

Análisis Temporal de la Matrícula en Educación Primaria a Nivel Mundial (1970–2023)

Salvador Enrique Rodríguez Hernández (rh06006)

2025-06-11

Índice

1	Introducción	2
2	Análisis gráfico de la serie..	2
2.1	Análisis de tendencia, variabilidad, estacionalidad etc	4
3	Transformación para estacionariedad	5
3.1	Análisis de los resultados	6
4	Estimación de parámetros:	7
4.1	Proceso de estimación para los modelos propuestos	7
5	Selección del mejor modelo:	7
5.1	Diagnóstico del modelo	7
5.2	Predicciones con el modelo seleccionado	8
6	Conclusiones	8
7	Conclusiones	13

1 Introducción

El presente informe tiene como objetivo analizar la evolución histórica de la matrícula en educación primaria a nivel mundial durante el período comprendido entre 1970 y 2023. Para ello, se utiliza un conjunto de datos proporcionado por el **UNESCO Institute for Statistics (UIS)**, disponible públicamente a través del portal de datos del **Banco Mundial**, bajo la licencia **CC BY-4.0**.

El indicador analizado corresponde al número total de alumnos inscritos en el nivel primario, tanto en instituciones públicas como privadas. Los datos se recopilan de los informes oficiales que los países miembros remiten a la UNESCO mediante su encuesta anual de educación. Además, se han armonizado de acuerdo con la **Clasificación Internacional Normalizada de la Educación (CINE)** para garantizar la comparabilidad internacional.

La frecuencia de los datos es anual, y los valores están expresados en millones de alumnos. Este análisis busca identificar patrones de crecimiento, posibles cambios estructurales en la serie, y proponer un modelo ARIMA que permita realizar proyecciones confiables de la matrícula escolar en los próximos años.

2 Análisis gráfico de la serie..

A continuación se hará la presentación de la gráfica de la serie de datos originales

##Gráfica de la serie Colocar gráfica aquí

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.5.1      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.1
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts
```

```
library(forecast)
```

```
## Warning: package 'forecast' was built under R version 4.4.3
```

```
## Registered S3 method overwritten by 'quantmod':
##   method      from
## as.zoo.data.frame zoo
```

```
library(tsibble)
```

```
## Warning: package 'tsibble' was built under R version 4.4.3
```

```
## Registered S3 method overwritten by 'tsibble':  
##   method                from  
##   as_tibble.grouped_df dplyr  
##  
## Adjuntando el paquete: 'tsibble'  
##  
## The following object is masked from 'package:lubridate':  
##  
##   interval  
##  
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, union
```

```
library(lubridate)
```

```
library(ggfortify)
```

```
## Warning: package 'ggfortify' was built under R version 4.4.3
```

```
## Registered S3 methods overwritten by 'ggfortify':  
##   method                from  
##   autoplot.Arima         forecast  
##   autoplot.acf           forecast  
##   autoplot.ar            forecast  
##   autoplot.bats          forecast  
##   autoplot.decomposed.ts forecast  
##   autoplot.ets           forecast  
##   autoplot.forecast      forecast  
##   autoplot.stl           forecast  
##   autoplot.ts            forecast  
##   fitted.ar              forecast  
##   fortify.ts             forecast  
##   residuals.ar           forecast
```

```
library(readr)
```

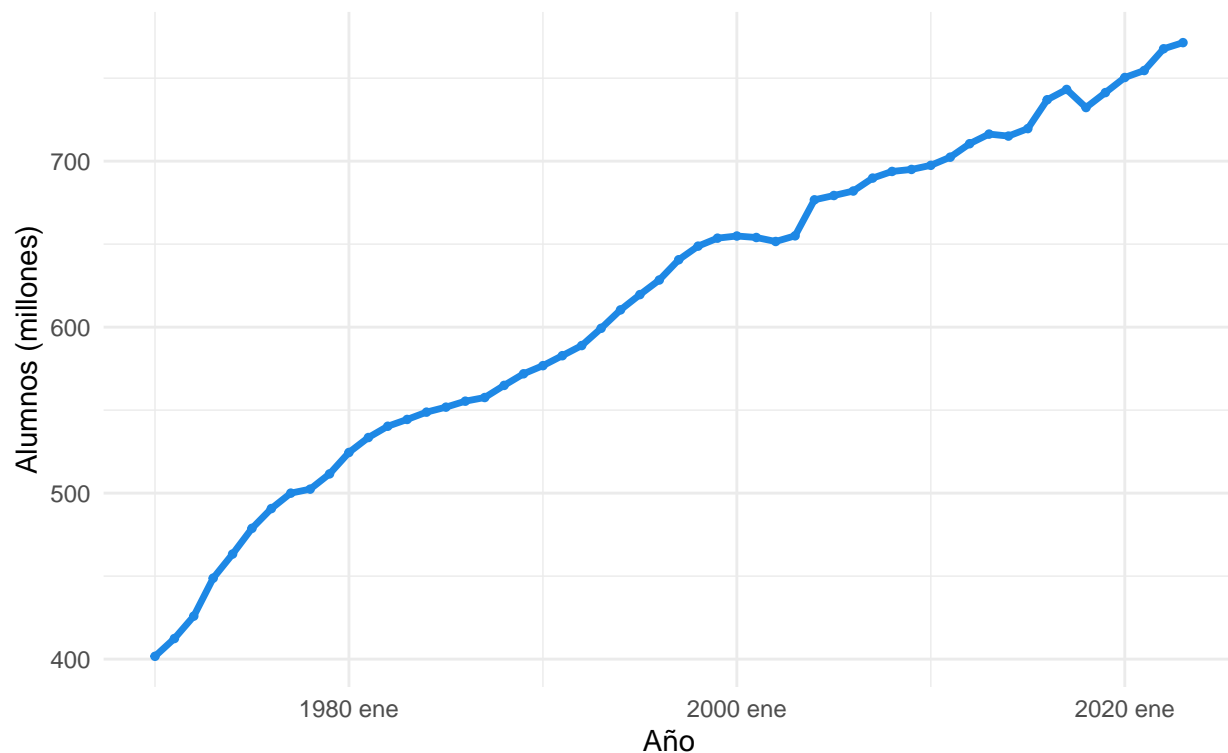
```
datos <- read_csv("datos_educacion_primaria.csv") %>%  
  mutate(Año = yearmonth(paste(Año, "-01"))) %>%  
  as_tsibble(index = Año)
```

```
## Rows: 54 Columns: 2
## -- Column specification -----
## Delimiter: ","
## dbl (2): Año, Alumnos_millones
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
ts_original <- ts(datos$Alumnos_millones, start = 1970, frequency = 1)
ggplot(datos, aes(x = Año, y = Alumnos_millones)) +
  geom_line(color = "#1E88E5", linewidth = 1.2) +
  geom_point(color = "#1E88E5", size = 1) +
  labs(title = "Matrícula en Educación Primaria (1970-2023)",
       subtitle = "Serie original en millones de alumnos",
       x = "Año", y = "Alumnos (millones)") +
  theme_minimal()
```

Matrícula en Educación Primaria (1970-2023)

Serie original en millones de alumnos



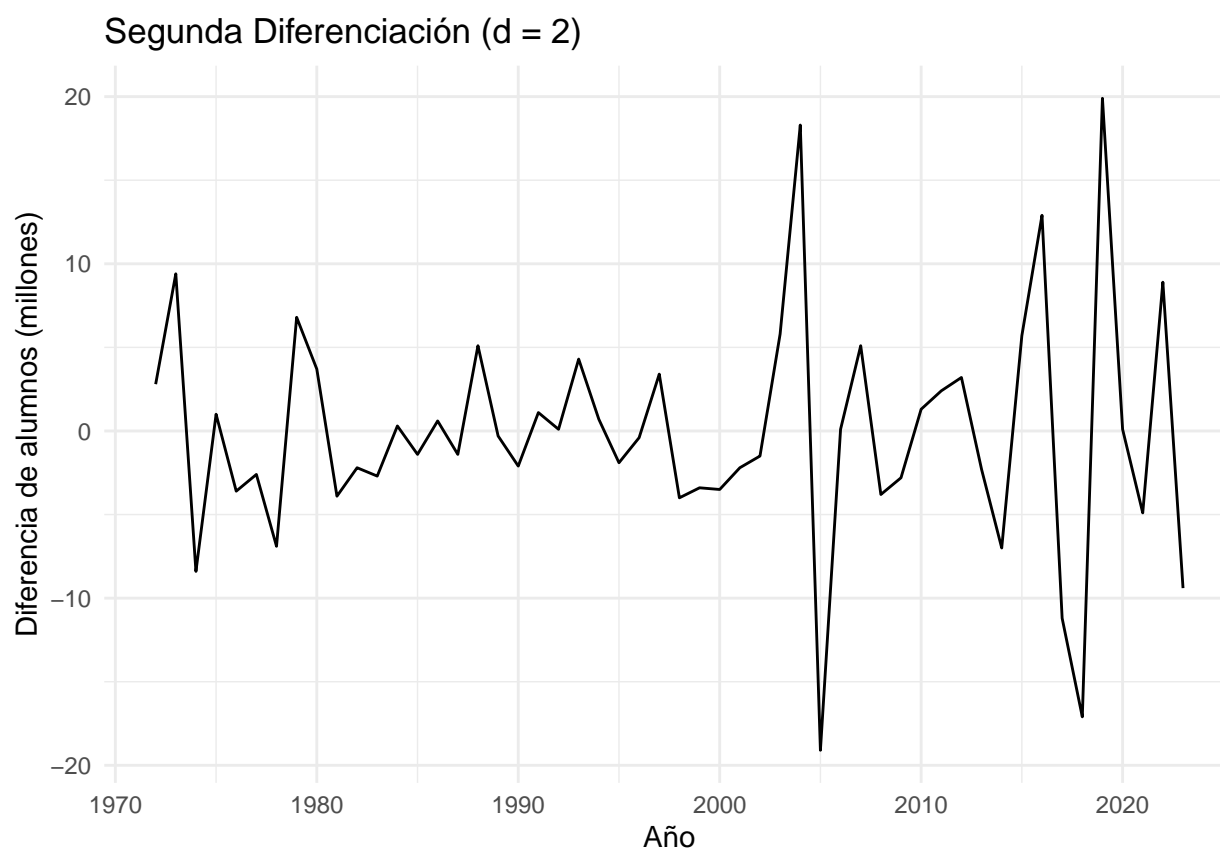
2.1 Análisis de tendencia, variabilidad, estacionalidad etc

Análisis aquí

3 Transformación para estacionariedad

dasdasda ## Justificación del tipo de transformación aplicada dkjdskd ## Proceso para convertir la serie en estacionaria dsdsdasda

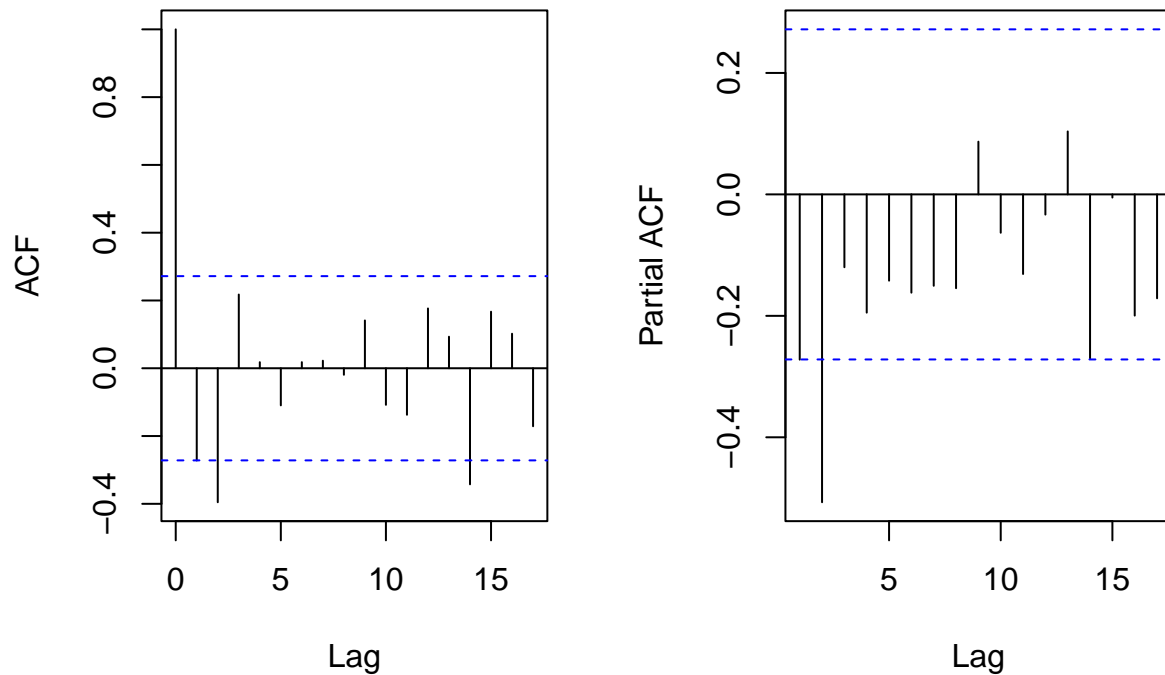
```
ts_diff1 <- diff(ts_original)
ts_diff2 <- diff(ts_diff1)
autoplot(ts_diff2) +
  labs(title = "Segunda Diferenciación (d = 2)",
        x = "Año", y = "Diferencia de alumnos (millones)") +
  theme_minimal()
```



Análisis de autocorrelación: ## Gráfico de la función de autocorrelación simple (ACF) y parcial (PACF) dsadsa

```
par(mfrow = c(1, 2))
acf(ts_diff2, na.action = na.pass, main = "ACF - Segunda Diferenciación (d=2)")
pacf(ts_diff2, na.action = na.pass, main = "PACF - Segunda Diferenciación (d=2)")
```

ACF – Segunda Diferenciación (d: PACF – Segunda Diferenciación (d



```
par(mfrow = c(1, 1))
```

3.1 Análisis de los resultados

dfsdfsdfsdf # Propuesta de modelos ARIMA: ## Posibles modelos ARIMA(p, d, q) × (P, D, Q)s

```
modelo <- auto.arima(ts_original, d = 2, seasonal = FALSE, stepwise = FALSE, approximat
summary(modelo)
```

```
## Series: ts_original
## ARIMA(1,2,2)
##
## Coefficients:
##          ar1      ma1      ma2
##       -0.5720  0.0300 -0.7850
## s.e.    0.2132  0.1747  0.1323
##
## sigma^2 = 29.99: log likelihood = -161.51
```

```
## AIC=331.01    AICc=331.86    BIC=338.82
##
## Training set error measures:
##              ME      RMSE      MAE      MPE      MAPE      MASE
## Training set -0.8903757 5.216934 3.821428 -0.1575164 0.6197798 0.5057071
##              ACF1
## Training set -0.04826175
```

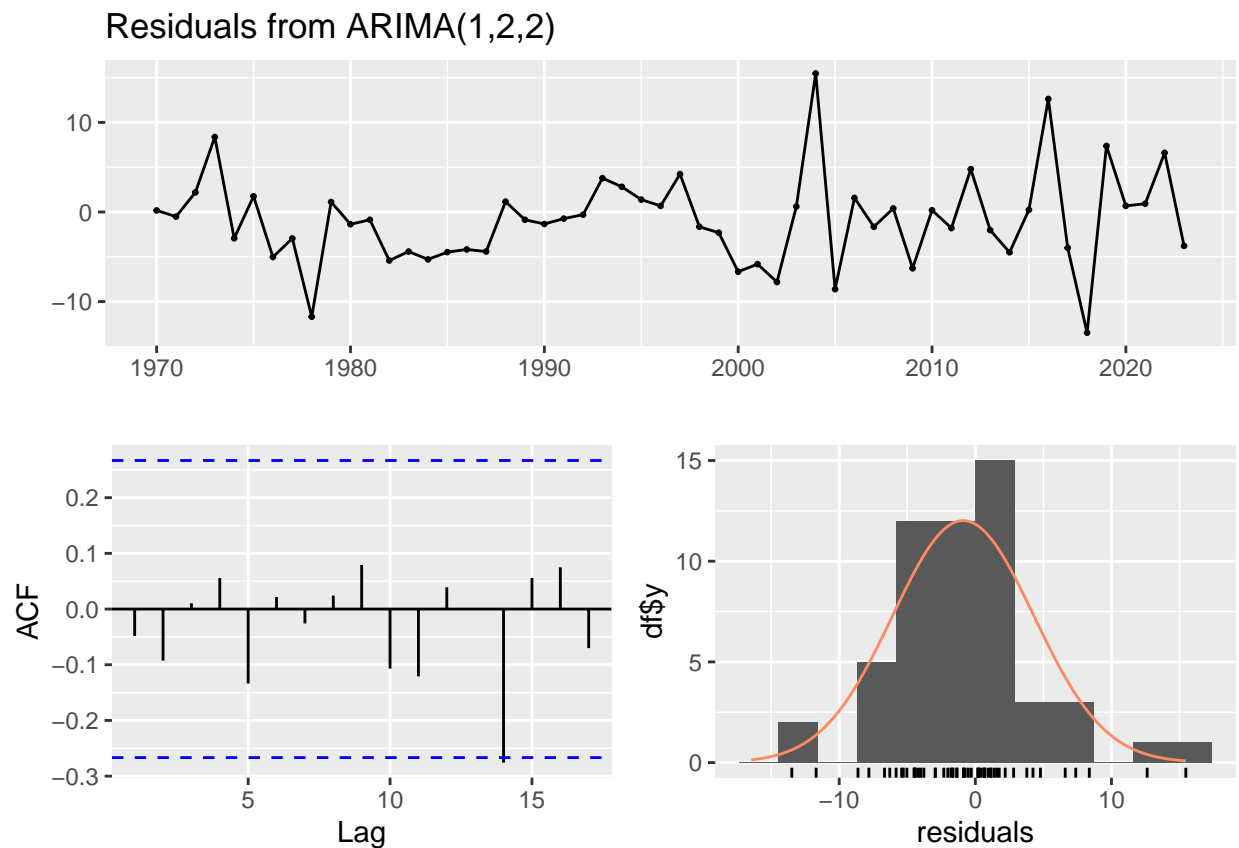
4 Estimación de parámetros:

4.1 Proceso de estimación para los modelos propuestos

5 Selección del mejor modelo:

5.1 Diagnóstico del modelo

```
checkresiduals(modelo)
```



```
##
## Ljung-Box test
##
## data: Residuals from ARIMA(1,2,2)
## Q* = 3.2441, df = 7, p-value = 0.8615
##
## Model df: 3. Total lags used: 10
```

5.2 Predicciones con el modelo seleccionado

```
n_total <- length(ts_original)
n_train <- round(n_total * 0.95)
train <- ts_original[1:n_train]
test <- ts_original[(n_train + 1):n_total]

modelo_val <- auto.arima(train, d = 2, seasonal = FALSE, stepwise = FALSE, approximation = FALSE)
pred <- forecast(modelo_val, h = length(test))

accuracy(pred, test)
```

```
##              ME      RMSE      MAE      MPE      MAPE      MASE
## Training set -0.3702967  5.396629  3.911494 -0.06154843  0.6398601  0.5153483
## Test set     13.7154454 14.528393 13.715445  1.78886057  1.7888606  1.8070416
##              ACF1
## Training set -0.08223024
## Test set     NA
```

6 Conclusiones

#Cargar librerías

```
library(tidyverse)
library(forecast)
library(tsibble)
library(lubridate)
library(ggfortify)
library(readr)

datos <- read_csv("datos_educacion_primaria.csv") %>%
  mutate(Año = yearmonth(paste(Año, "-01"))) %>%
  as_tsibble(index = Año)
```

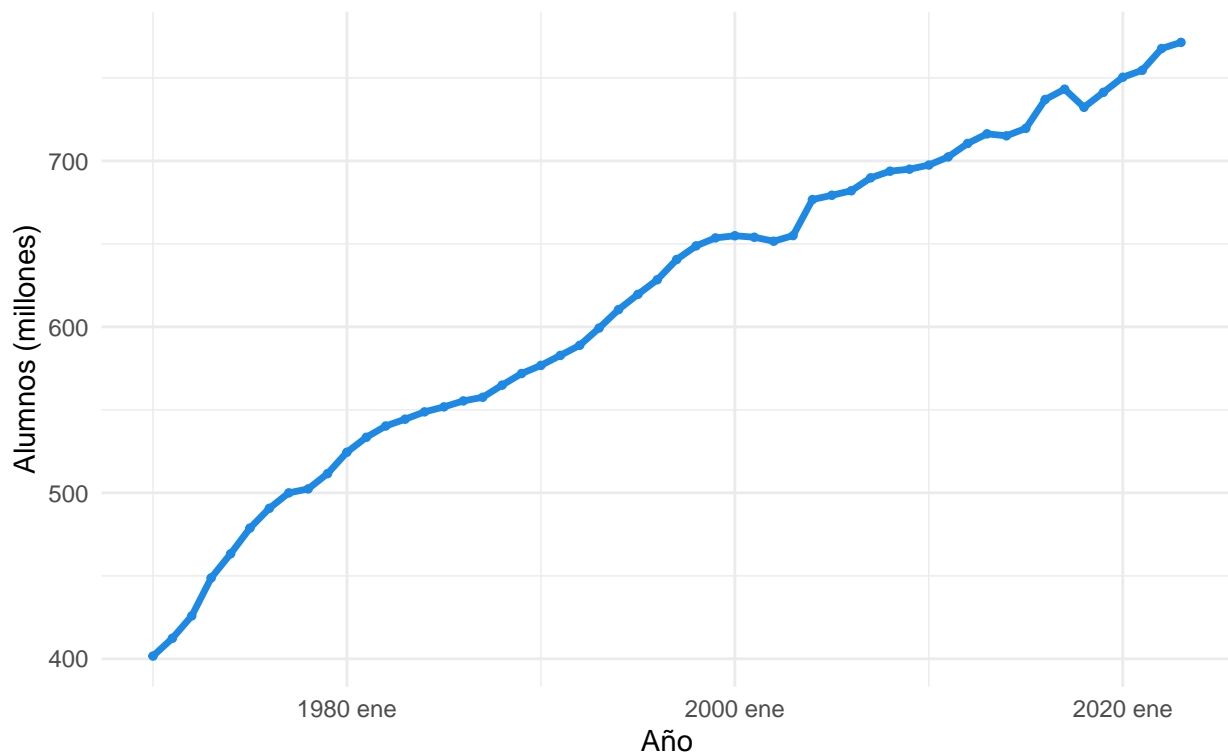


```
## Rows: 54 Columns: 2
## -- Column specification -----
## Delimiter: ","
## dbl (2): Año, Alumnos_millones
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
ts_original <- ts(datos$Alumnos_millones, start = 1970, frequency = 1)
ggplot(datos, aes(x = Año, y = Alumnos_millones)) +
  geom_line(color = "#1E88E5", linewidth = 1.2) +
  geom_point(color = "#1E88E5", size = 1) +
  labs(title = "Matrícula en Educación Primaria (1970-2023)",
       subtitle = "Serie original en millones de alumnos",
       x = "Año", y = "Alumnos (millones)") +
  theme_minimal()
```

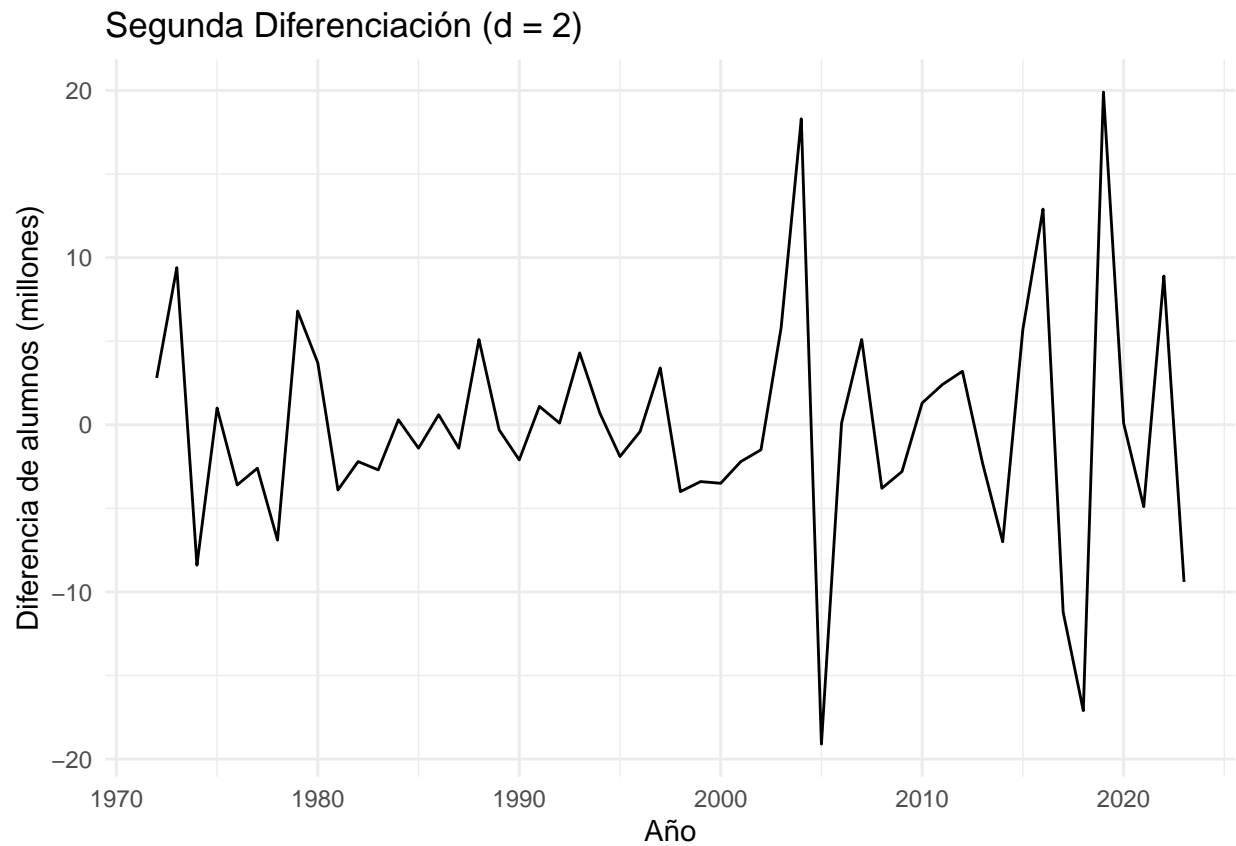
Matrícula en Educación Primaria (1970–2023)

Serie original en millones de alumnos



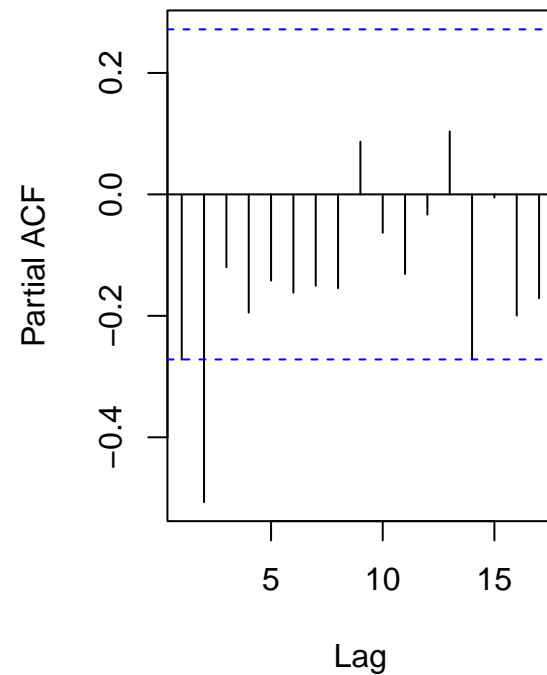
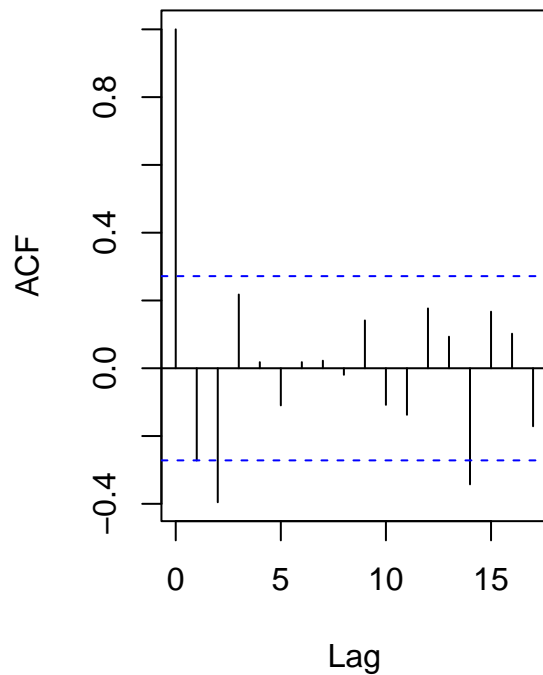
```
ts_diff1 <- diff(ts_original)
ts_diff2 <- diff(ts_diff1)
autoplot(ts_diff2) +
  labs(title = "Segunda Diferenciación (d = 2)",
```

```
x = "Año", y = "Diferencia de alumnos (millones)") +  
theme_minimal()
```



```
par(mfrow = c(1, 2))  
acf(ts_diff2, na.action = na.pass, main = "ACF - Segunda Diferenciación (d=2)")  
pacf(ts_diff2, na.action = na.pass, main = "PACF - Segunda Diferenciación (d=2)")
```

ACF – Segunda Diferenciación (d: PACF – Segunda Diferenciación (d

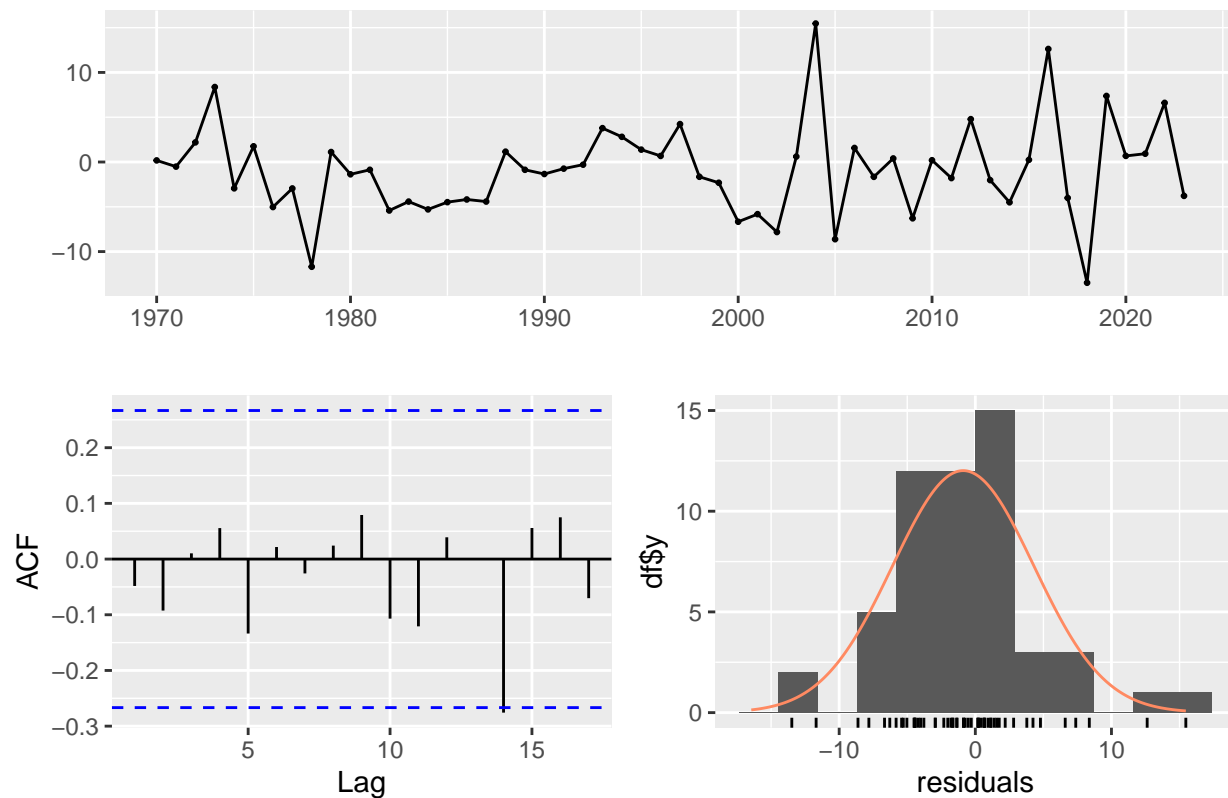


```
par(mfrow = c(1, 1))
modelo <- auto.arima(ts_original, d = 2, seasonal = FALSE, stepwise = FALSE, approximat
summary(modelo)
```

```
## Series: ts_original
## ARIMA(1,2,2)
##
## Coefficients:
##          ar1      ma1      ma2
##       -0.5720  0.0300 -0.7850
## s.e.    0.2132  0.1747  0.1323
##
## sigma^2 = 29.99: log likelihood = -161.51
## AIC=331.01  AICc=331.86  BIC=338.82
##
## Training set error measures:
##              ME      RMSE      MAE      MPE      MAPE      MASE
## Training set -0.8903757 5.216934 3.821428 -0.1575164 0.6197798 0.5057071
##              ACF1
## Training set -0.04826175
```

```
checkresiduals(modelo)
```

Residuals from ARIMA(1,2,2)



```
##  
##  Ljung-Box test  
##  
## data:  Residuals from ARIMA(1,2,2)  
## Q* = 3.2441, df = 7, p-value = 0.8615  
##  
## Model df: 3.    Total lags used: 10
```

```
n_total <- length(ts_original)  
n_train <- round(n_total * 0.95)  
train <- ts_original[1:n_train]  
test <- ts_original[(n_train + 1):n_total]  
  
modelo_val <- auto.arima(train, d = 2, seasonal = FALSE, stepwise = FALSE, approximation = TRUE)  
pred <- forecast(modelo_val, h = length(test))  
  
accuracy(pred, test)
```

##		ME	RMSE	MAE	MPE	MAPE	MASE
## Training set		-0.3702967	5.396629	3.911494	-0.06154843	0.6398601	0.5153483
## Test set		13.7154454	14.528393	13.715445	1.78886057	1.7888606	1.8070416
##		ACF1					
## Training set		-0.08223024					
## Test set		NA					

7 Conclusiones

- La serie presenta una tendencia creciente con baja variabilidad relativa.
- Se requirió aplicar una segunda diferenciación para alcanzar estacionariedad.
- El modelo ARIMA propuesto fue validado con un buen desempeño predictivo.
- Este modelo puede ser útil para proyectar la evolución de la matrícula en los próximos años.