Fit: aov(formula = dataset_p_año$promedios ~ dataset_p_año$turnos)

$`dataset_p_año$turnos`
      diff      lwr      upr    p adj
MAÑANA-COMPLETO 0.2021051 -0.30598161 0.7101918 0.6188066
TARDE-COMPLETO  0.5739612  0.05976203 1.0881605 0.0242373
TARDE-MAÑANA    0.3718561  0.18250640 0.5612059 0.0000140
```

Luego, en el caso de sexto año nuevamente para la asignatura de Lengua, vemos en el gráfico 46 que el p-value obtenido para el contraste Anova es muy inferior a 0.05, por lo que procedemos a rechazar H_0 y decir que hay una diferencia significativa en los promedios entre los turnos. Con el test de Tukey, notamos que las diferencias se encuentran entre los turnos Tarde y Completo, junto con Tarde y Mañana.

Para los años restantes que no se encuentran en este análisis con su respectiva asignatura es porque no se encontró diferencia significativa. Analizando las potencias de estos test, podemos notar que no dan muy alta, entre 0.13 y 0.28, lo cual es bastante bajo.

Ahora bien, procederemos a comprobar el cumplimiento de las condiciones de los contrastes Anovas, donde calcularemos la normalidad de cada uno de los años para cada turno utilizando Lilliefors para aquellos con más de 30 datos, y Shapiro en caso contrario. Si alguno cumple con la normalidad, calcularemos la varianza de los residuos del modelo del contraste Anova.

Gráfico 47

Test de normalidad para cada turno de cuarto año de la asignatura de Lengua.

```
[1] "Test de normalidad, LENGUA para 4° año"
[[1]]

      Lilliefors (Kolmogorov-Smirnov) normality test

data:  a_testear_normalidad
D = 0.083034, p-value = 1.486e-06

[[2]]

      Lilliefors (Kolmogorov-Smirnov) normality test

data:  a_testear_normalidad
D = 0.14414, p-value = 4.462e-15

[[3]]

      Shapiro-Wilk normality test

data:  a_testear_normalidad
W = 0.94812, p-value = 0.5701
```

Gráfico 48

Test de normalidad para cada turno de quinto año de la asignatura de Lengua.

```
[1] "-----"
[1] "Test de normalidad, LENGUA para 5º año"
[[1]]

      Lilliefors (Kolmogorov-Smirnov) normality test

data:  a_testear_normalidad
D = 0.11195, p-value = 2.114e-14

[[2]]

      Lilliefors (Kolmogorov-Smirnov) normality test

data:  a_testear_normalidad
D = 0.11075, p-value = 4.298e-08

[[3]]

      Shapiro-Wilk normality test

data:  a_testear_normalidad
W = 0.87056, p-value = 0.02773
```

Gráfico 49

Test de normalidad para cada turno de sexto año de la asignatura de Lengua.

```
[1] "Test de normalidad, LENGUA para 6º año"
[[1]]

      Lilliefors (Kolmogorov-Smirnov) normality test

data:  a_testear_normalidad
D = 0.080443, p-value = 1.058e-06

[[2]]

      Lilliefors (Kolmogorov-Smirnov) normality test

data:  a_testear_normalidad
D = 0.11843, p-value = 1.596e-10

[[3]]

      Shapiro-Wilk normality test

data:  a_testear_normalidad
W = 0.93746, p-value = 0.1293
```

Podemos observar en los gráficos 47, 48 y 49 que en todos los casos, al menos uno de los turnos no cumplen con el supuesto de normalidad. Por lo tanto el modelo ANOVA no es válido y se deben utilizar pruebas no paramétricas para cada uno de ellos. Utilizaremos el test de Kruskal-Wallis para este cálculo junto con su contraste posterior.

Gráfico 50

Test de Kruskal-Wallis y su contraste posterior para la asignatura de Lengua para cuarto año.

```
[1] "Test de Kruskal-Wallis para la materia LENGUA para 4º año"

      Kruskal-Wallis rank sum test

data:  dataset_p_año$promedios and dataset_p_año$turnos
Kruskal-Wallis chi-squared = 11.733, df = 2, p-value = 0.002833

      Pairwise comparisons using Wilcoxon rank sum test

data:  dataset_p_año$promedios and dataset_p_año$turnos

      COMPLETO MAÑANA
MAÑANA 0.5063      -
TARDE  0.9575    0.0025

P value adjustment method: holm
```

Gráfico 51

Test de Kruskal-Wallis y su contraste posterior para la asignatura de Lengua para quinto año.

```
[1] "Test de Kruskal-Wallis para la materia LENGUA para 5º año"

      Kruskal-Wallis rank sum test

data:  dataset_p_año$promedios and dataset_p_año$turnos
Kruskal-Wallis chi-squared = 9.5916, df = 2, p-value = 0.008264

      Pairwise comparisons using Wilcoxon rank sum test

data:  dataset_p_año$promedios and dataset_p_año$turnos

      COMPLETO MAÑANA
MAÑANA 0.120      -
TARDE  0.045    0.039

P value adjustment method: holm
```

Gráfico 52

Test de Kruskal-Wallis y su contraste posterior para la asignatura de Lengua para sexto año.

```
[1] "Test de Kruskal-Wallis para la materia LENGUA para 6° año"

      Kruskal-Wallis rank sum test

data:  dataset_p_año$promedios and dataset_p_año$turnos
Kruskal-Wallis chi-squared = 22.701, df = 2, p-value = 1.176e-05

      Pairwise comparisons using Wilcoxon rank sum test

data:  dataset_p_año$promedios and dataset_p_año$turnos

      COMPLETO MAÑANA
MAÑANA 0.31      -
TARDE  0.03      2.1e-05

P value adjustment method: holm
```

Como podemos observar en los gráficos 50, 51 y 52, como el p-value del test de Kruskal-Wallis es menor a 0.05, por lo que se rechaza la hipótesis nula, la cual es que la variable de respuesta es la misma en todas las poblaciones valoradas, y en los casos que se nombran a continuación, la diferencia es significativa.

- Cuarto año - Lengua: $p = 0.002$
- Quinto año - Lengua: $p = 0.008$
- Sexto año - Lengua: $p = 1.17 \cdot 10^{-5}$

Donde podemos observar en los mismos gráficos, en qué turnos hay diferencia significativa gracias al contraste posterior, que son los siguientes:

- Cuarto año - Lengua: Tarde-Mañana
- Quinto año - Lengua: Tarde-Mañana y Tarde-Completo
- Sexto año - Lengua: Tarde-Completo y Tarde-Mañana

Gráfico 53

Test de Kruskal-Wallis y su contraste posterior para la asignatura de Matemática para tercer año.

```
[1] "Test de Kruskal-Wallis para la materia MATEMÁTICA para 3º año"

      Kruskal-Wallis rank sum test

data:  dataset_p_año$promedios and dataset_p_año$turnos
Kruskal-Wallis chi-squared = 6.7259, df = 2, p-value = 0.03463

      Pairwise comparisons using Wilcoxon rank sum test

data:  dataset_p_año$promedios and dataset_p_año$turnos

      COMPLETO MAÑANA
MAÑANA 0.502      -
TARDE  0.226      0.087

P value adjustment method: holm
```

Además, este estudio se realizó para aquellos donde la diferencia no era significativa, y se encontró que para el tercer año en Matemática hay diferencia significativa con el test de Kruskal-Wallis, como podemos observar en el gráfico 53, pero su contraste posterior nos indica que no existe diferencia significativa en el mismo, lo cual es interesante, y esto puede deberse a la proximidad al valor de 0.05 en el p-value del test de Kruskal-Wallis.

Después se realizó un análisis comparativo entre las notas de los trimestres de primer año de secundaria para la materia de Matemática, para analizar el desempeño de los alumnos en el mismo y detectar si ha habido una mejora o no.

Gráfico 54

Boxplot exploratorio para el desempeño por trimestre de primer año para Matemática

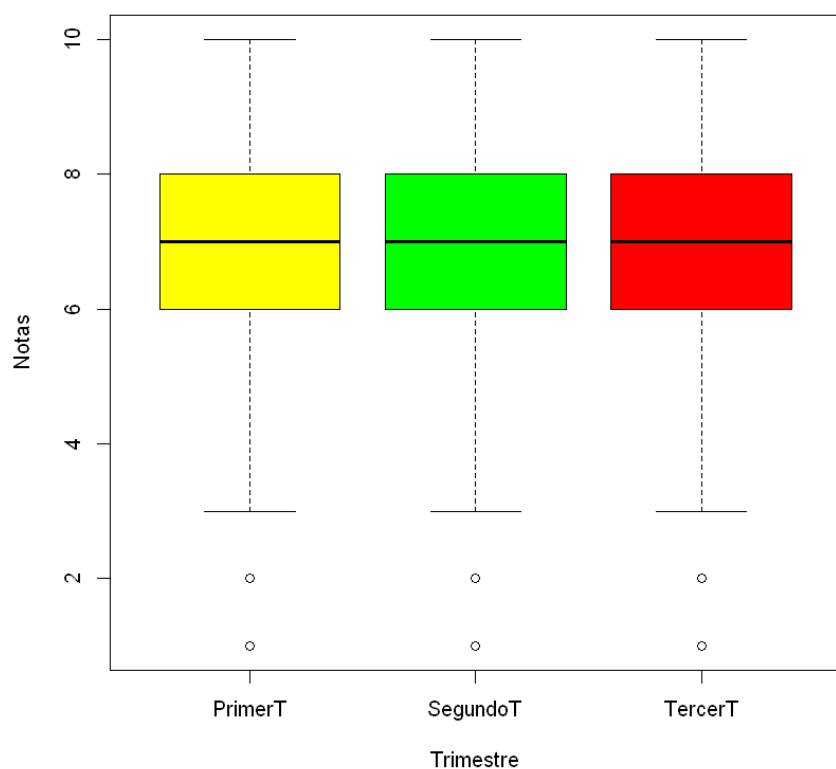


Tabla 8

Descriptivos de las notas de cada trimestre para primer año de Matemática

Trimestre	Media	Mediana
1°	6,73	7
2°	6,89	7
3°	6,80	7

Primeramente se hizo un análisis de descriptivos junto con un análisis gráfico, donde viendo los resultados dados en el gráfico 54 y la tabla 8, a simple vista no se puede notar una diferencia significativa, donde observamos que las medianas son iguales en los 3 grupos y las medias son similares. En este caso se tomaron que las muestras son pareadas, por lo que se utilizará el test de Friedman para ver si hay diferencias significativas en los desempeños de los trimestres.

Gráfico 55

Test de Friedman y su contraste posterior.

```
Friedman rank sum test

data:  Nota and Trimestre and Documento
Friedman chi-squared = 13.413, df = 2, p-value = 0.001223

Pairwise comparisons using Wilcoxon signed rank test

data:  data_secu_1$Nota and data_secu_1$Trimestre

      PrimerT SegundoT
SegundoT 0.00035 -
TercerT   0.13193 0.13193

P value adjustment method: holm
```

Sin embargo, utilizando el test de Friedman mostrado en el gráfico 55, como el p-value obtenido es menor a 0.05 podemos decir que hay diferencia significativa y, utilizando su contraste posterior, decimos que esta se encuentra entre el primer y segundo trimestre.

Análisis Bivariado Cualitativo

En primer lugar, antes de realizar el análisis, se crearon 4 nuevas variables;

- CondiciónFinal: Consta de los valores “APROBADO” y “DESAPROBADO”, donde el alumno se lo marca como aprobado en caso de tener un promedio mayor o igual a 6 y una nota mayor a 6 en el tercer trimestre, y desaprobado en caso de que no se cumpla ninguna de esas dos condiciones.
- Desempeño: Consta de los valores “INSUFICIENTE”, “REGULAR”, “APROBADO”, “BUENO”, “MUY BUENO” y “DISTINGUIDO”, donde cada uno pertenece a un rango dado por el promedio final de cada asignatura.
- Eximido: Consta de los valores “Eximido”, “ParcialmenteEximido” y “NoEximido”, donde un alumno será eximido si aprobó todas las asignaturas, parcialmente eximido si desaprobó menos de 4 y no eximido si desaprobó 4 materias o más.
- MejoriaT: Consta de los valores “MEJORO”, “IGUAL” y “EMPEORO”, donde un alumno va a mejorar en una asignatura si la nota del tercer trimestre es mejor a la del primero, igual si la nota es la misma y empeora si la nota del tercer trimestre es más baja que la del primero.

Luego, para nuestro análisis estudiaremos si aprobar Matemática en primer año de secundaria depende de si el alumno asiste a una escuela privada o no.

Gráfico 56

Tabla de contingencia de totales, de fila y de columna.

Cell Contents			

Count			
Row Percent			
Column Percent			
Total Percent			

Total Observations in Table: 884			
dataset_prueba_4\$EsPrivada	dataset_prueba_4\$CondicionFinal		
	APROBADO	DESAPROBADO	Row Total

FALSE	539	216	755
	71.4%	28.6%	85.4%
	83.1%	91.9%	
	61.0%	24.4%	

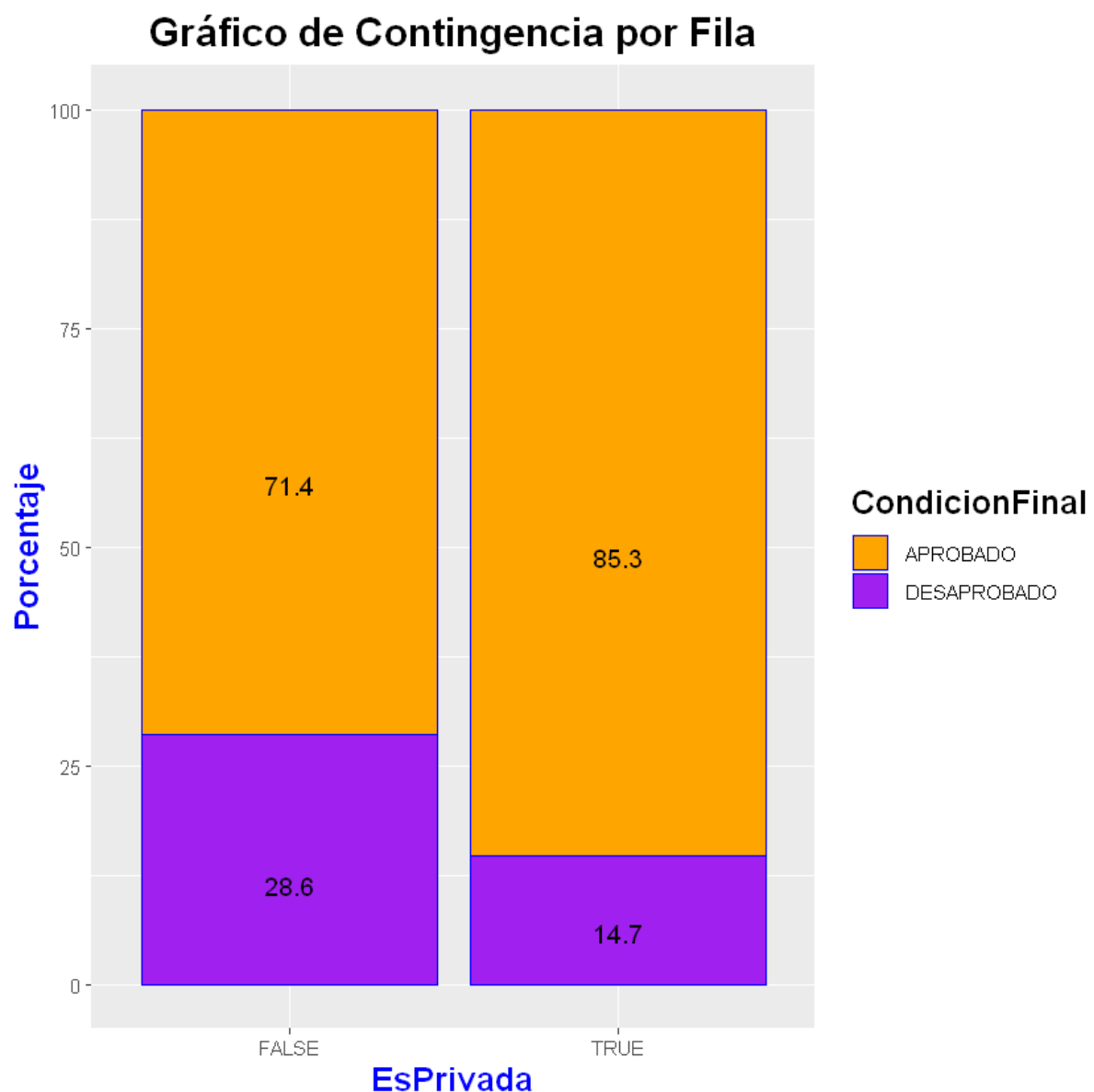
TRUE	110	19	129
	85.3%	14.7%	14.6%
	16.9%	8.1%	
	12.4%	2.1%	

Column Total	649	235	884
	73.4%	26.6%	

En el gráfico 56, podemos ver la tabla de contingencia que contiene los totales, de filas y de columnas, donde para la tabla de contingencia por filas podemos ver cómo se distribuyen las condiciones finales con respecto a si la escuela a la que asiste el alumno es privada o no. Para la tabla de contingencia por columnas, podemos ver cómo se distribuye la gestión de la escuela con respecto a la condición final de la asignatura.

Gráfico 57

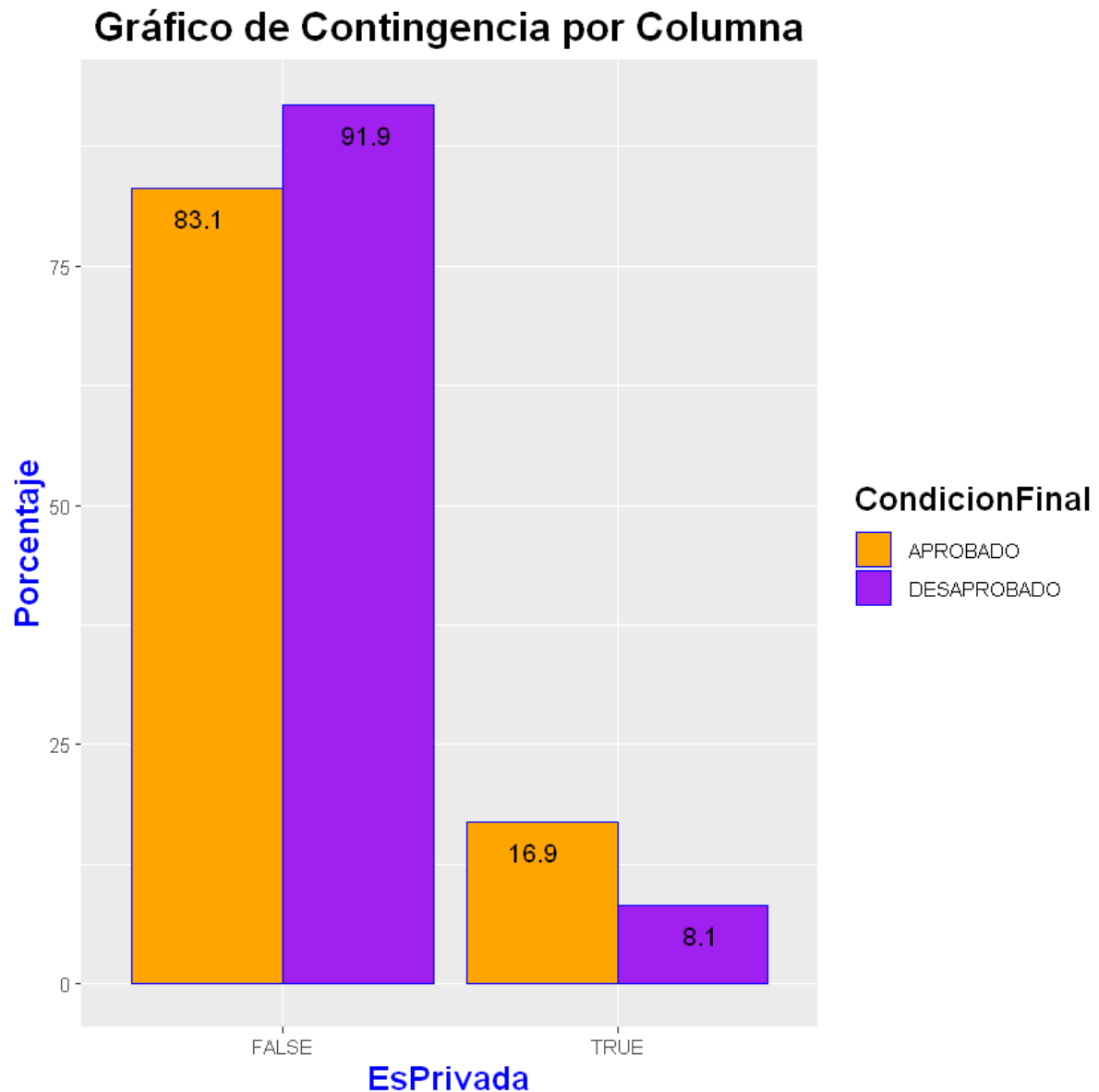
Gráfico de contingencia por fila para las condiciones finales de Matemática con respecto a la gestión.



Podemos observar en el gráfico 57, que el porcentaje de aprobados para las escuelas públicas es de 71.4%, y en el caso de las escuelas privadas es de 85.3%, por lo que podemos suponer que hay una asociación.

Gráfico 58

Gráfico de contingencia por columna para las condiciones finales de Matemática con respecto a la gestión.



En el gráfico 58, podemos observar que para la columna “APROBADO” el 83,1 % pertenece a las escuelas públicas. Por lo que planteamos como hipótesis que la condición final para la asignatura Matemática de primer año no está asociado a si el alumno va o no a una escuela privada.

Se realizó el test chi-cuadrado para determinar si hay evidencia estadística suficiente para rechazar la hipótesis planteada anteriormente.

Gráfico 59

Test Chi-Cuadrado.

Pearson's Chi-squared test with Yates' continuity correction

```
data: tabla_mat_con$EsPrivada and tabla_mat_con$CondicionFinal  
X-squared = 10.177, df = 1, p-value = 0.001422
```

En el gráfico 59 observamos que el test nos arrojó un p-value menor a 0.05, por lo que hay evidencia suficiente para rechazar nuestra hipótesis nula y decimos que existe asociación entre la condición final y la gestión de la escuela. A partir de esto, analizamos con el test de Cramer la fuerza de esta asociación encontrada, donde se vio un resultado de 0.15, que es una fuerza de asociación leve. Se verá la tabla de valores esperados.

Gráfico 60

Tabla de valores esperados

	APROBADO	DESAPROBADO
FALSE	554	201
TRUE	95	34

En la tabla del gráfico 60, vemos que se esperan menos aprobados para las escuelas privadas que los dados en la tabla de contingencia del gráfico 56, al igual que más desaprobados en las escuelas públicas. Acto seguido, se analizó la dirección de la fuerza de asociación con los residuos.

Gráfico 61

Tabla de residuos corregidos del test Chi-Cuadrado

	tabla_mat_con\$CondicionFinal	
tabla_mat_con\$EsPrivada	APROBADO	DESAPROBADO
FALSE	-3.3	3.3
TRUE	3.3	-3.3

En el gráfico 61, podemos observar en la tabla de residuos corregidos que la fuerza de asociación es positiva para los aprobados de una escuela privada y para los desaprobados en una escuela pública. De manera contraria, es negativa para los aprobados en una escuela pública y para los desaprobados en una escuela privada.

Gráfico 62

Tabla de razones condicionales

	FALSE	TRUE	Total
APROBADO	539.0	110.0	649.0
DESAPROBADO	216.0	19.0	235.0
Total	755.0	129.0	884.0
razones_condicionales	2.5	5.8	2.8

En el gráfico 62 tenemos la tabla de razones condicionales, donde podemos decir que de cada 250 aprobados en la escuela pública, tenemos 100 desaprobados, mientras que en la escuela privada cada 580 aprobados, tenemos 100 desaprobados.

Otro análisis que se propuso realizar es si existe asociación entre la mejoría entre trimestres de un alumno y el tipo de escuela, privada o pública, en la asignatura lengua y literatura para todos los años de secundaria.

Gráfico 63

Tabla de contingencia de totales, de fila y de columna

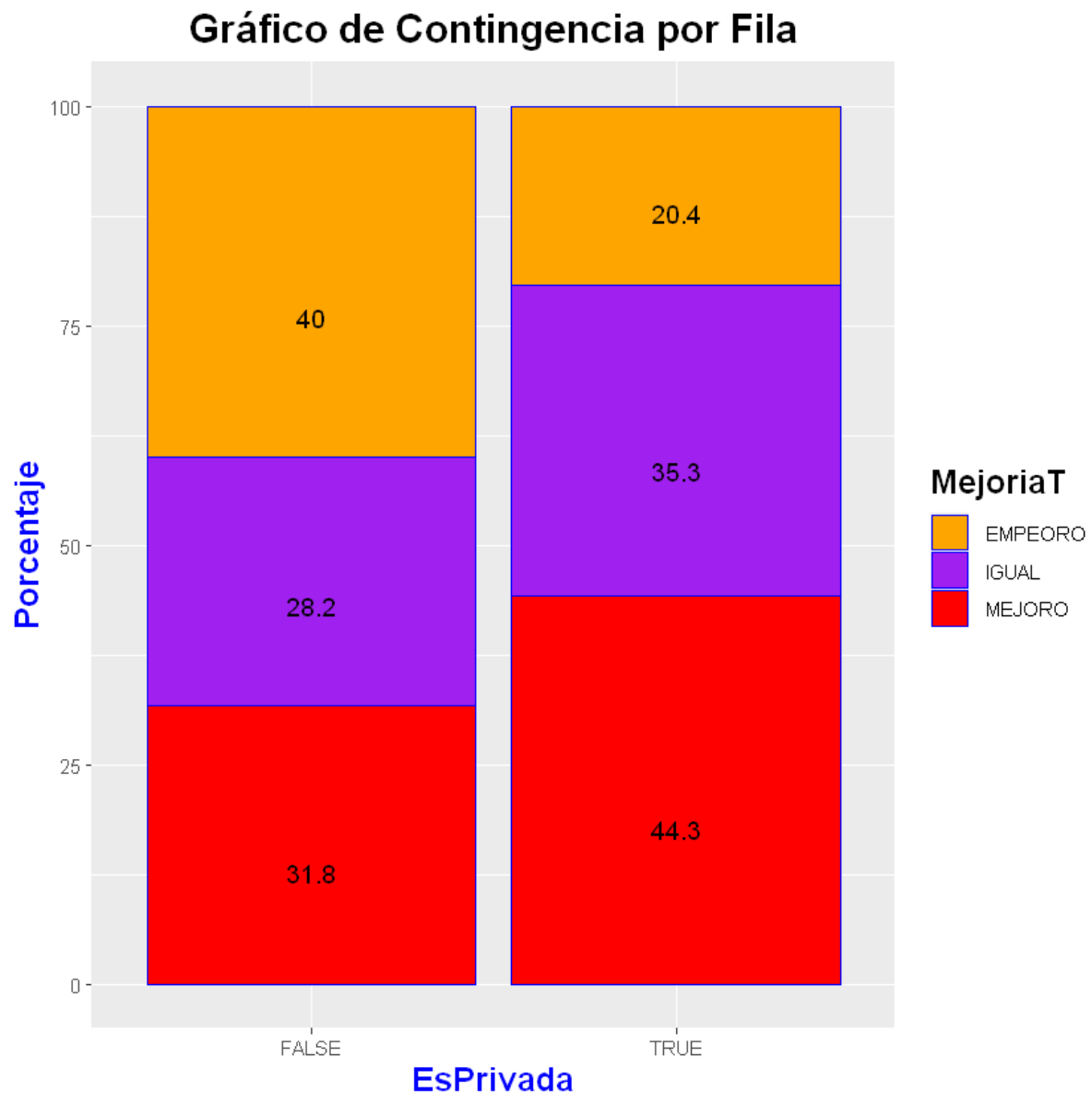
Cell Contents				
Count				
Row Percent				
Column Percent				
Total Percent				
Total Observations in Table: 4043				
dataset_prueba_5\$EsPrivada	dataset_prueba_5\$MejoriaT			Row Total
	EMPEORO	IGUAL	MEJORO	
FALSE	1408	993	1118	3519
	40.0%	28.2%	31.8%	87.0%
	92.9%	84.3%	82.8%	
	34.8%	24.6%	27.7%	
TRUE	107	185	232	524
	20.4%	35.3%	44.3%	13.0%
	7.1%	15.7%	17.2%	
	2.6%	4.6%	5.7%	
Column Total	1515	1178	1350	4043
	37.5%	29.1%	33.4%	

En el gráfico 63, podemos ver la tabla de contingencia que contiene los totales, de filas y de columnas, donde para la tabla de contingencia por filas podemos ver cómo se distribuyen las mejoras por trimestre con respecto a si la escuela a la que asiste el alumno es privada o

no. Para la tabla de contingencia por columnas, podemos ver cómo se distribuye la gestión de la escuela con respecto a la mejoría por trimestre de la asignatura.

Gráfico 64

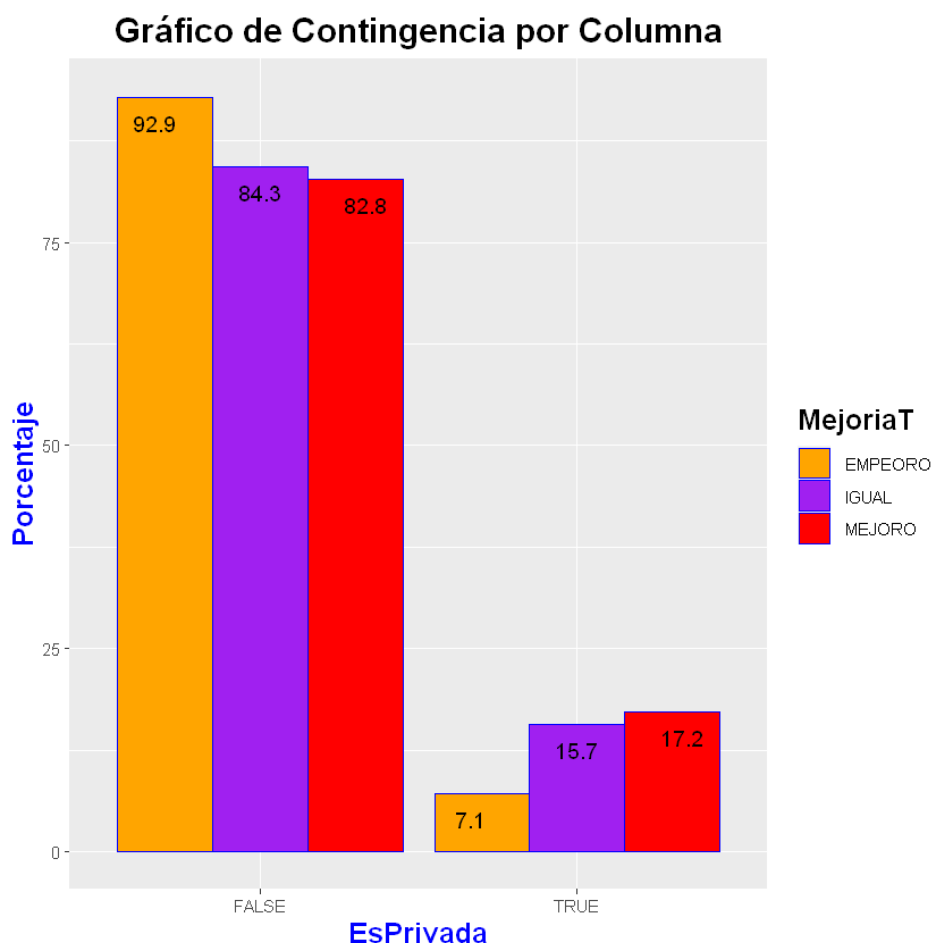
Gráfico de contingencia por fila para las mejoras trimestrales de Lengua y Literatura con respecto a la gestión.



Podemos observar en el gráfico 64, que el porcentaje de alumnos que mejoró al final del trimestre es del 31,8 % con respecto al total de alumnos de escuelas públicas. En cambio para los alumnos de escuelas privadas el 44,3 % mejoró hacia el final del cuatrimestre, por lo que podemos suponer que hay una asociación.

Gráfico 65

Gráfico de contingencia por columna para la mejoría trimestral de lengua y Literatura con respecto a la gestión.



En el gráfico 65, podemos observar que para la columna “MEJORO” el 82,8% pertenece a las escuelas públicas. También vemos que del 100% de alumnos que se encuentran en la categoría de “EMPEORO” el 92,9% pertenecen a escuelas públicas 7,1% a privadas .Por lo que planteamos como hipótesis que la mejoría trimestral para la asignatura Lengua y Literatura de primer año no está asociado a si el alumno va o no a una escuela privada. Se realizó el test chi-cuadrado para determinar si hay evidencia estadística suficiente para rechazar la hipótesis planteada anteriormente.

Gráfico 66

Tabla de residuos corregidos del test Chi-Cuadrado

Pearson's Chi-squared test

```
data: tabla_mejora$EsPrivada and tabla_mejora$MejoriaT  
X-squared = 75.936, df = 2, p-value < 2.2e-16
```

Podemos ver en el gráfico 66 que el p-valor proporcionado por el test es menor al nivel de significancia de 0.05 por lo que hay evidencia estadística suficiente para rechazar la hipótesis nula y podemos decir que existe una asociación entre la mejoría trimestral y la gestión de la escuela a la que asista el alumno. En última instancia medimos la fuerza de la asociación utilizando el test de Cramer, donde se vio un resultado de 0.19, que es una fuerza de asociación leve. Se verá la tabla de valores esperados.

Gráfico 67

Tabla de valores esperados

	EMPEORO	IGUAL	MEJORO
FALSE	1319	1025	1175
TRUE	196	153	175

En el gráfico 67, observamos que en la tabla se espera que menos estudiantes empeoren que los observados en las escuelas públicas, a su vez que se esperan más estudiantes que estén igual o mejor. Para las escuelas privadas sucede lo contrario, es decir se esperan que bastantes más estudiantes empeoren sus notas, mientras que se espera que menos estudiantes igualen o mejoren sus notas.

Ahora bien, se analizará la dirección de la asociación entre las variables estudiadas.

Gráfico 68

Tabla de residuos corregidos del test Chi-Cuadrado

```

      tabla_mejora$MejoriaT
tabla_mejora$EsPrivada EMPEORO IGUAL MEJORO
      FALSE      8.6  -3.3  -5.7
      TRUE     -8.6   3.3   5.7

```

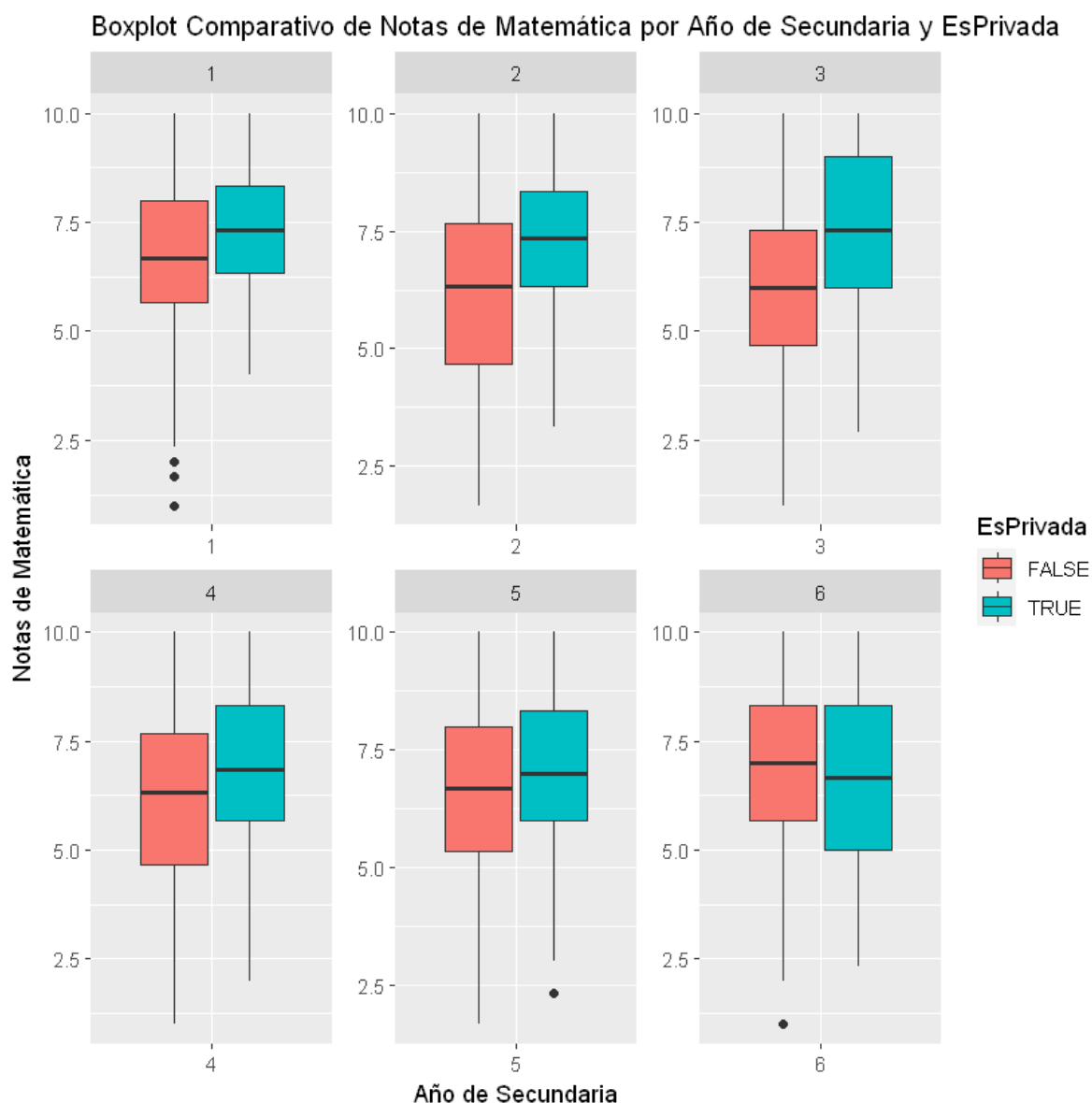
Viendo en el gráfico 68 la tabla de residuos corregidos del test hecho anteriormente, podemos observar que es más probable que un estudiante que asiste a una escuela pública empeore sus notas de Lengua y Literatura en el transcurso del año y menos probable que las iguale o las mejore. Caso contrario a las escuelas privadas, donde se puede observar que es más probable que un alumno mejore o iguale sus notas.

Análisis Multivariado Cuantitativo

En esta sección estudiamos si existe alguna posible relación entre variables, individuos o ambos. En primer lugar comparamos las notas de Matemática para cada año de secundaria segmentado por gestión.

Gráfico 69

Boxplot comparativo de las notas de Matemática por año de secundaria por gestión.

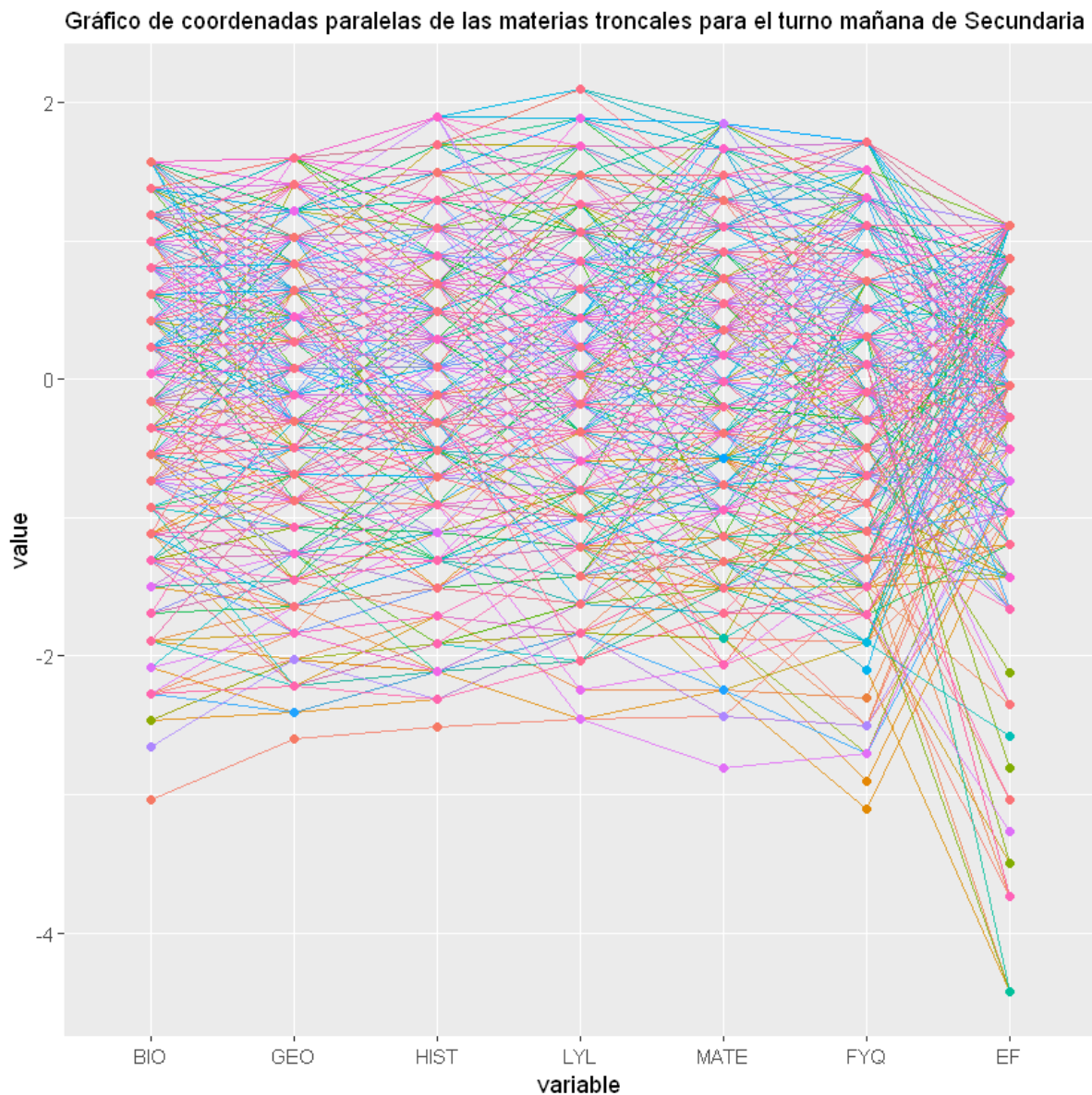


En el gráfico 69, se puede observar una posible diferencia entre las medias de los años de escuelas de gestión pública y privada. Queda pendiente de estudio la verificación de la existencia de las diferencia de medias mediante contrastes.

Analizamos las notas obtenidas por los estudiantes del trueno mañana en las materias troncales de primer año de Secundaria.

Gráfico 70

Gráfico de coordenadas paralelas de las materias troncales para el Turno Mañana de Secundaria de Primer Año

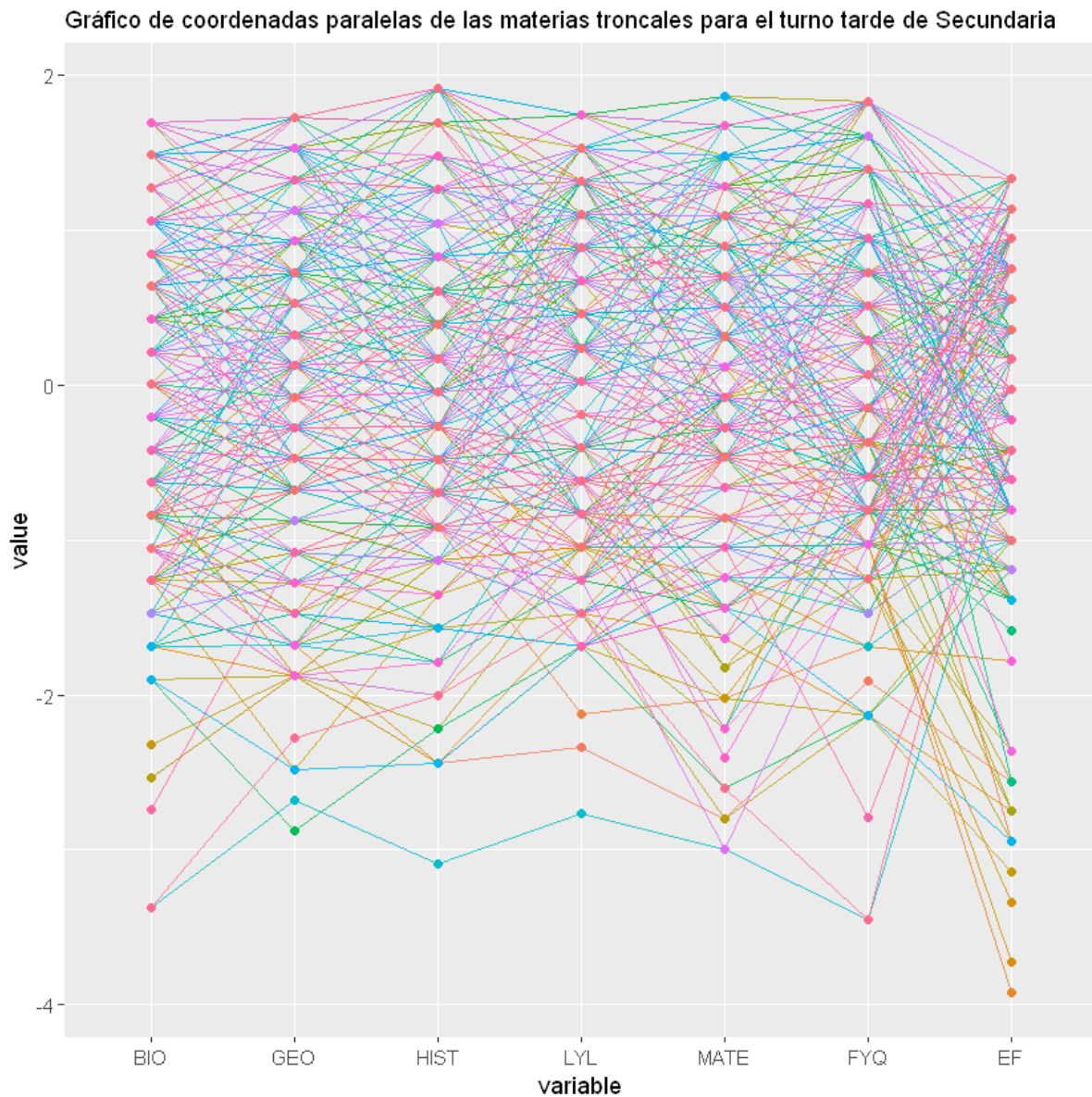


En el gráfico 70, observamos que la asignatura educación física posee valores atípicos menores. Por otro lado, notamos que la asignatura Lengua y Literatura tiene valores atípicos altos.

Realizamos el mismo análisis pero para el turno tarde.

Gráfico 71

Gráfico de coordenadas paralelas de las materias troncales para el Turno Tarde de Secundaria de Primer Año



En el gráfico 71 vemos que la asignatura biología y educación física tienen valores atípicos menores notorios. También, destacan las asignaturas historia y matemáticas por sus claros valores atípicos altos.

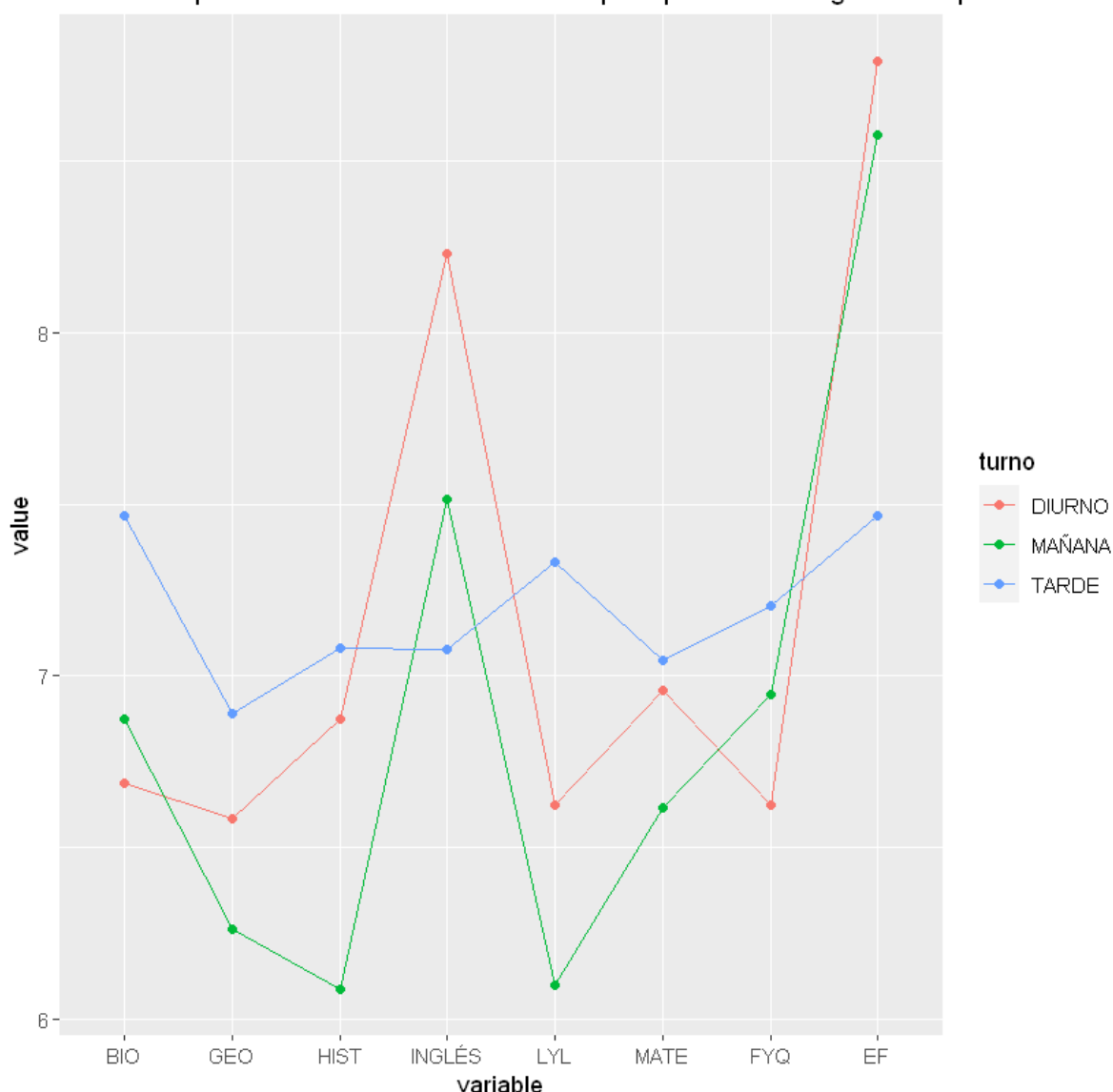
Tomando ambos turnos se realizó un gráfico de perfiles para analizar los valores medios o medianos de cada una de las variables observadas en distintos individuos o en las diferentes categorías en las que se clasifica al grupo o a los individuos.

Esto nos permitirá comparar la posición central de estas variables en los distintos individuos o grupos definidos.

Gráfico 72

Gráfico de perfiles de las materias troncales para Primer Año segmentado por Turnos

Gráfico de perfiles de las materias troncales para primer año segmentado por turnos

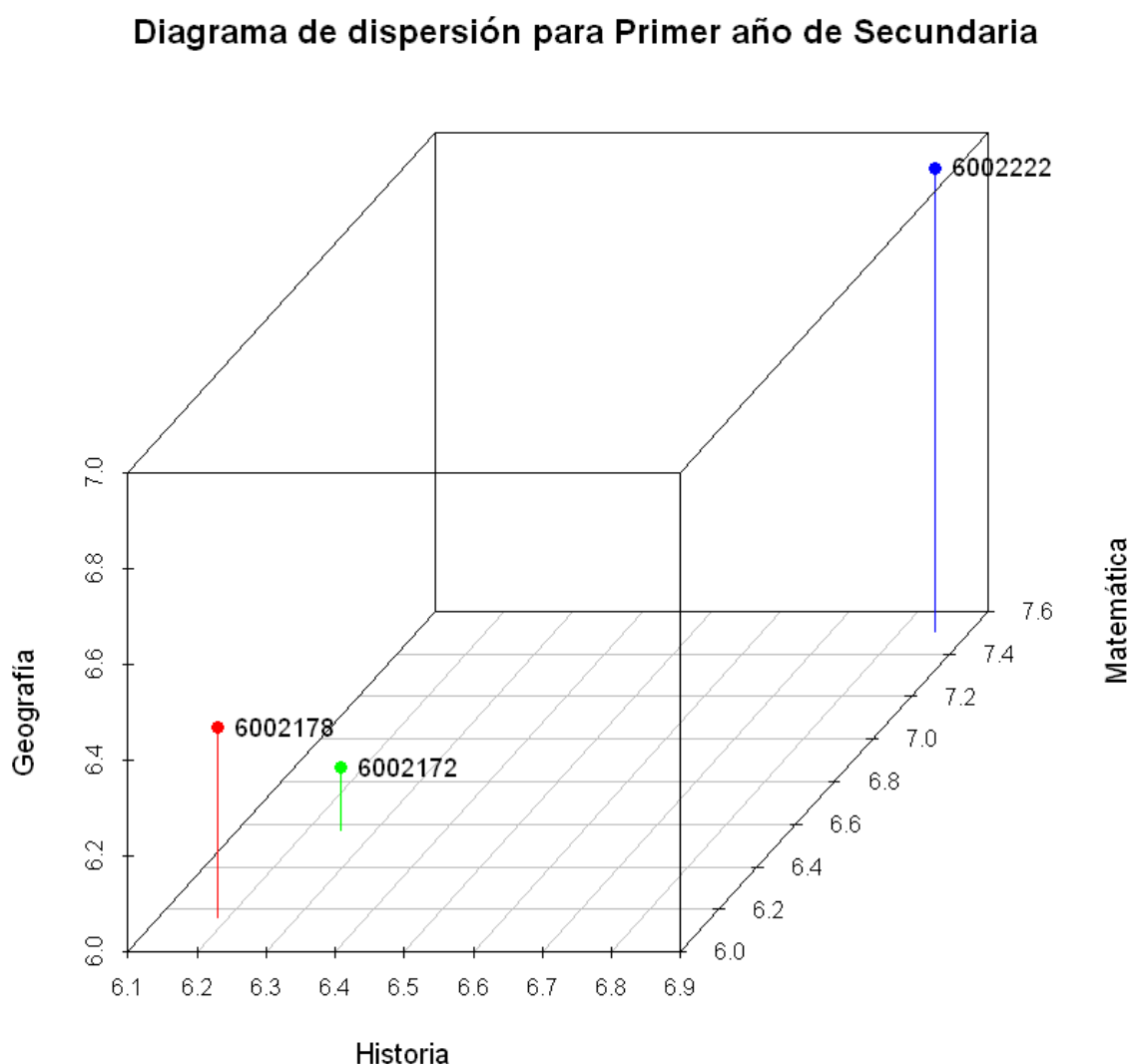


En el gráfico 72, podemos observar que el turno diurno presenta picos altos de sus valores medios para las asignaturas inglés y educación física, y picos por debajo de 7 en las asignaturas geografía, lengua y literatura y física y química. Para el turno mañana observamos picos de valores medios cercanos a 6 en las asignaturas historia y lengua y literatura, por debajo de 7 en matemática y un pico alto para educación física.

En el siguiente estudio, se analizará las medias de las notas de Geografía, Historia y Matemática entre las escuelas de primer año de secundaria.

Gráfico 73

Diagrama de dispersión en 3D para Primer Año de Secundaria para las materias de Geografía, Historia y Matemática según la escuela.



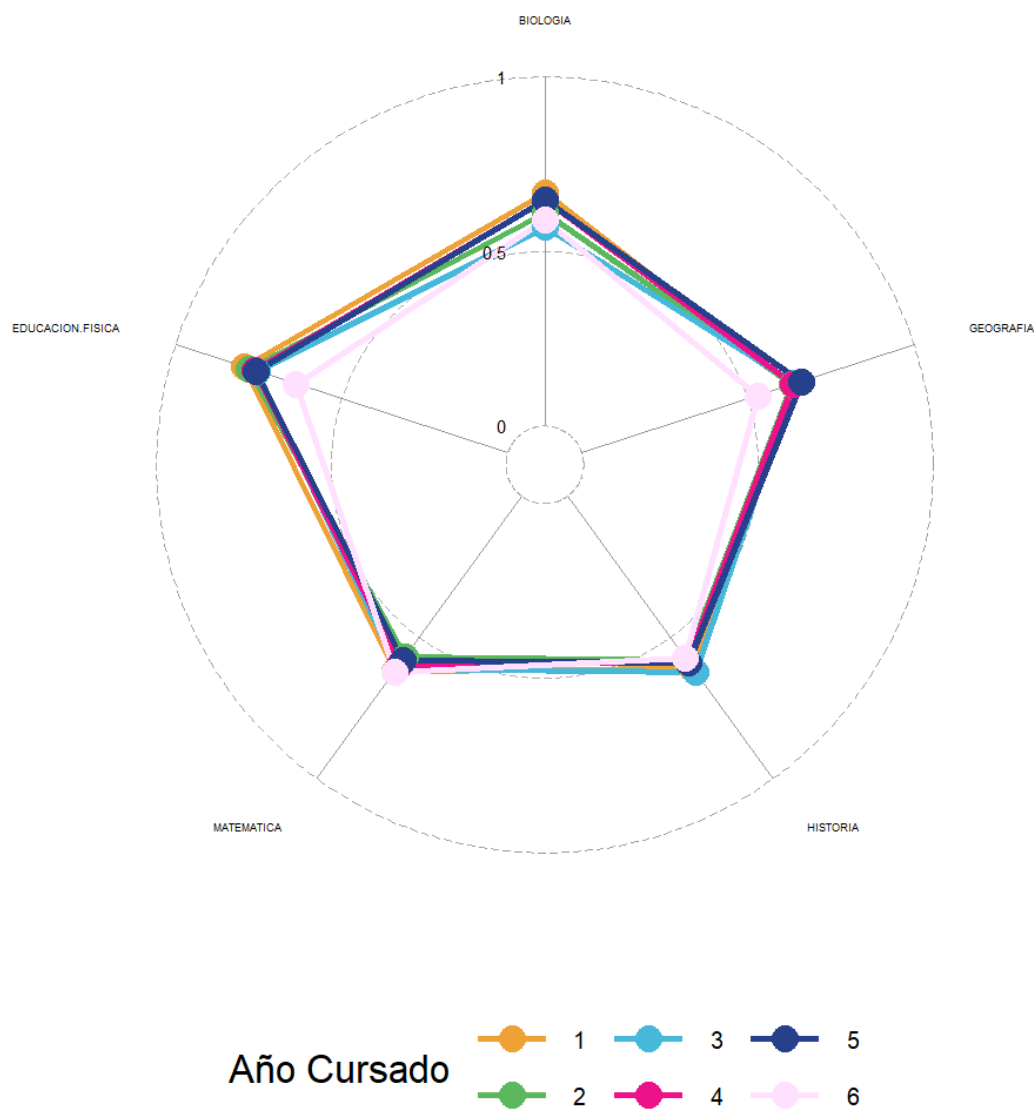
Podemos observar en el gráfico 73, que la escuela con el CUE 6002222 presenta mejores medias en las notas de Geografía, Historia y Matemática, mientras que la escuela con CUE 6002178, presenta peores notas en Historia y Matemática que las otras dos pero presenta mejores notas de Geografía que la escuela con CUE 6002172.

Luego se realizó un análisis de las materias troncales de Secundaria segmentadas por año, para ello se tomaron las materias comunes entre los 6 años, las cuales son Matemática, Biología, Geografía, Educación Física e Historia.

Gráfico 74

Gráfico de diagrama radial para las notas de las materias troncales segmentadas por Año.

Comparación de las notas de las materias troncales por Año



En el gráfico 74 podemos observar que uno de los años con mayor desempeño en todas las notas comparadas fue Primero, seguido por Quinto, ya que son los que presentan los picos más cercanos al borde del círculo (es decir, al valor 1, porque los valores están escalados). Mientras que Sexto presenta mayoritariamente los promedios más bajos. Los más bajos en de notas en las materias fueron:

- Biología: Tercero
- Geografía: Sexto
- Historia: Sexto
- Matemática: Segundo
- Educación Física: Sexto

Conclusiones

En primer lugar y dado al análisis exploratorio realizado estamos en condiciones de decir que en la implementación del boletín virtual hace falta una mejora a la hora de cargar datos ya que en un principio el estado general de la base de datos era malo y en consecuencia demanda tiempo en la limpieza y reorganización de los datos. El cual se podría destinar a un análisis más profundo.

En cuanto al análisis del rendimiento académico, en el caso de las escuelas primarias observamos en la sección de exploración que el promedio para las materias de matemática y lengua en los seis grados es aproximadamente 8, y sus mínimos en todos los casos es 6 por lo que podemos decir que el rendimiento académico para estas materias de los estudiantes de primaria del departamento La Paz es muy bueno.

En un análisis similar para las escuelas secundarias, obtenemos que la media de los promedios ronda alrededor de 6 en ambas materias, distinguiéndose en lengua y literatura quinto año con una media de 7,51 y sexto año con una media 8,30. También notamos que los mínimos llegan a la nota mínima (1). Dado esto concluimos que el rendimiento académico en estas materias es regular.

En la segunda instancia de este trabajo se realizaron análisis comparativos de medias. Este estudio permitió identificar diferencias significativas en el rendimiento académico de los estudiantes en distintas asignaturas, años escolares y escuelas, donde se encontró que para la primera hipótesis que planteaba que los promedios de Matemática de primer año eran mayores en el turno mañana, que no hay evidencia estadística suficiente para decir que existe una diferencia significativa entre los promedios de primer año del turno mañana y tarde de secundaria en la asignatura Matemática.

También se encontró que existe diferencia significativa entre los promedios de Matemática entre las escuelas privadas y públicas, siendo las notas de matemática superiores en las escuelas de gestión privada.

Otro análisis comparativo que se realizó fue sobre el desempeño de la asignatura Lengua y Matemática de los estudiantes de primaria, considerando los diferentes turnos, donde se encontró realizando pruebas no paramétricas, que para los años cuarto, quinto y sexto existen diferencias significativas en los promedios de la asignatura Lengua, pero no así para Matemática según el turno. Y con un contraste posterior se ve que para cuarto año en la asignatura de Lengua hay diferencias entre el turno tarde y mañana, para quinto año hay diferencias entre Tarde-Mañana y Tarde-Completo y para sexto año hay diferencias entre los turnos Tarde-Completo y Tarde-Mañana.

Por último se analizó si existe diferencia entre los promedios de los trimestres de primer año de secundaria para la materia Matemática, para analizar el desempeño de los estudiantes y detectar si hubo mejora. Se encontró que hay diferencia significativa entre el Primer y Segundo trimestre en la asignatura Matemática en el primer año de secundaria por lo que los estudiantes tienden a mejorar sus promedios llegando a fin de año.

En esta instancia del trabajo nos preguntamos si existe asociación entre aprobar Matemáticas e ir a una escuela de gestión privada. Por lo que realizamos un análisis bivariado donde se encontró, que el porcentaje de aprobados para las escuelas públicas es de 71.4%, y en el caso de las escuelas privadas es de 85.3%, por lo que suponiendo que hay una asociación, se encontró que los estudiantes de primer año de las escuelas privadas tienen una tendencia favorable de aprobar Matemática, mientras que los de las escuelas públicas tienen tendencia a desaprobado, dando como resultado que de cada 250 aprobados en una escuela pública tenemos 100 desaprobados, mientras que en las escuelas privadas de cada 580 aprobados tenemos 100 desaprobados.

En otro análisis similar se estudió si existe asociación entre la mejoría entre trimestres de un alumno y el tipo de escuela, privada o pública, en la asignatura Lengua y Literatura para todos los años de secundaria. Pudimos observar que el porcentaje de alumnos que mejora hacia el final del año para aquellos que asisten a una escuela pública es del 31,8%. En cambio, el porcentaje de alumnos que asisten a escuelas privadas y mejora hacia el final del año es del 44,3%. Finalmente se concluyó que aquellos estudiantes de secundaria que tienen una mayor probabilidad de mejorar hacia el final del año en la asignatura Lengua y Literatura son los que asistan a escuelas de gestión privada.

Del análisis multivariado concluimos que para las materias troncales en el Turno Mañana de Secundaria de Primer Año la asignatura educación física posee valores atípicos cercanos al extremo inferior. Por otro lado, la asignatura Lengua y Literatura destaca por tener valores atípicos cercanos al extremo superior. En la misma línea para el turno tarde, en las asignaturas biología y educación física destacan valores atípicos cercanos al extremo inferior. También, destacan las asignaturas historia y matemáticas por sus claros valores atípicos hacia el extremo superior.

Para concluir el análisis de las materias troncales para Primer Año segmentado por Turnos El turno diurno destaca en valores medios altos en las asignaturas Inglés y Educación Física, y está debajo de 7 en las asignaturas de Geografía, Lengua y Literatura y Física y Química. El turno mañana destaca por sus bajos valores medios cercanos a 6 en las asignaturas Historia y Lengua y Literatura, por debajo de 7 en matemática y un pico alto para Educación Física, y el turno tarde destaca por ser el más homogéneo de los tres ya que no posee picos exagerados.

La escuela que mejor promedio tiene para el primer año de secundaria en las materias de Geografía, Historia y Matemática es la que se identifica con el CUE 6002222.

Uno de los mejores desempeños en las materias troncales comparadas para todos los años de secundaria fue Primero, seguido por Quinto, y el que peor desempeño tuvo fue Sexto.

Anexo

Gráfico 7

Primaria - Histograma de promedio de Matemática - 1 año.

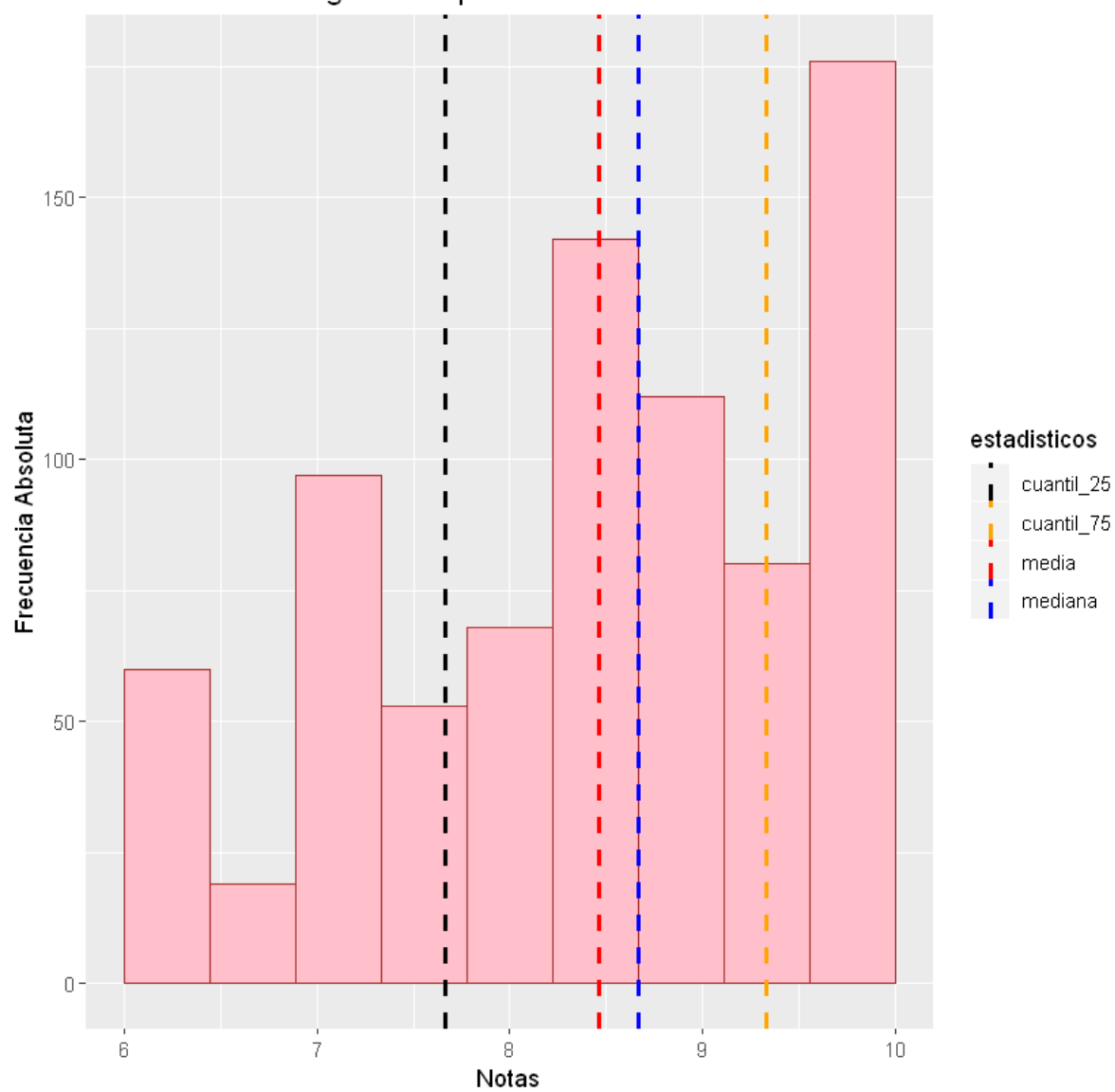


Gráfico 8

Primaria - Histograma de promedio de Lengua - 1 año.

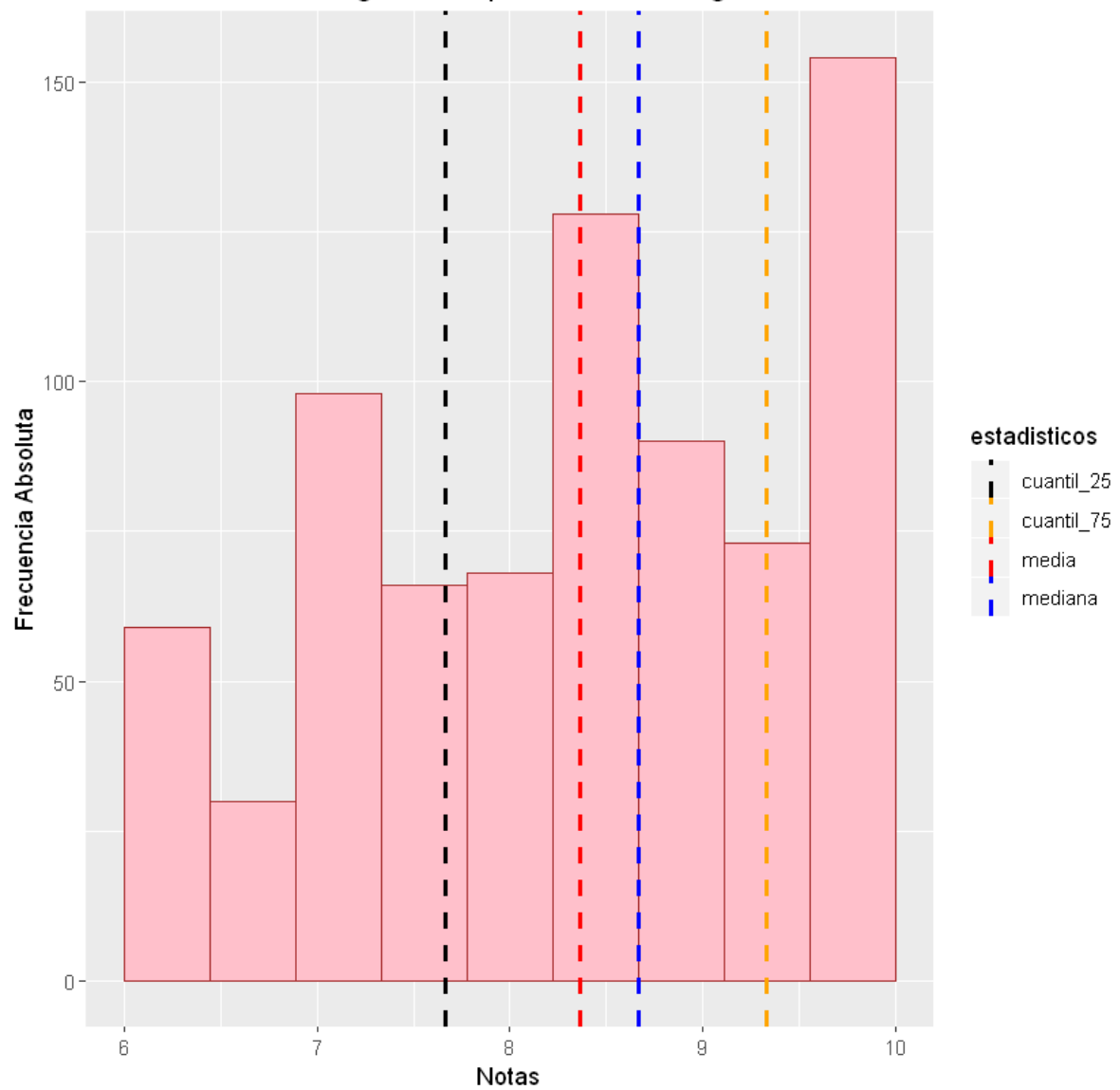


Gráfico 9

Primaria - Histograma de promedio de Matemática - 2 año.

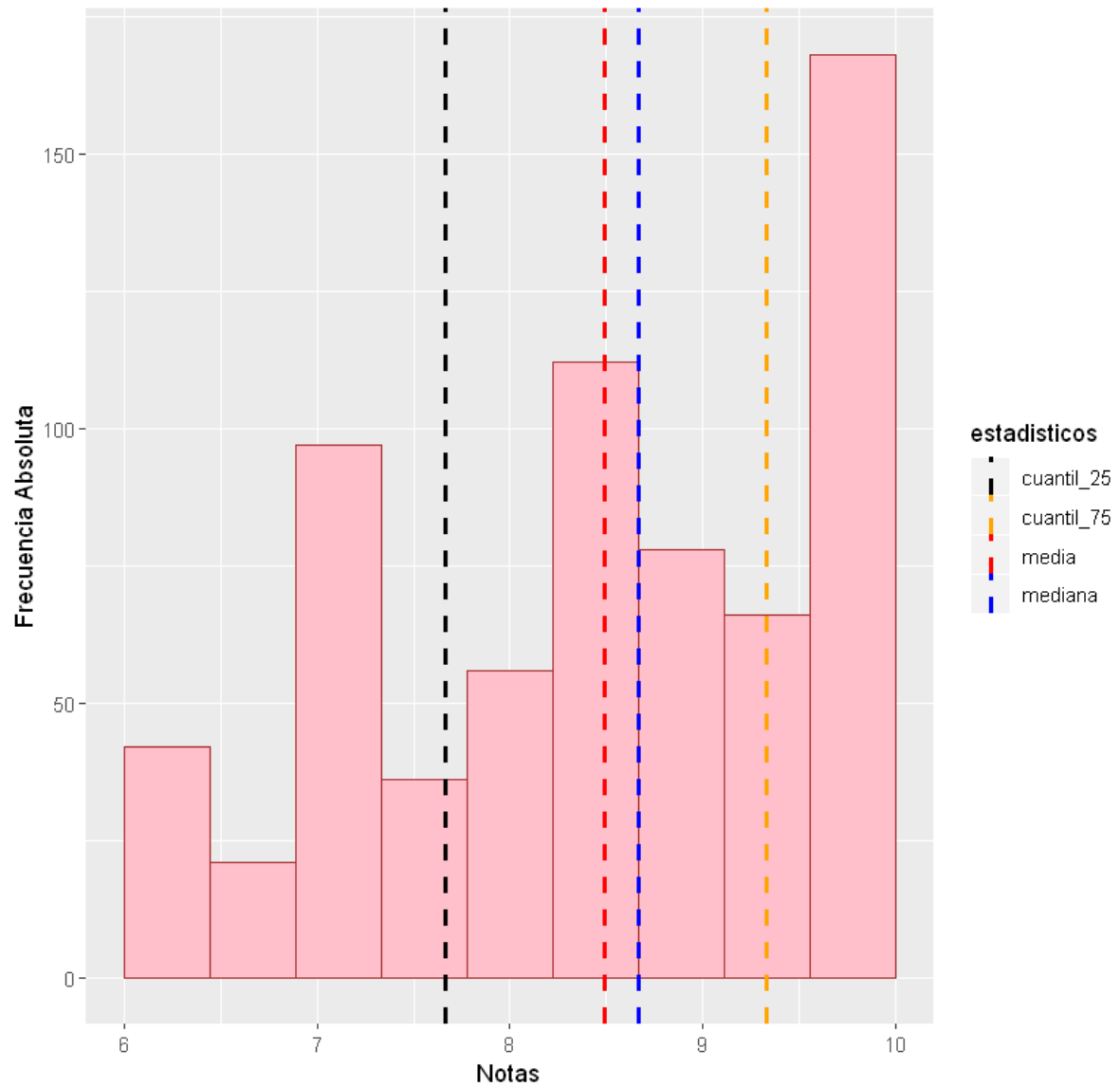


Gráfico 10

Primaria - Histograma de promedio de Lengua - 2 año.

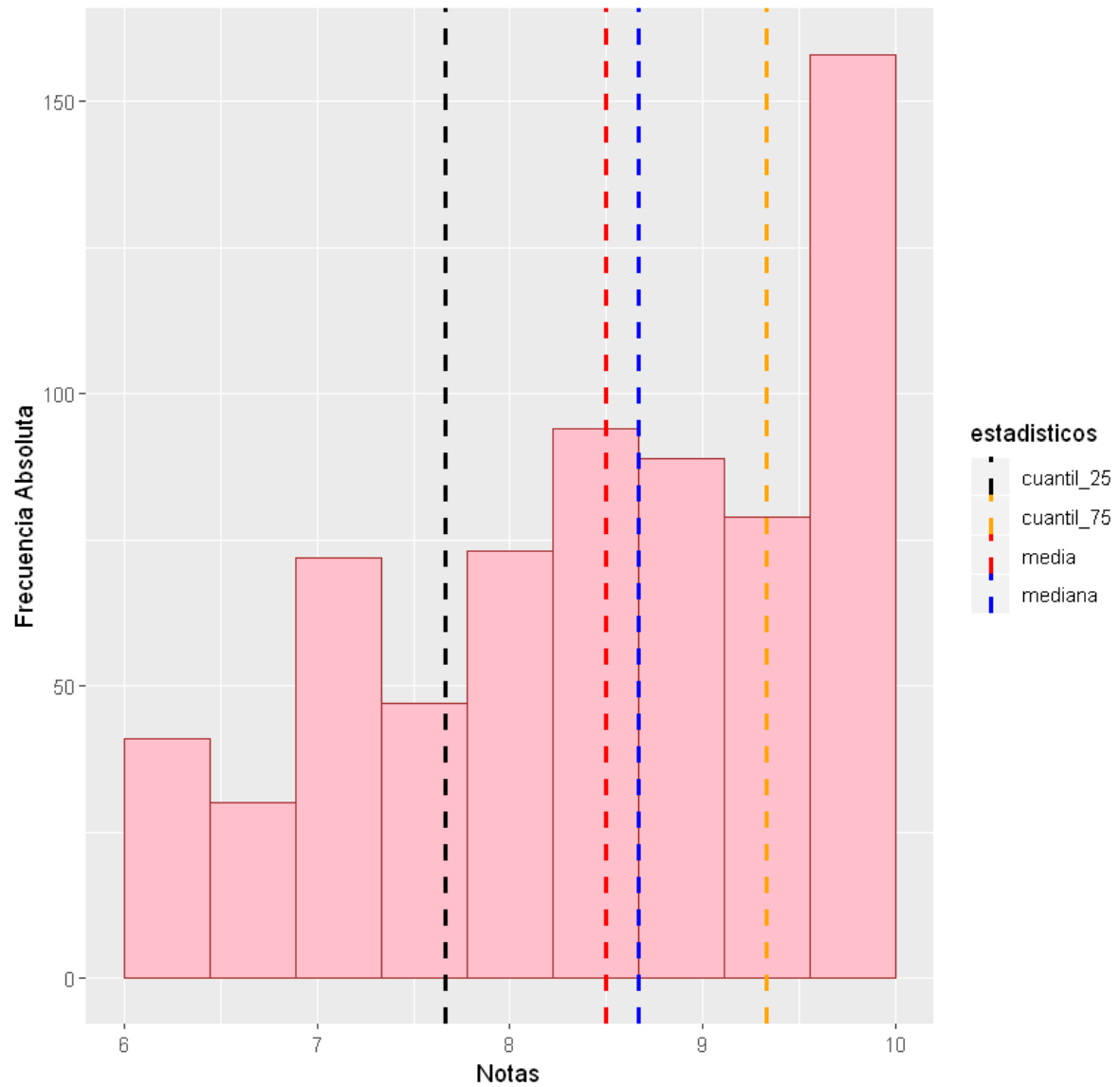


Gráfico 11

Primaria - Histograma de promedio de Matemática - 3 año.

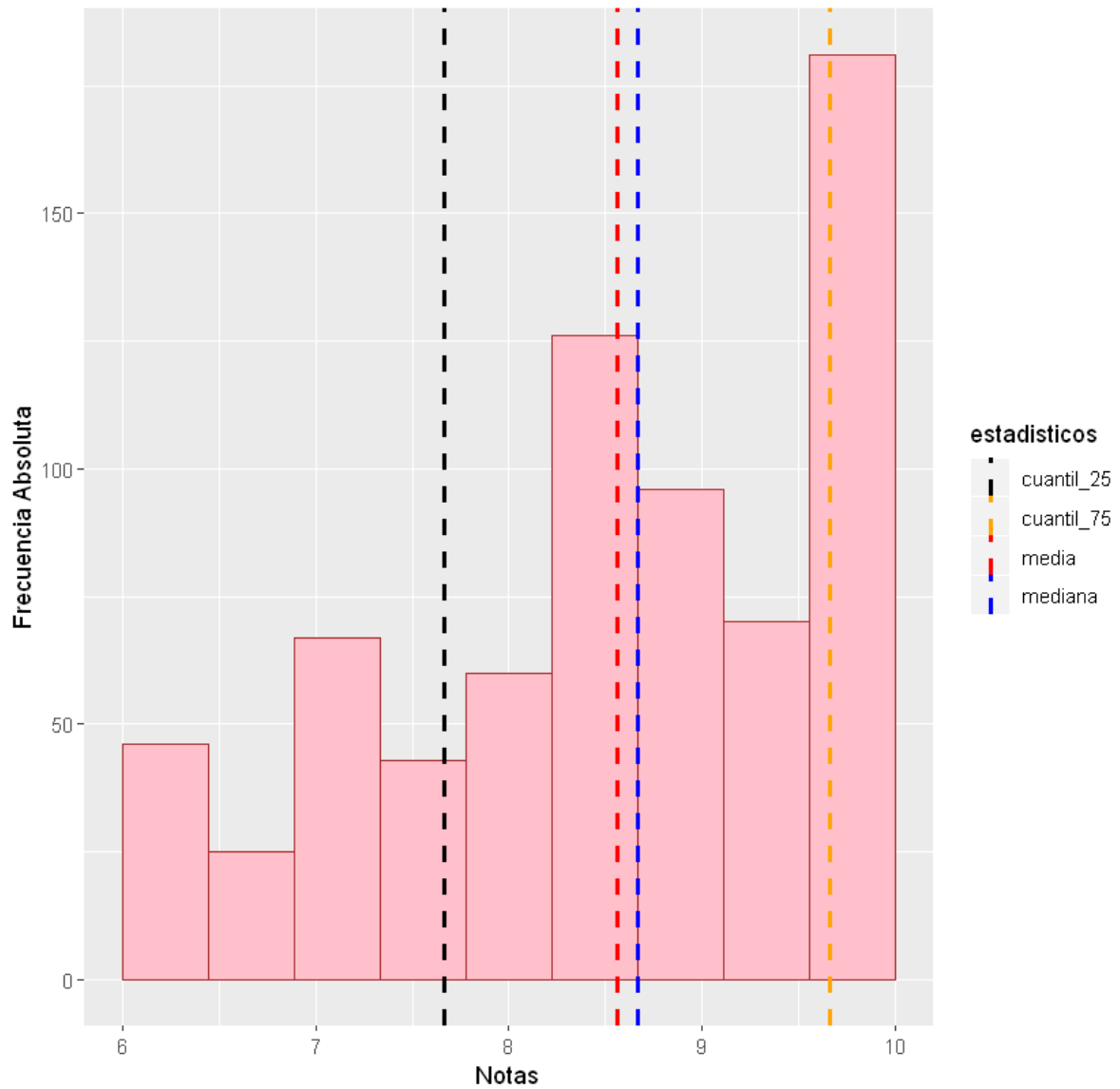


Gráfico 12

Primaria - Histograma de promedio de Lengua - 3 año.

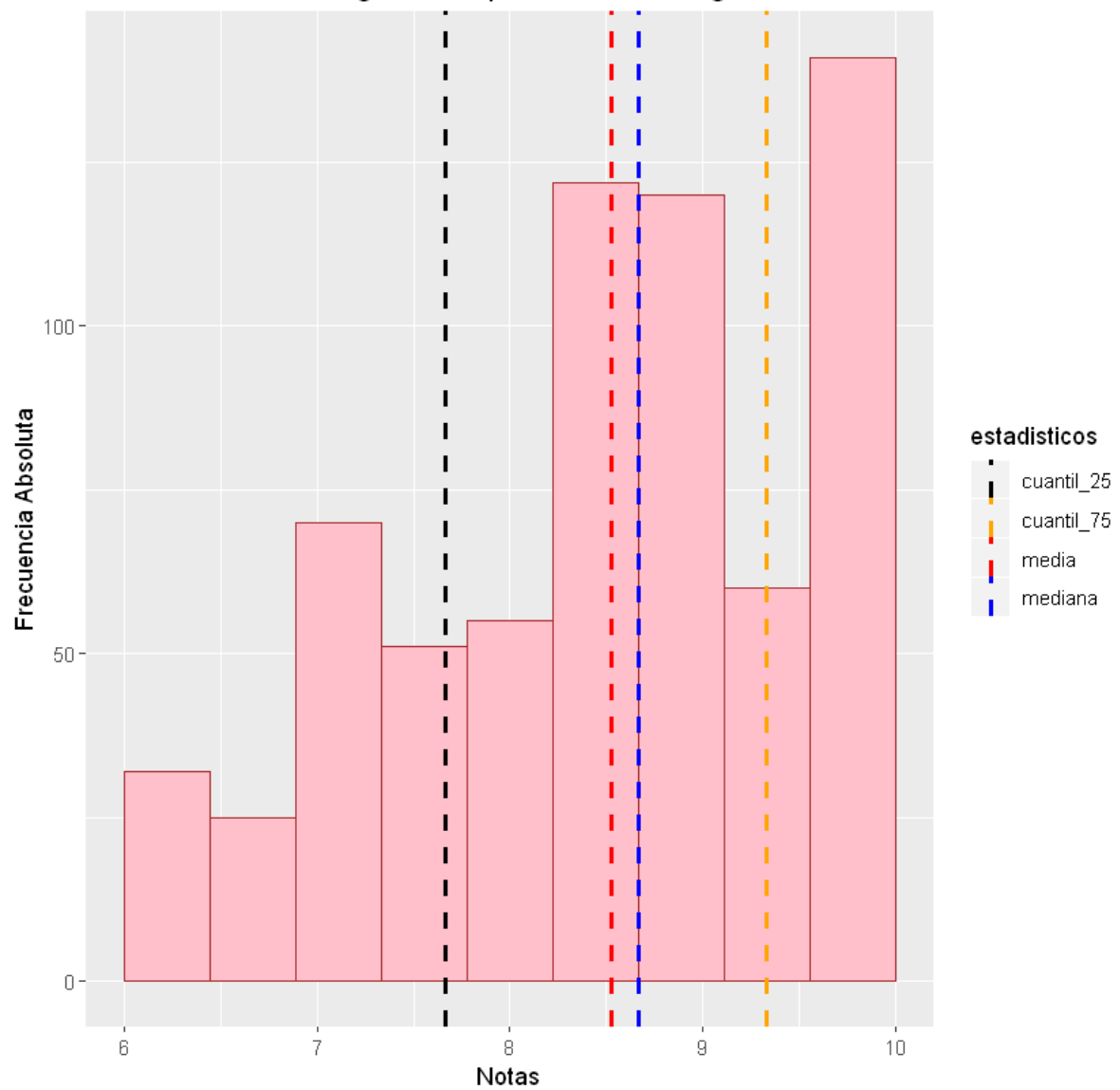


Gráfico 13

Primaria - Histograma de promedio de Matemática - 4 año.

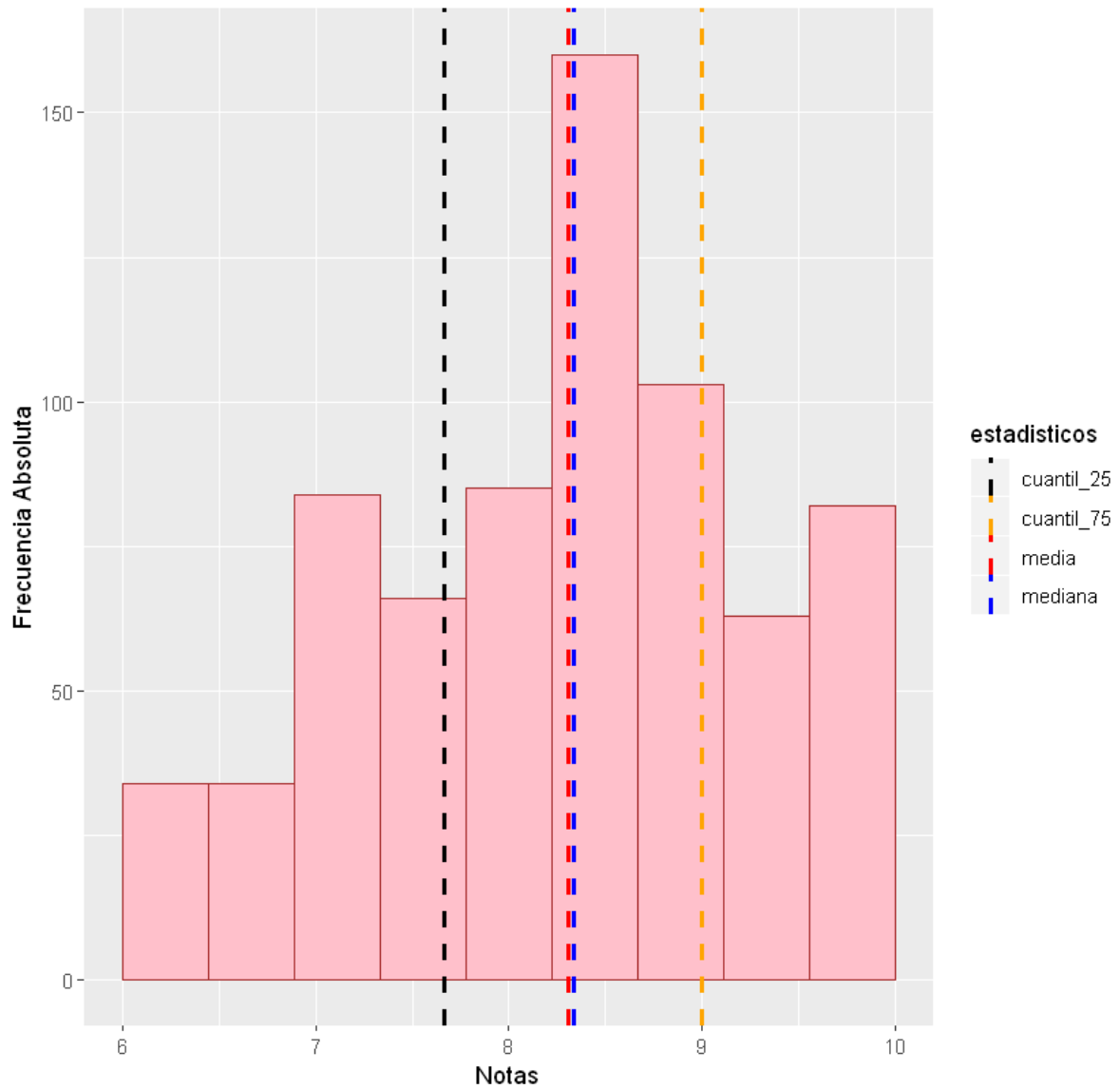


Gráfico 14

Primaria - Histograma de promedio de Lengua - 4 año.

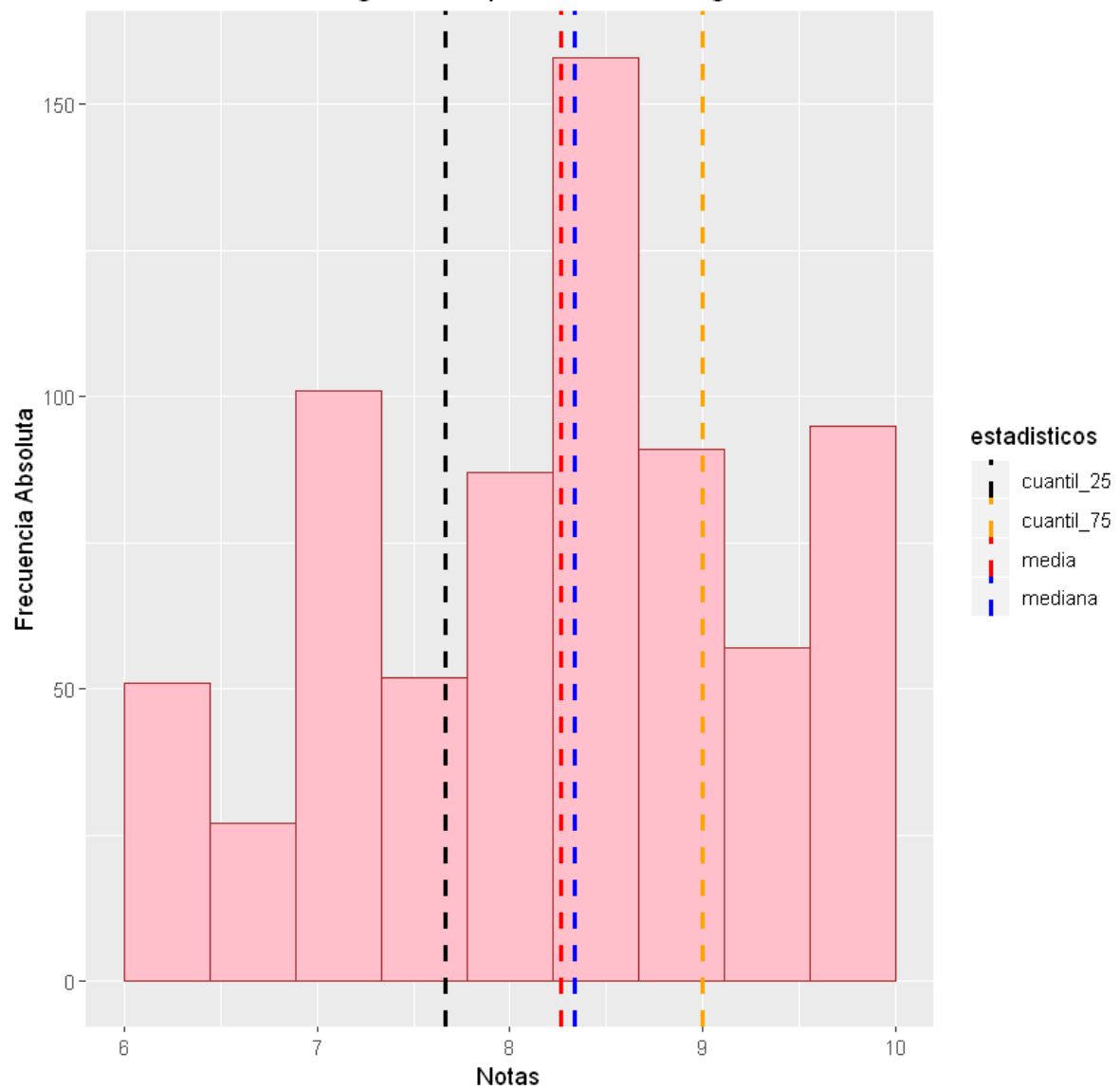


Gráfico 15

Primaria - Histograma de promedio de Matemática - 5 año.

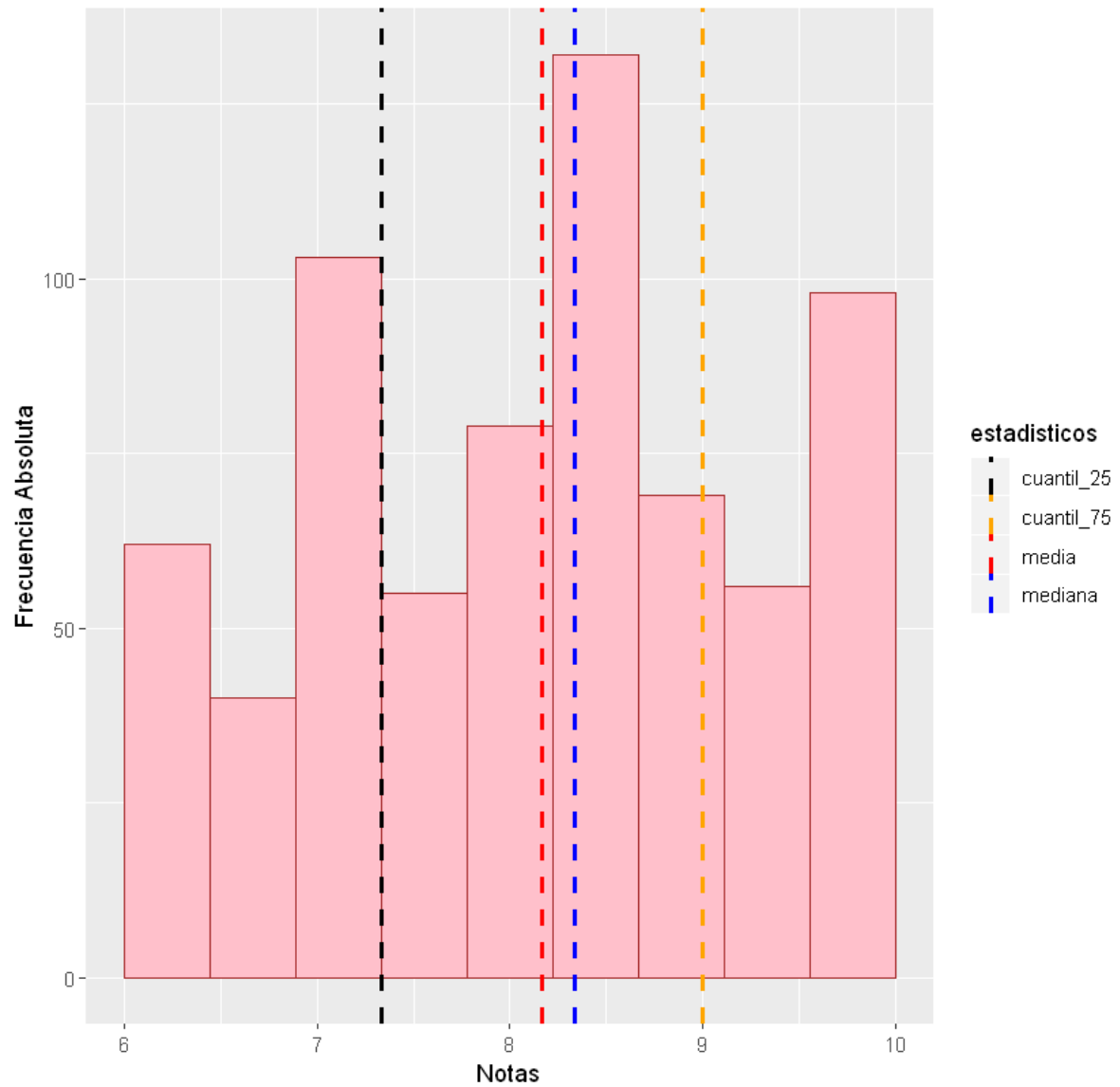


Gráfico 16

Primaria - Histograma de promedio de Lengua - 5 año.

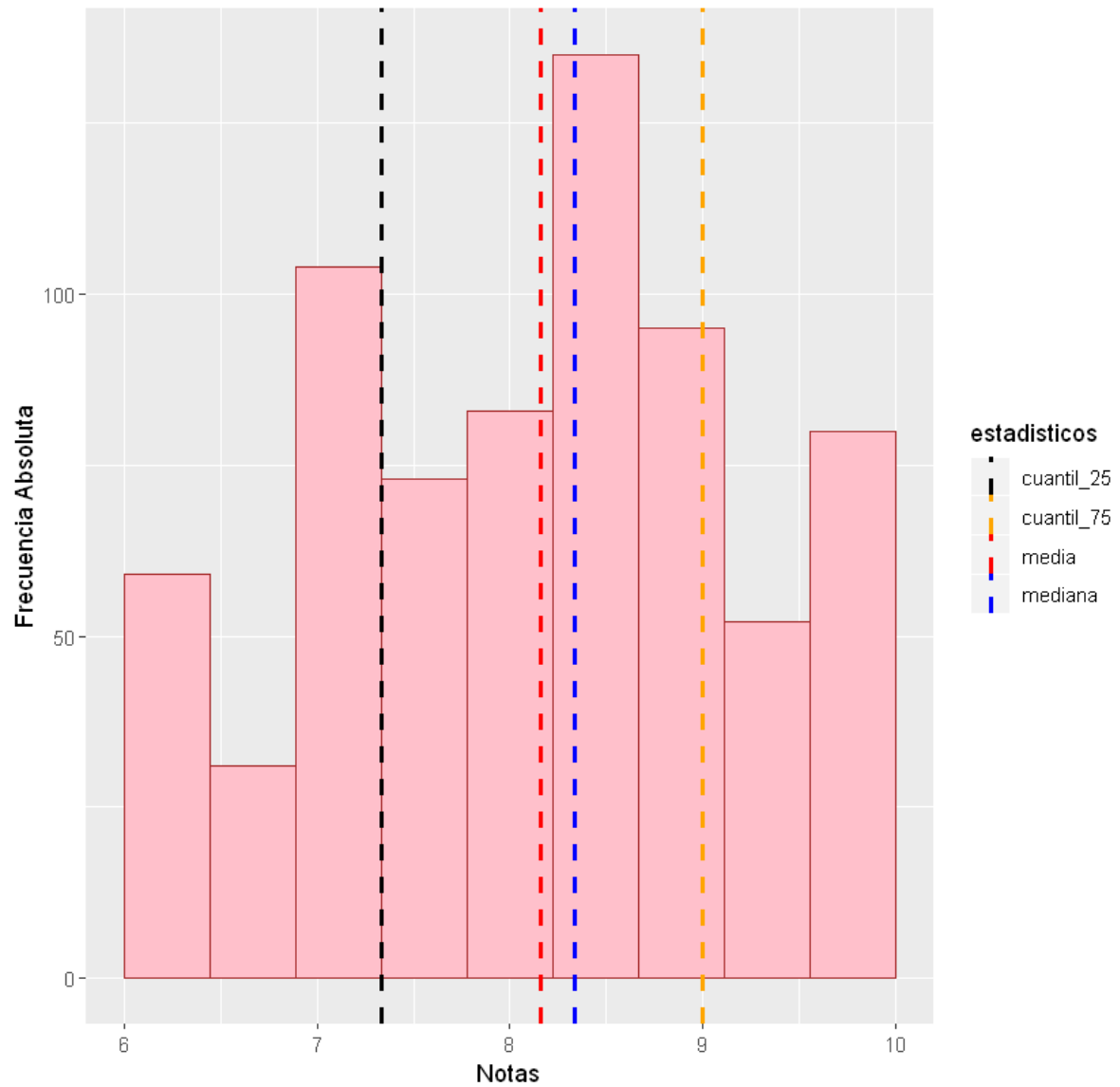


Gráfico 17

Primaria - Histograma de promedio de Matemática - 6 año.

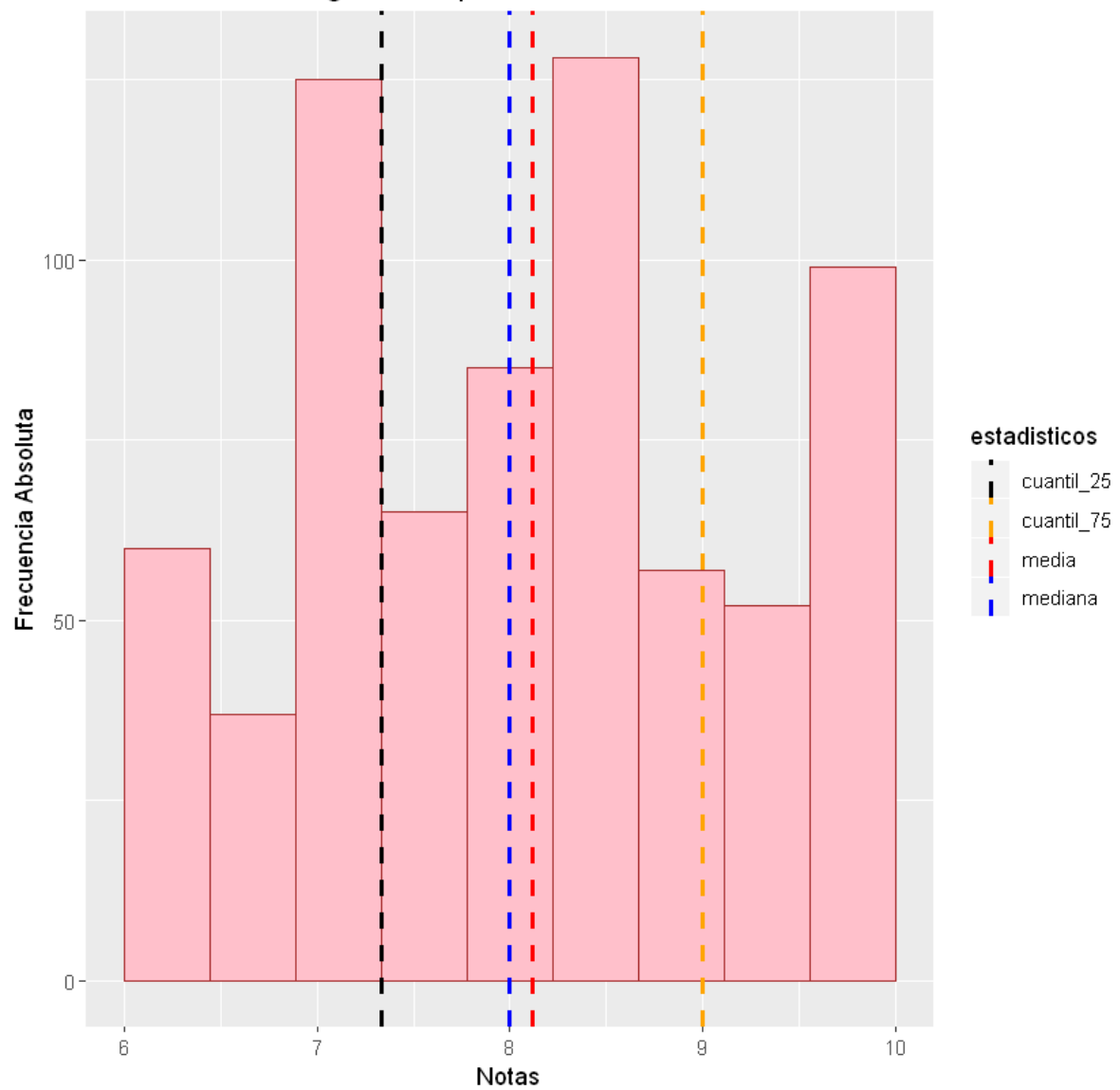


Gráfico 18

Primaria - Histograma de promedio de Lengua - 6 año.

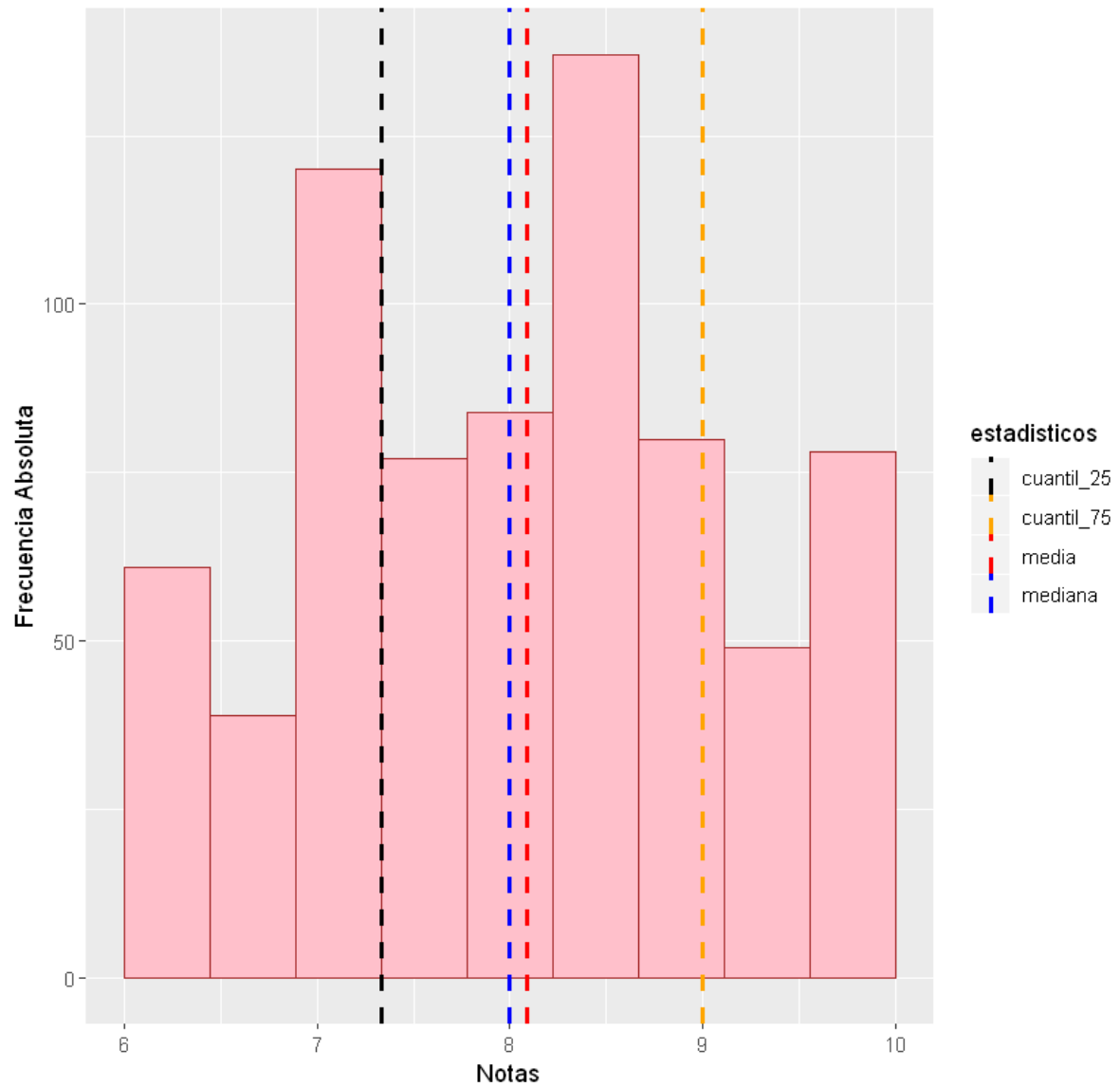


Gráfico 19

Secundaria - Histograma de promedio de Matemática - 1 año.

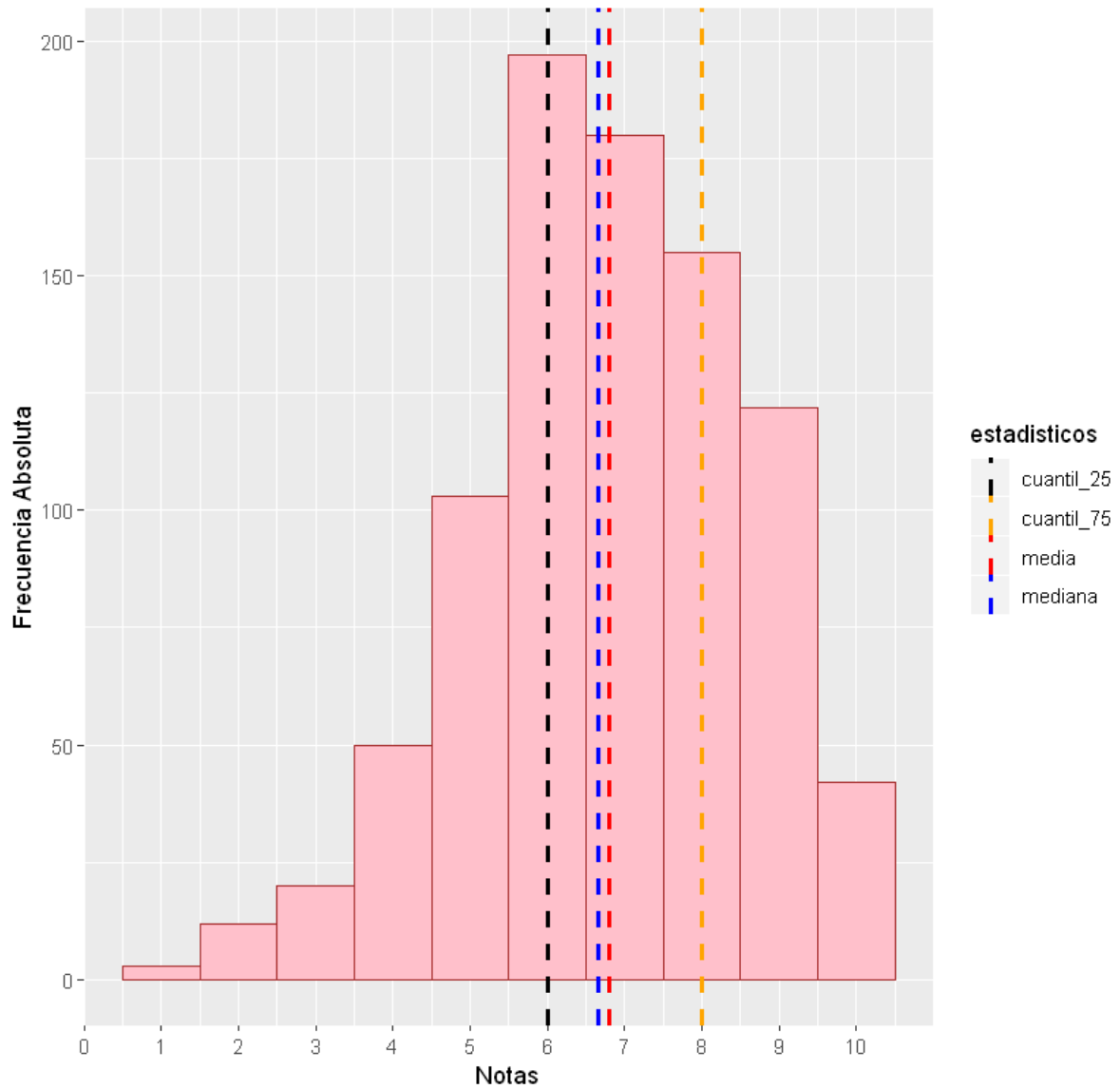


Gráfico 20

Secundaria - Histograma de promedio de Lengua y Literatura - 1 año.

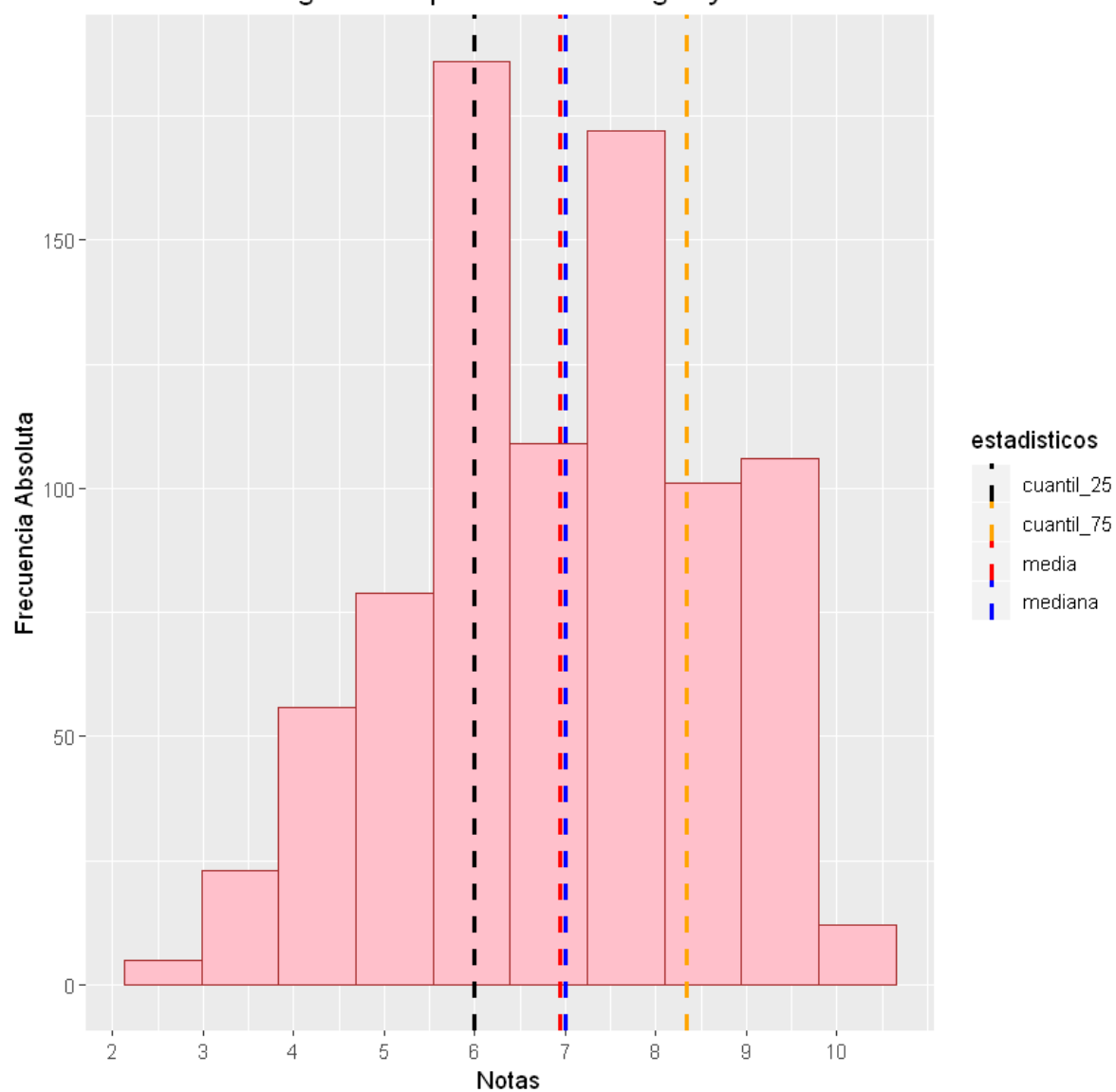


Gráfico 21

Secundaria - Histograma de promedio de Matemática - 2 año.

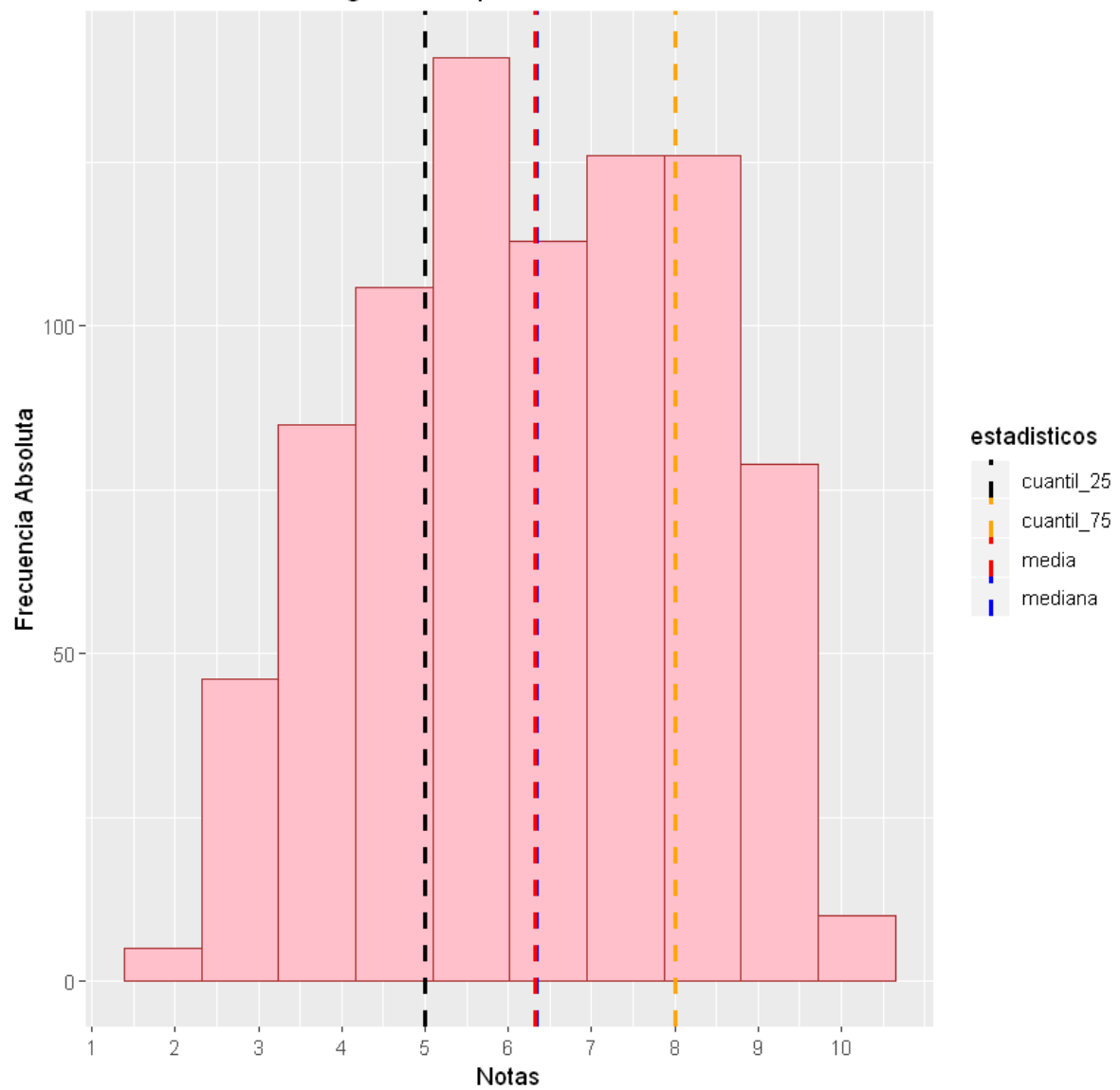


Gráfico 22

Secundaria - Histograma de promedio de Lengua y Literatura - 2 año.

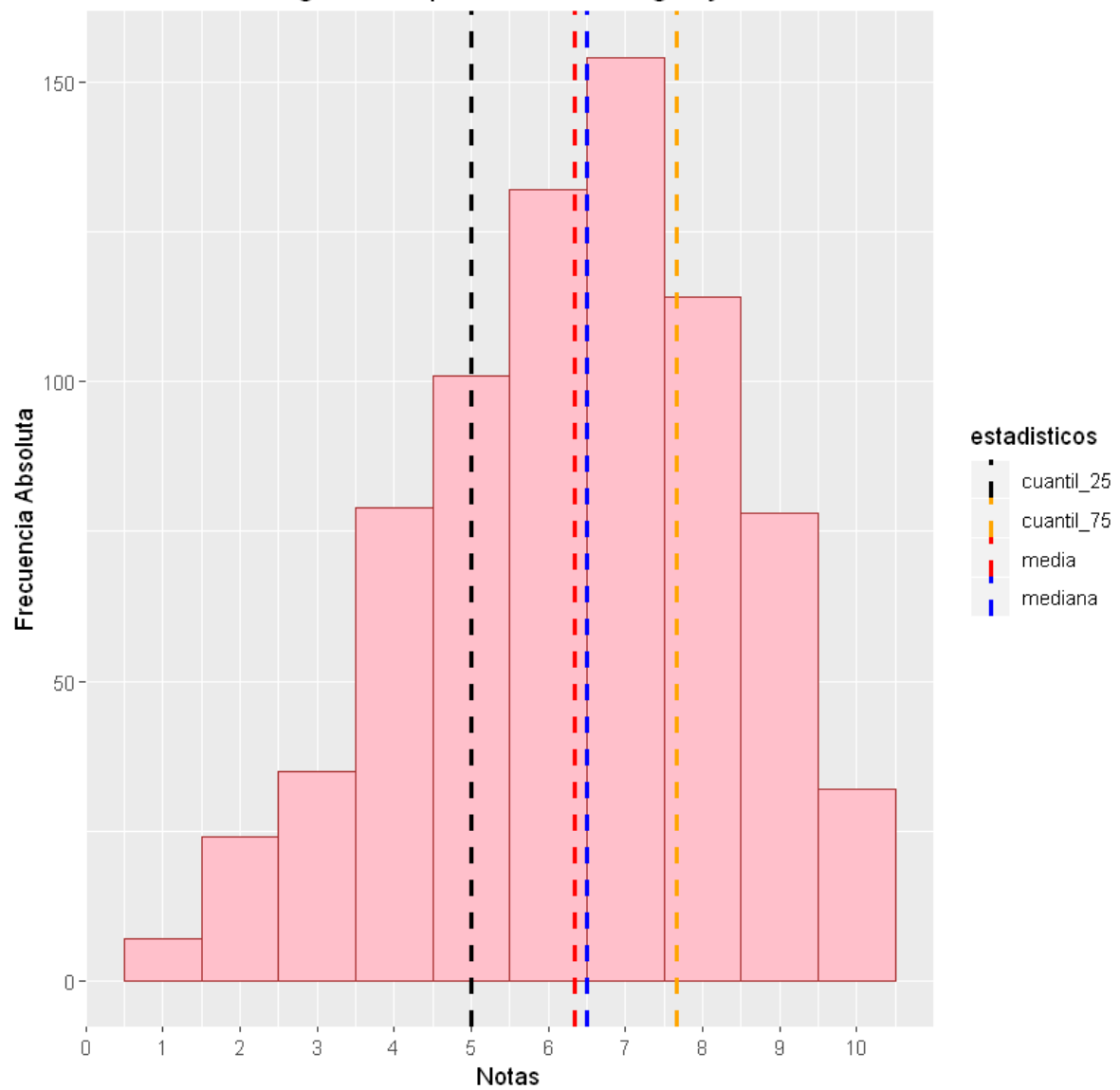


Gráfico 23

Secundaria - Histograma de promedio de Matemática - 3 año.

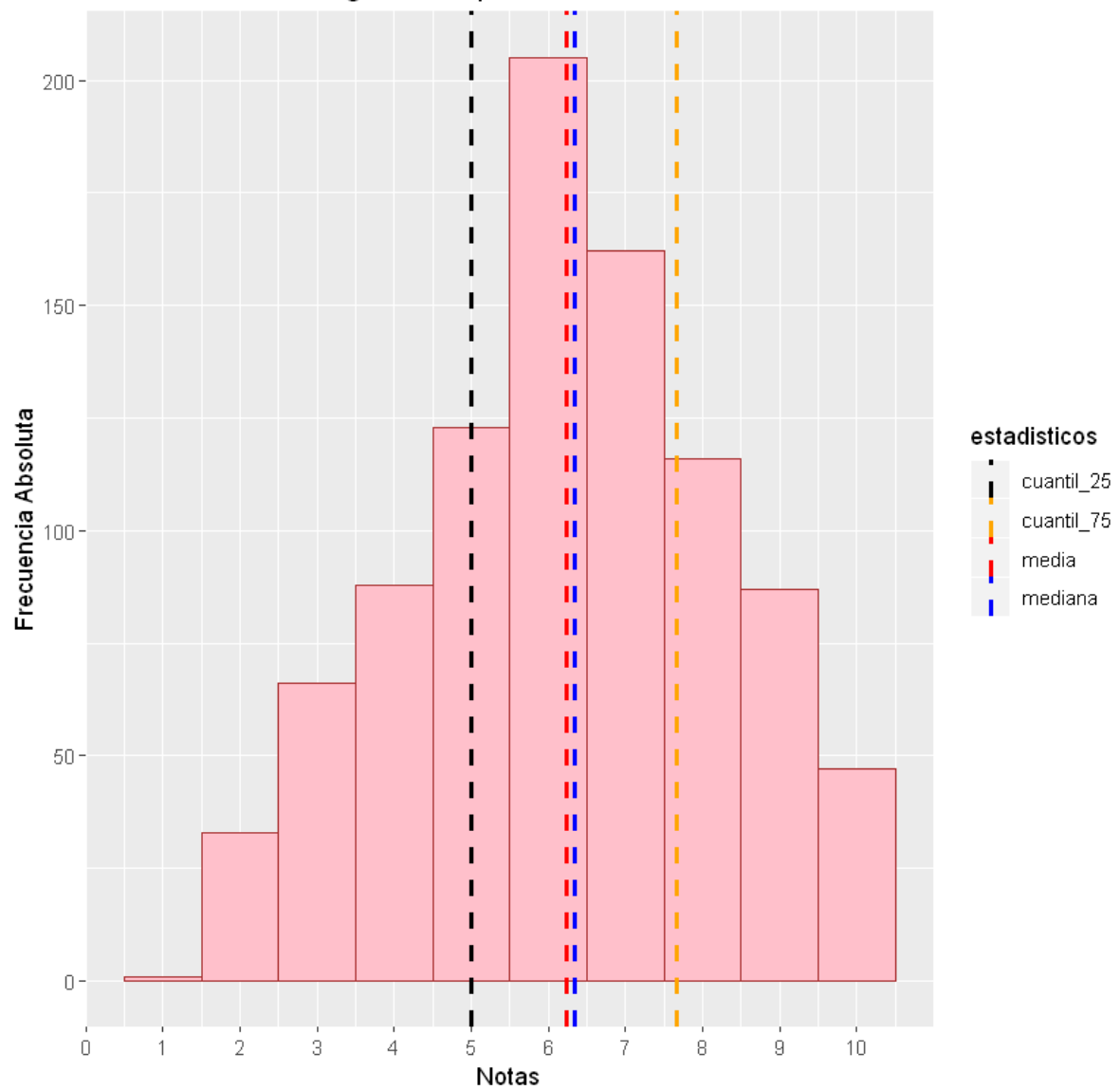


Gráfico 24

Secundaria - Histograma de promedio de Lengua y Literatura - 3 año.

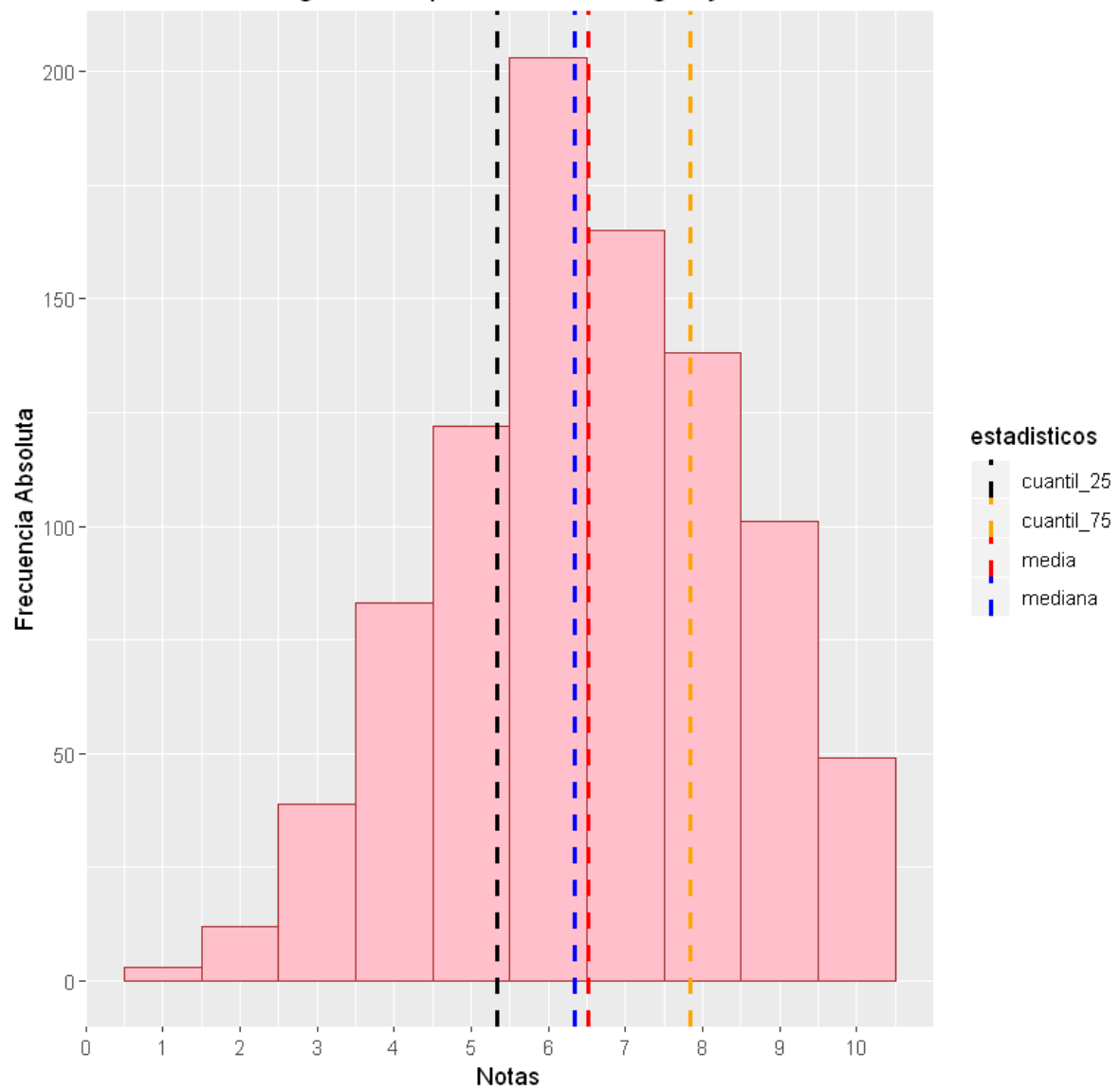


Gráfico 25

Secundaria - Histograma de promedio de Matemática - 4 año.

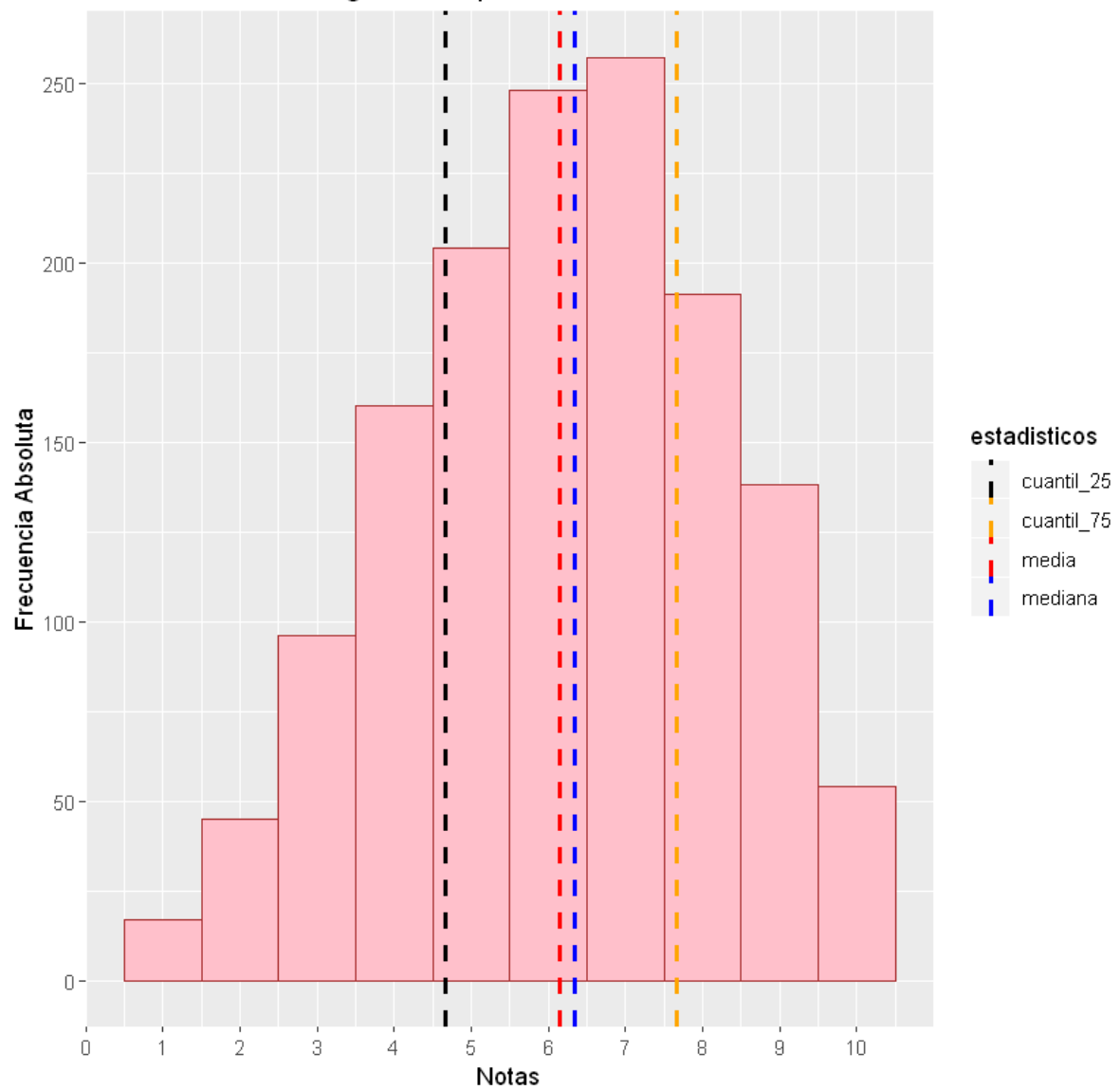


Gráfico 26

Secundaria - Histograma de promedio de Lengua y Literatura - 4 año.

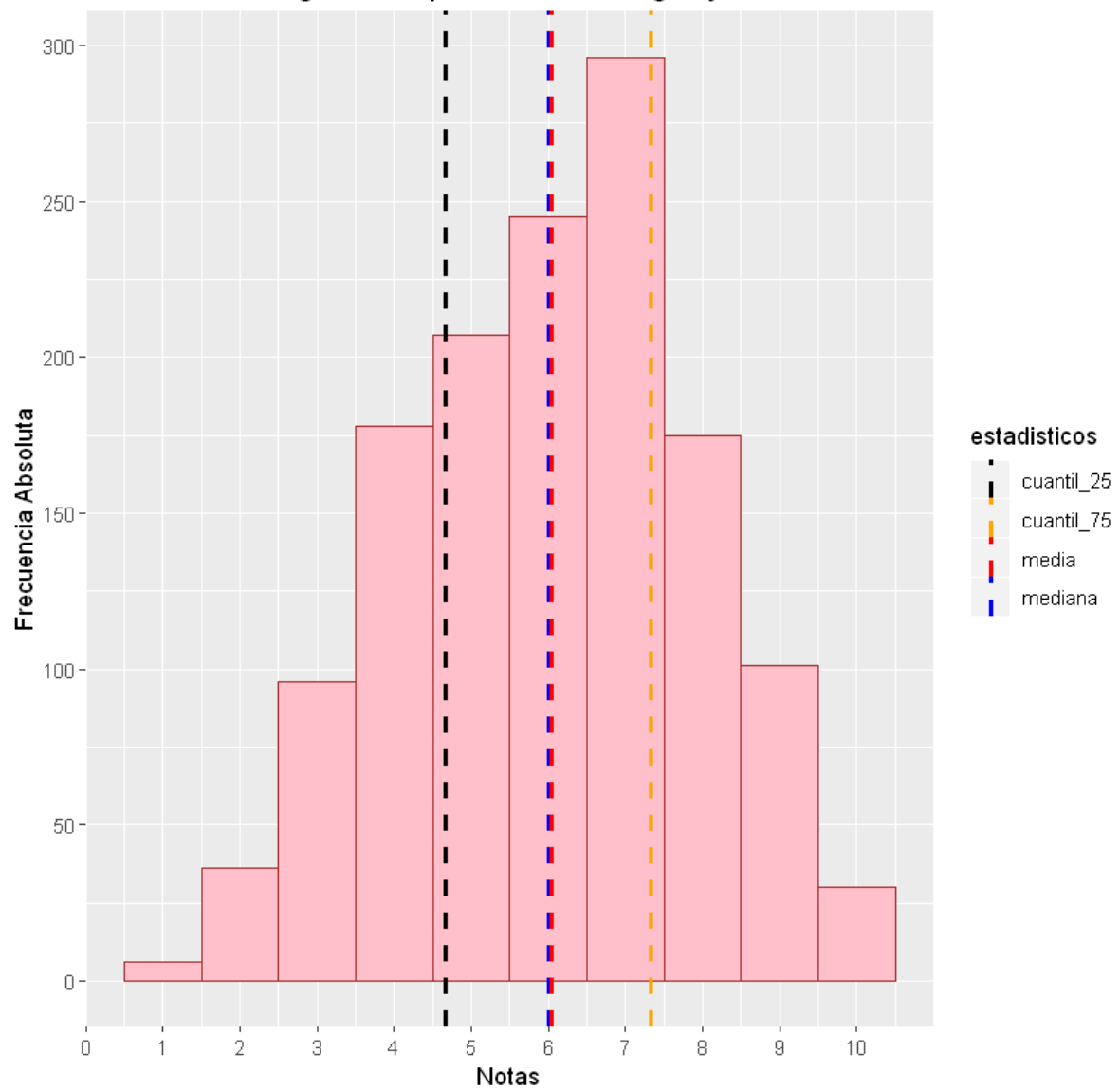


Gráfico 27

Secundaria - Histograma de promedio de Matemática - 5 año.

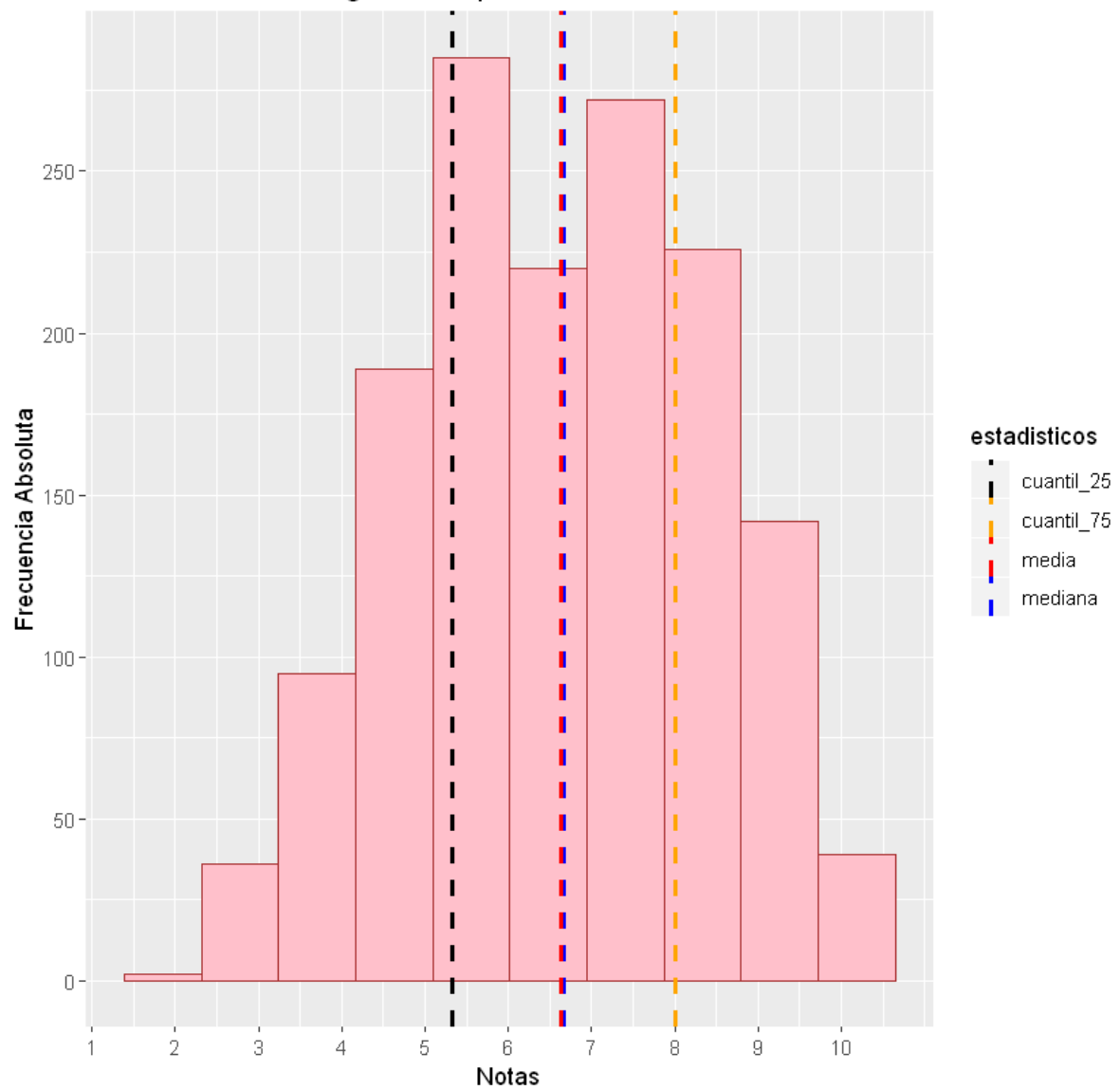


Gráfico 28

Secundaria - Histograma de promedio de Lengua y Literatura - 5 año.

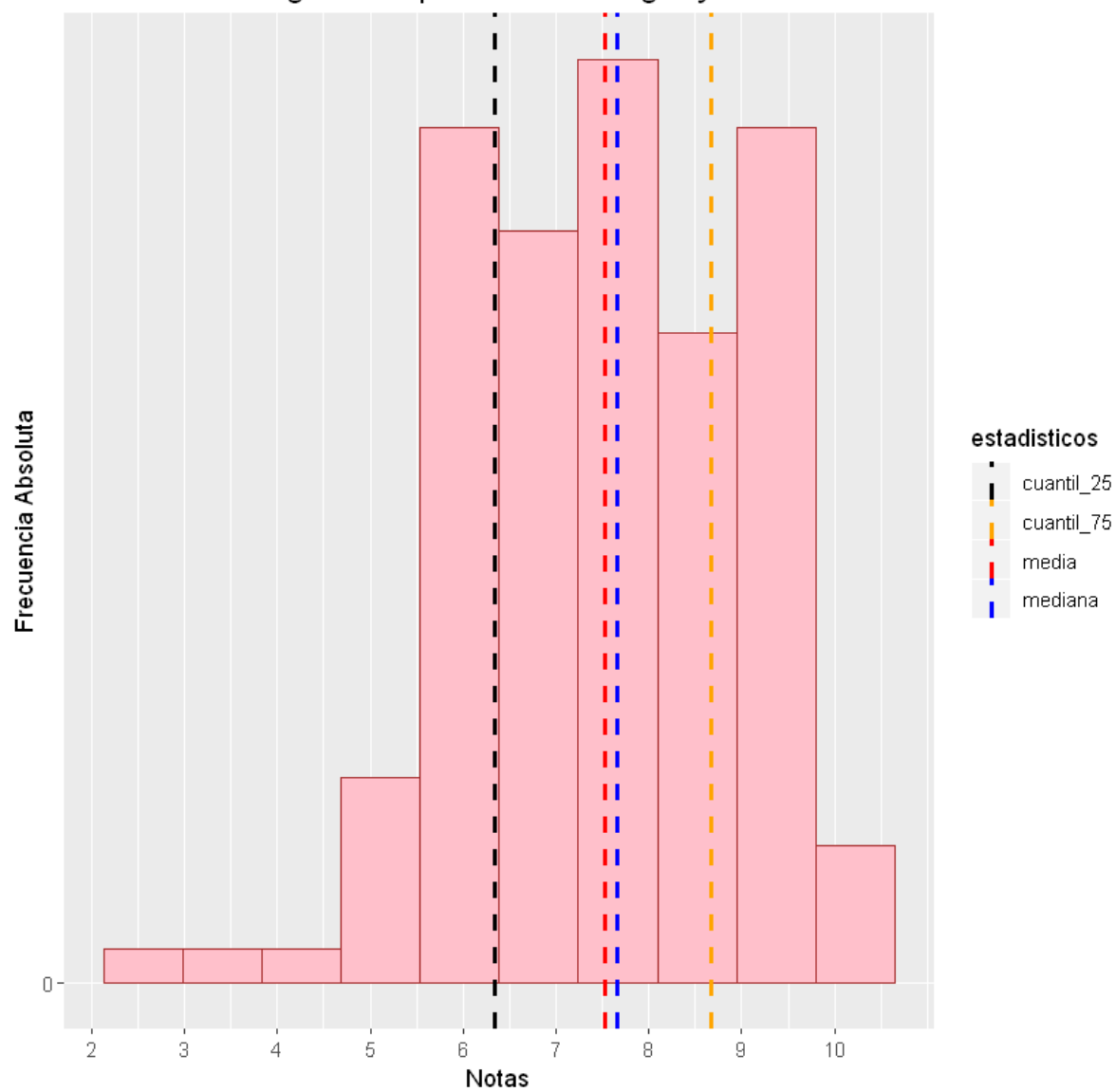


Gráfico 29

Secundaria - Histograma de promedio de Matemática - 6 año.

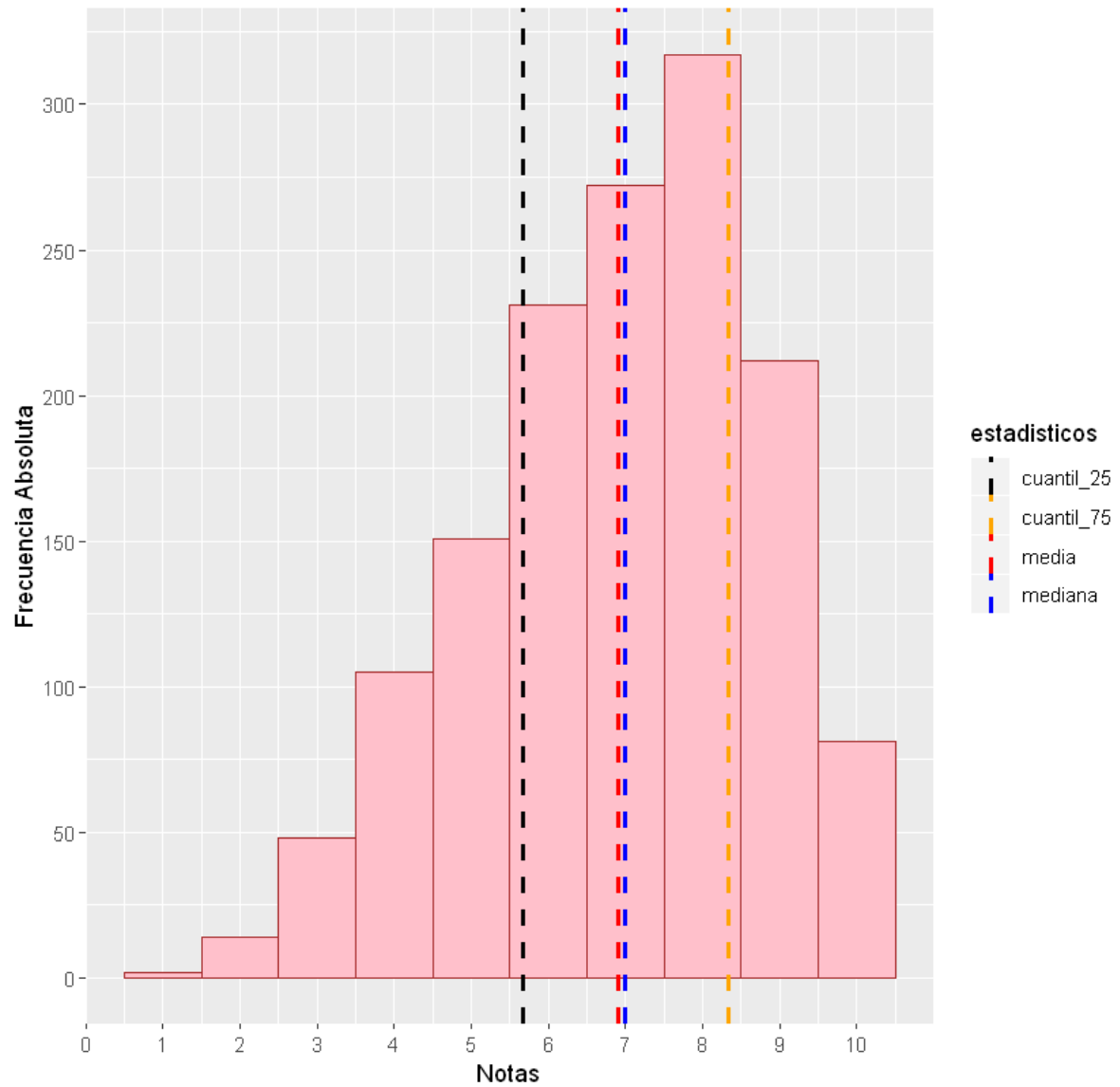


Gráfico 30

Secundaria - Histograma de promedio de Lengua y Literatura - 6 año.

