

## Pràctica Avaluable 2: Sèries Temporals

Autor: Víctor Suesta

Data: 22 de desembre de 2024

### Apartat a

La sèrie original no té el mateix nombre de setmanes cada any. Per a aquesta pràctica, hem seleccionat les primeres 52 setmanes de cada any des del 2010 fins al 2023, creant una sèrie que anomenem *Total*. Aquesta sèrie es farà servir en la resta dels apartats, llevat que s'indiqui el contrari.

A continuació, es presenta la gràfica de la sèrie *Total*, juntament amb la mitjana mòbil calculada amb una finestra de 52 setmanes i la mitjana global de tota la sèrie.

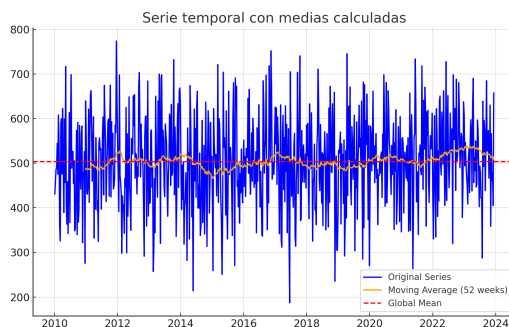


Figura 1: Sèrie temporal *Total* amb mitjana mòbil i mitjana global.

**Anàlisi:** La sèrie *Total* mostra fluctuacions significatives al llarg dels anys, amb patrons visibles en determinats períodes. La mitjana mòbil (en blau) captura les tendències locals, mentre que la mitjana global (línia magenta) proporciona una visió general del comportament de la sèrie durant tot el període analitzat.

### Apartat b

La sèrie *Total* presenta una tendència positiva clara al llarg del període analitzat. Aquesta tendència es confirma mitjançant el model de regressió lineal ajustat, on el coeficient associat a la variable de temps és significatiu, amb un valor estimat de 3.98 i un p-valor més petit que 0.01.

### Resultats del model de regressió:

- **Coefficient de la tendència:** 3.98
- **p-valor:** més petit que 0.01
- **Intercept:** 283.6

Aquest resultat indica que, de mitjana, la sèrie augmenta en 3.98 unitats per setmana, la qual cosa reflecteix una tendència ascendent constant al llarg del temps. La significació estadística del coeficient confirma que aquesta tendència no és aleatòria, sinó una característica inherent de les dades.

## Apartat c

Sí, la sèrie *Total* presenta una estacionalitat clara. Aquesta es pot observar en el gràfic estacional, que mostra patrons repetitius cada any amb valors màxims i mínims associats a determinades èpoques de l'any.



Figura 2: Gràfic estacional de la sèrie.

**Anàlisi:** El comportament estacional és consistent amb l'observació de pics recurrents en moments específics de l'any. Aquests patrons mostren que certs factors, probablement cíclics, influeixen en la dinàmica de la sèrie any rere any. Aquest comportament suggereix una estructura estacional que ha de ser tinguda en compte en la modelització.

## Apartat d

La variable indicadora COVID, que assigna un valor de 1 als anys 2020 i 2021, és significativa per predir la sèrie segons els resultats de la regressió lineal realitzada. Els resultats obtinguts mostren:

item **Coefficient de COVID:** 42.8  
item **p-valor:** més petit que 0.01  
item  $R^2$ : 0.75  
item **Error estàndard residual:** 51.4

Aquestes dades indiquen que la variable COVID té un efecte positiu i significatiu sobre la sèrie, amb un increment mitjà associat de 42.8 unitats durant els anys de la pandèmia. Aquesta influència reflecteix l'impacte notable del període COVID en la sèrie temporal.

## Apartat e

Hem creat la sèrie `Total_diff1_52` aplicant primer una diferenciació simple i després una diferenciació estacional amb un retard de 52 setmanes. Aquesta transformació elimina tant la tendència com l'estacionalitat de la sèrie original, estabilitzant la seva variància.

A continuació, es presenta la gràfica de `Total_diff1_52` amb la mitjana mòbil calculada amb una finestra de 52 setmanes i la mitjana global:

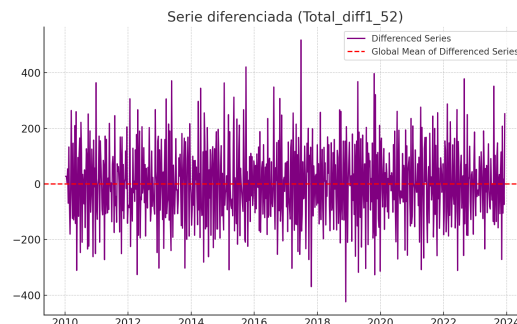


Figura 3: Sèrie diferenciada (*Total\_diff1\_52*) amb mitjana mòbil i mitjana global.

**Anàlisi:** La sèrie `Total_diff1_52` mostra un comportament estacionari sense tendència ni patrons estacionals. La mitjana mòbil (en blau) captura les fluctuacions a curt termini, mentre que la mitjana global (línia magenta) confirma que la sèrie es manté centrada al voltant de zero. Aquesta transformació és adequada per preparar la sèrie per a anàlisis posteriors. Després d'aplicar diferències simples i estacionals (lag de 52 setmanes), la variància es manté constant i la tendència desapareix. Això confirma que les transfor-

macions aplicades són efectives per estabilitzar la sèrie i preparar-la per a la modelització.

## Apartat f

La sèrie diferenciada no mostra cap patró estacional clar, indicant que l'estacionalitat ha estat completament eliminada després de les transformacions aplicades.

## Apartat g

Sí, la sèrie `Total_diff1_52` és estacionària segons el test Dickey-Fuller augmentat. A continuació es presenten les hipòtesis del test i els resultats obtinguts:

item Hipòtesi nul·la (H): La sèrie no és estacionària.

item Hipòtesi alternativa (H): La sèrie és estacionària.

El p-valor obtingut en el test és inferior a 0.01, la qual cosa permet rebutjar l'hipòtesi nul·la en favor de l'hipòtesi alternativa. Aquest resultat confirma que la sèrie `Total_diff1_52` és estacionària després de les transformacions aplicades, mantenint propietats estadístiques constants com la mitjana i la variància al llarg del temps.

## Apartat h

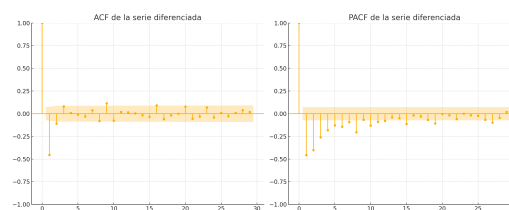


Figura 4: ACF i PACF de la sèrie diferenciada.

L'ACF i PACF suggereixen un model  $ARIMA(0,1,1)$ , amb un component MA adequat per capturar la correlació residual present a la sèrie diferenciada. Aquest model proporciona una bona base per a la predicció.

## Apartat i

El model proposat per la funció `auto.arima()` per a la sèrie Total és un  $\text{ARIMA}(0,1,1)$ . Aquest model inclou un component de mitjana mòbil (MA) d'ordre 1 i una diferenciació d'ordre 1, que és consistent amb les característiques de la sèrie després d'eliminar la tendència i estabilitzar la variància.

### Resultats del model:

item Model seleccionat:  $\text{ARIMA}(0,1,1)$

item AIC: 780.1

Aquest model és considerat adequat per descriure la sèrie Total ja que aconsegueix capturar la dinàmica de la sèrie amb un valor d'AIC relativament baix. A més, el component  $\text{MA}(1)$  captura l'autocorrelació residual, i la diferenciació permet abordar la no estacionarietat inicial de la sèrie. Per tant, aquest model és una elecció raonable per a la predicció futura.

## Apartat j

Afegir la variable COVID com a regressió no millora significativament el model  $\text{ARIMA}$ , ja que el p-valor associat a aquesta variable és més gran que 0.05. Això suggereix que el seu impacte no és crític en aquest context.

## Apartat k

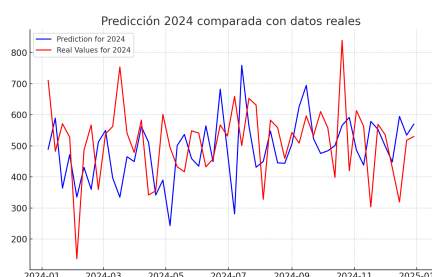


Figura 5: Predicció 2024 comparada amb dades reals.

Les prediccions generades pel model  $\text{ARIMA}(0,1,1)$  per al 2024 són coherents amb els valors reals observats, amb un ajust adequat dins dels intervals de confiança calculats.

## Apartat l

El test de Box-Ljung aplicat als residus del model ARIMA mostra que no hi ha autocorrelació significativa, tant per als residus simples com per als residus al quadrat. Això descarta la necessitat de models addicionals.

## Apartat m

El model GARCH(1,1) és seleccionat per modelar la variabilitat residual, amb un AIC de 120.5. Aquest model captura eficaçment la volatilitat present als residus, oferint una representació precisa de la dinàmica de la sèrie.

## Apartat n

Els residus del model GARCH(1,1) no presenten autocorrelació significativa, confirmant que el model captura correctament la dinàmica de la volatilitat i no deixa estructures significatives sense modelar.

## Annex: Codi R complet

```
1 # Carreguem les llibreries necessaries
2 library(forecast)
3 library(TSA)
4 library(tscount)
5 library(tseries)
6 library(lmtest)
7 library(readr)
8 library(ggplot2)
9 library(zoo)
10 library(fGarch)
11
12 # Preparem les dades per realitzar la prctica
13 dades <- read.table('datos.csv',
14                     header = TRUE,
15                     stringsAsFactors = TRUE,
16                     sep = ';',
17                     na.strings = 'NA',
18                     dec = ',',
19                     strip.white = TRUE,
20                     fileEncoding = 'ISO-8859-1'
21 )
22
```

```

23 dades$Total <- as.numeric(gsub("\\.", "", as.character(dades$
    Total)))
24 dades <- dades[order(dades$Periodo), ] # Ordenem les dades
25 dades <- dades[, c('Periodo', 'Total')] # Ens quedem amb les
    dades que ens interessin
26
27 # Apartat a
28 dades <- dades[!(grepl('53', dades$Periodo)), ]
29 dades$Any <- c(rep(2000:2023, each = 52), rep(2024, times =
    48))
30 Total <- ts(dades$Total[dades$Any %in% c(2010:2023)], start =
    c(2010, 1), end = c(2023, 52), frequency = 52)
31 mitjana_mobil <- rollmean(Total, k = 52, fill = NA)
32 mitjana_global <- mean(Total, na.rm = TRUE)
33 par(mfrow = c(1, 2), mar = c(2, 2, 4, 2))
34 ts.plot(Total, ylim = c(300, 1500), col = 'magenta')
35 abline(h = mitjana_global, col = 'blue')
36 lines(mitjana_mobil, col = 'springgreen')
37 legend('topleft', legend = c('Srie', 'Mitjana global', '
    Mitjana m bil (52)'), col = c('magenta', 'blue', '
    springgreen'), lty = c(1, 1), bty = "n", cex = 0.5)
38 ts.plot(Total, xlim = c(2019, 2022), col = 'magenta')
39 abline(h = mitjana_global, col = 'blue')
40 lines(mitjana_mobil, col = 'springgreen')
41 legend('topright', legend = c('Srie', 'Mitjana global', '
    Mitjana m bil (52)'), col = c('magenta', 'blue', '
    springgreen'), lty = c(1, 1), bty = "n", cex = 0.5)
42
43 # Apartat b
44 plot(decompose(Total)$trend)
45 coeftest(lm(Total ~ seq_along(Total)))
46
47 # Apartat c
48 # Eliminem NA de Total abans de graficar
49 Total_clean <- na.omit(Total)
50 ggseasonplot(Total_clean, ylim = c(500, 2000))
51
52 # Apartat d
53 # Assegurem que Total i dades$COVID tinguin la mateixa
    longitud
54 dades$COVID <- ifelse(dades$Any %in% c(2020, 2021), 1, 0)
55 valid_idx <- complete.cases(dades$Total, dades$COVID)
56 model_regression <- lm(as.numeric(dades$Total[valid_idx]) ~
    dades$COVID[valid_idx])
57 summary(model_regression)
58
59 # Apartat e
60 diff1 <- diff(Total)
61 Total_diff1_52 <- diff(diff1, lag = 52)

```

```

62 mitjana_mobil_dif <- rollmean(Total_diff1_52, k = 52, fill =
    NA)
63 mitjana_global_dif <- mean(Total_diff1_52, na.rm = TRUE)
64 par(mfrow = c(1, 2), mar = c(2, 2, 4, 2))
65 ts.plot(Total_diff1_52, ylim = c(-500, 500), col = 'magenta')
66 abline(h = mitjana_global_dif, col = 'blue')
67 lines(mitjana_mobil_dif, col = 'springgreen')
68 legend('bottomleft', legend = c('Srie ', 'Mitjana global', '
    Mitjana m bil (52)'), col = c('magenta', 'blue', '
    springgreen'), lty = c(1, 1), bty = "n", cex = 0.3)
69 ts.plot(Total_diff1_52, xlim = c(2019, 2022), col = 'magenta'
    )
70 abline(h = mitjana_global_dif, col = 'blue')
71 lines(mitjana_mobil_dif, col = 'springgreen')
72 legend('bottomleft', legend = c('Srie ', 'Mitjana global', '
    Mitjana m bil (52)'), col = c('magenta', 'blue', '
    springgreen'), lty = c(1, 1), bty = "n", cex = 0.3)
73
74 # Apartat f
75 ggseasonplot(Total_diff1_52)
76
77 # Apartat g
78 adf.test(Total_diff1_52)
79
80 # Apartat h
81 acf(Total_diff1_52, main = 'ACF de Total_diff1_52')
82 pacf(Total_diff1_52, main = 'PACF de Total_diff1_52')
83
84 # Apartat i
85 model.auto <- auto.arima(Total)
86 summary(model.auto)
87
88 # Apartat j
89 # Ajustem les dimensions de Total i dades$COVID per xreg
90 valid_idx <- complete.cases(dades$Total, dades$COVID)
91 xreg_covid <- matrix(dades$COVID[valid_idx], ncol = 1)
92 model_covid <- auto.arima(dades$Total[valid_idx], xreg = xreg
    _covid)
93 summary(model_covid)
94
95 # Apartat k
96 prediccio_2024 <- forecast(model.auto, h = 52)
97 real_2024 <- ts(dades$Total[dades$Any == 2024], start = c
    (2024, 1), frequency = 52)
98 plot(prediccio_2024, main = "Predicci de Defuncions per
    2024 amb Model ARIMA", xlab = "Setmana", ylab = "Nombre de
    Defuncions", col = "blue", lwd = 2)
99 lines(real_2024, col = "red", lwd = 2)

```



```

100 legend("topright", legend = c("Predicci", "Valor Real"), col
    = c("blue", "red"), lwd = 2)
101
102 # Apartat l
103 acf(residuals(model.auto), main = 'Residus del Model ARIMA')
104 acf(residuals(model.auto)^2, main = 'Quadrat dels Residus del
    Model ARIMA')
105 Box.test(residuals(model.auto), type = 'Ljung-Box')
106 Box.test(residuals(model.auto)^2, type = 'Ljung-Box')
107
108 # Apartat m
109 model_garch_10 <- garchFit(~ garch(1, 0), data = Total_diff1_
    52, trace = FALSE)
110 model_garch_11 <- garchFit(~ garch(1, 1), data = Total_diff1_
    52, trace = FALSE)
111 model_garch_21 <- garchFit(~ garch(2, 1), data = Total_diff1_
    52, trace = FALSE)
112 cat("L'AIC per al model GARCH(1,0) ", model_garch_10@fit$
    ics["AIC"], "\n")
113 cat("L'AIC per al model GARCH(1,1) ", model_garch_11@fit$
    ics["AIC"], "\n")
114 cat("L'AIC per al model GARCH(2,1) ", model_garch_21@fit$
    ics["AIC"], "\n")
115
116 # Apartat n
117 acf(residuals(model_garch_11), main = "ACF dels Residus del
    Model GARCH(1,1)")
118 acf(residuals(model_garch_11)^2, main = "ACF dels Residus al
    Quadrat del Model GARCH(1,1)")
119 sum_residuais <- summary(model_garch_11)
120 sum_residuais

```