# =============================================

# Análisis de la Matrícula en Educación Primaria (1970–2023)

# Universidad de El Salvador - Proyecto de Investigación

# =============================================

# 1. Cargar librerías —-

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

# 2. Cargar datos —-

datos <- read\_csv(“datos\_educacion\_primaria.csv”) %>% mutate(Año = yearmonth(paste(Año, “-01”))) %>% as\_tsibble(index = Año)

ts\_original <- ts(datos$Alumnos\_millones, start = 1970, frequency = 1)

# 3. Análisis gráfico de la serie —-

# Serie original

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()

# 4. Transformación para estacionariedad —-

ts\_diff1 <- diff(ts\_original) ts\_diff2 <- diff(ts\_diff1)

# Gráfico de la segunda diferenciación

autoplot(ts\_diff2) + labs(title = “Segunda Diferenciación (d = 2)”, x = “Año”, y = “Diferencia de alumnos (millones)”) + theme\_minimal()

# 5. Análisis de autocorrelación —-

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)”) par(mfrow = c(1, 1))

# 6. Modelado ARIMA —-

modelo <- auto.arima(ts\_original, d=2, seasonal=FALSE, stepwise=FALSE, approximation=FALSE) summary(modelo)

# 7. Diagnóstico de residuos —-

checkresiduals(modelo)

# 8. Validación del modelo —-

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))

# Métricas de validación

accuracy(pred, test)

# 9. Conclusiones (en comentarios) —-

# - 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.