

Analiza Seriilor Temporale

Buzilă Tatiana - gr. SD-231
Marusic Galina - conf. univ. dr.





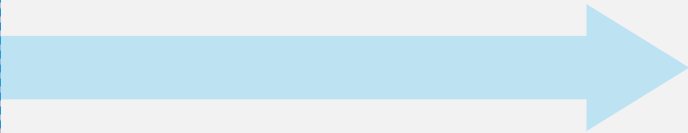
Introducere

Ați încercat vreodată să **preziceți** cum va fi **vremea** mâine doar uitându-vă la zilele anterioare? Sau poate v-ați întrebat de ce **prețurile produselor** fluctuează de la o lună la alta? **Seriile temporale** sunt peste tot: în **economie**, **meteorologie**, **marketing**, **transport** și **sănătate**. Analizându-le, putem înțelege tiparele și face **predicții precise**.



Serii temporale

O **serie temporală** este un set de date colectate și înregistrate în **ordine cronologică**, la **intervale de timp egale** (zilnic, lunar, anual etc.), cu scopul de a analiza **evoluția unui fenomen** în timp. Aceste date permit identificarea unor tipare, cum ar fi **tendința**, **sezonalitatea** și **fluctuațiile ciclice**.



Unde Se Folosește Analiza Seriilor Temporale?

Economie și Finanțe

Prognoza prețurilor acțiunilor și riscurilor financiare.

Meteorologie

Previzionarea vremii și a fenomenelor extreme.

Sănătate

Monitorizarea bolilor și prognoza epidemilor.

Marketing

Analiza vânzărilor și comportamentului clienților.

Inginerie

Prezicerea defectării echipamentelor și întreținerea preventivă.

Energie

Prognoza consumului de energie și gestionarea rețelelor electrice.

Componentele unei Serii Temporale

01

Tendința (Trend) - Schimbări pe termen lung ale valorilor, cum ar fi creșterea sau scăderea generală a datelor.



02

Sezonalitatea (Seasonality) – Repetarea anumitor tipare la intervale regulate (zilnic, lunar, anual).



03

Ciclurile (Cyclicity) – Fluctuații pe termen lung, dar fără o regularitate precisă, cum ar fi perioadele de expansiune și recesiune economică.

04

Reziduurile (Noise) – Variabilitate aleatorie care nu poate fi explicată de celelalte componente, de obicei din cauza evenimentelor imprevizibile.

Exemplu: Numărul de Pasageri Zilnici pe Traseul X

În fiecare zi, la aceeași oră, se colectează date despre numărul de pasageri care călătoresc pe un anumit traseu de autobuz. Aceasta înseamnă că, de exemplu, la **ora 8:00** dimineața se numără pasagerii care urcă în autobuz pe traseul X.

Aceste date constituie o **serie temporală**, deoarece sunt colectate la intervale de timp regulate (**în acest caz, zilnic la aceeași oră**).

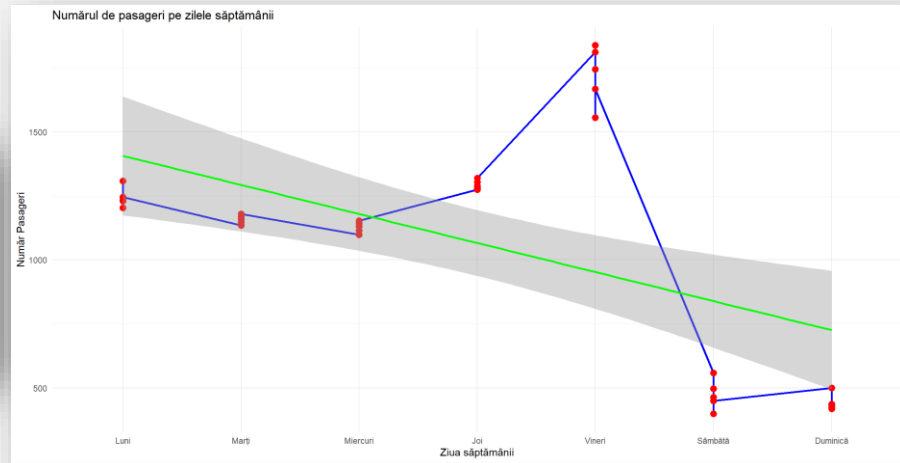
Ziua	Luni	Marți	Miercuri	Joi	Vineri	Sâmbătă	Duminică
Săptămâna 1	1308	1135	1098	1274	1812	559	499
Săptămâna 2	1230	1147	1115	1281	1838	463	419
Săptămâna 3	1203	1159	1130	1290	1745	497	425
Săptămâna 4	1234	1168	1142	1304	1556	399	431
Săptămâna 5	1245	1179	1153	1320	1667	450	437

Analiză exemplu: Identificarea Tendinței

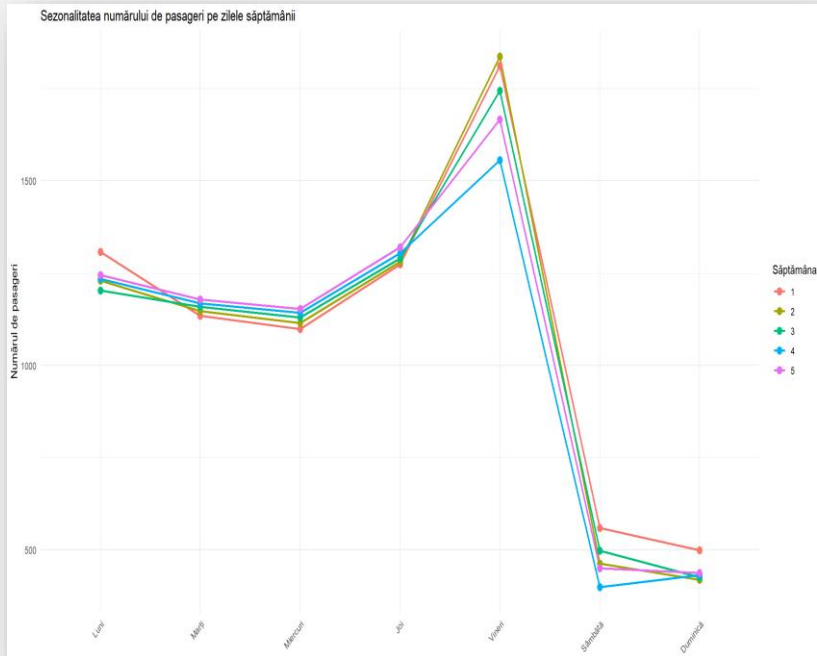
Zilele lucrătoare (Luni-Vineri) prezintă un număr de pasageri mai mare, ceea ce indică o cerere crescută pentru transportul public datorită activităților zilnice de muncă sau școală.

Zilele de weekend (Sâmbătă-Duminică) un număr semnificativ mai mic de pasageri, ceea ce sugerează că, în general, transportul public este utilizat mai puțin în acest interval, deoarece mulți oameni nu depind de acesta pentru activitățile lor din weekend.

```
> ggplot(date_pasageri, aes(x = Zi, y = Număr_Pasageri, group = 1)) +
+   geom_line(color = "blue", size = 1) +
+   geom_point(color = "red", size = 3) +
+   geom_smooth(method = "lm", color = "green", size = 1) +
# Linia de trend
+   labs(title = "Numărul de pasageri pe zilele săptămânii",
+         x = "Ziua săptămânii",
+         y = "Număr Pasageri") +
+   theme_minimal() +
+   scale_x_discrete(limits = c("Luni", "Marți", "Miercuri",
+ "Joi", "Vineri", "Sâmbătă", "Duminică"))
```



Analiză exemplu: Identificarea Sezonalității



```
> install.packages("ggplot2") # Dacă nu este deja instalat
library(ggplot2)

data <- data.frame(
  Saptamana = rep(1:5, each = 7), # Etichetăm fiecare săptămână
  Ziua = rep(c("Luni", "Marți", "Miercuri", "Joi", "Vineri", "Sâmbătă", "Duminică"),
    times = 5),
  Pasageri = c(1308, 1135, 1098, 1274, 1812, 559, 499,
    1230, 1147, 1115, 1281, 1838, 463, 419,
    1203, 1159, 1130, 1290, 1745, 497, 425,
    1234, 1168, 1142, 1304, 1556, 399, 431,
    1245, 1179, 1153, 1320, 1667, 450, 437)
)

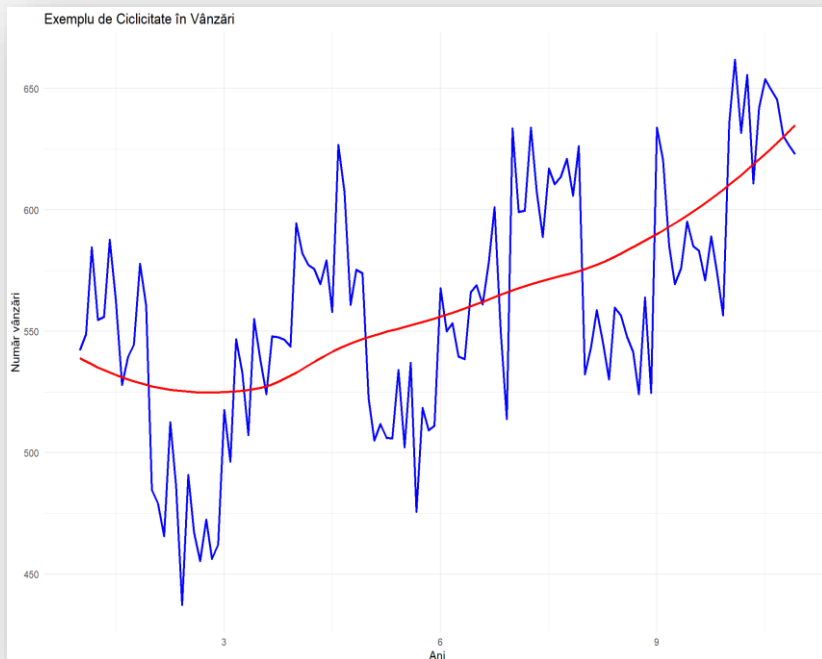
data$Ziua <- factor(data$Ziua, levels = c("Luni", "Marți", "Miercuri", "Joi", "Vineri",
  "Sâmbătă", "Duminică"))

ggplot(data, aes(x = Ziua, y = Pasageri, group = Saptamana, color = as.factor(Saptamana)
)) +
  geom_line(size = 1) +
  geom_point(size = 3) +
  theme_minimal() +
  labs(title = "Sezonalitatea numărului de pasageri pe zilele săptămânii",
    x = "Ziua săptămânii",
    y = "Numărul de pasageri",
    color = "Săptămâna") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Vineri se înregistrează cel mai mare număr de pasageri, indicând o cerere mai mare pentru transportul public înainte de weekend, în timp ce alte zile ale săptămânii au valori mai mici.

Deci, faptul că vineri se înregistrează mereu un număr mai mare de pasageri indică **sezonaltate săptămânală**.

Analiză: Exemplu de ciclicitate în vânzări



```

> library(ggplot2)
> set.seed(123)
> ani <- rep(1:10, each=12)
> luni <- rep(1:12, times=10)
>
> ciclu <- sin(2 * pi * ani / 3) * 50 # variație la fiecare 3 ani
>
> trend <- ani * 10
> zgomot <- rnorm(120, mean=0, sd=20)
> vanzari <- 500 + trend + ciclu + zgomot
>
> date <- data.frame(An=ani, Luna=luni, Vanzari=vanzari)
>
> ggplot(date, aes(x=An + (Luna-1)/12, y=Vanzari)) +
+   geom_line(color="blue", size=1) +
+   geom_smooth(method="loess", color="red", se=FALSE) + # Linie
+   labs(title="Exemplu de Ciclicitate în Vânzări",
+         x="Ani", y="Număr vânzări") +
+   theme_minimal()

```

Linia albastră = vânzările lunare fluctuante
Linia roșie = tendința generală pe termen lung
Modelul ciclic la fiecare 3 ani este vizibil sub forma unor creșteri și scăderi neregulate.

Modele de Analiză a Seriilor de Timp

<u>ARIMA</u>	Folosit pentru prognoze atunci când datele au tendințe, dar nu au schimbări sezoniere.
<u>SARIMA</u>	O variantă a ARIMA, folosită atunci când datele au atât tendințe, cât și schimbări sezoniere.
<u>SARIMAX</u>	Similar cu SARIMA, dar ia în considerare și factori externi care influențează datele.
<u>Holt-Winters</u>	Folosește netezirea pentru a prezice datele, ținând cont de tendință și sezonalitate.
<u>VAR</u>	Analizează cum se influențează mai multe seturi de date între ele.
<u>GARCH</u>	Analizează fluctuațiile rapide ale datelor, folosit mai ales în piețele financiare

ARIMA(p, d, q)

```

> # Incărcarea pachetelor necesare
library(ggplot2)
library(forecast)

# 1 Incărcarea setului de date (modifică path-ul cu locația exactă a fișierului)
data <- read.csv("C:\\Users\\tatia\\Desktop\\Anul 2\\Semestru 2\\Analiza sistemelor și vizualizarea datelor\\archive (6)\\champagne.csv")

# 2 Transformarea în serie temporală
ts_data <- ts(data$Sales, start = c(1964, 1), frequency = 12)

# 3 Descompunerea seriei temporale
decomposed <- decompose(ts_data)

# 4 Grafic pentru tendință
trend_plot <- ggplot(data = data.frame(Time = time(decomposed$trend), Trend = decomposed$trend),
  aes(x = Time, y = Trend)) +
  geom_line(color = "blue") +
  ggtitle("Tendința vânzărilor de șampanie") +
  theme_minimal()

# 5 Grafic pentru sezonabilitate
seasonal_plot <- ggplot(data = data.frame(Time = time(decomposed$seasonal), Seasonal = decomposed$seasonal),
  aes(x = Time, y = Seasonal)) +
  geom_line(color = "green") +
  ggtitle("Sezonabilitatea vânzărilor de șampanie") +
  theme_minimal()

# 6 Grafic pentru reziduuri
residuals_plot <- ggplot(data = data.frame(Time = time(decomposed$random), Residuals = decomposed$random),
  aes(x = Time, y = Residuals)) +
  geom_line(color = "red") +
  ggtitle("Reziduurile seriei temporale") +
  theme_minimal()

# 7 Modelarea și Predicția cu ARIMA
arima_model <- auto.arima(ts_data)
forecast_values <- forecast(arima_model, h = 12)

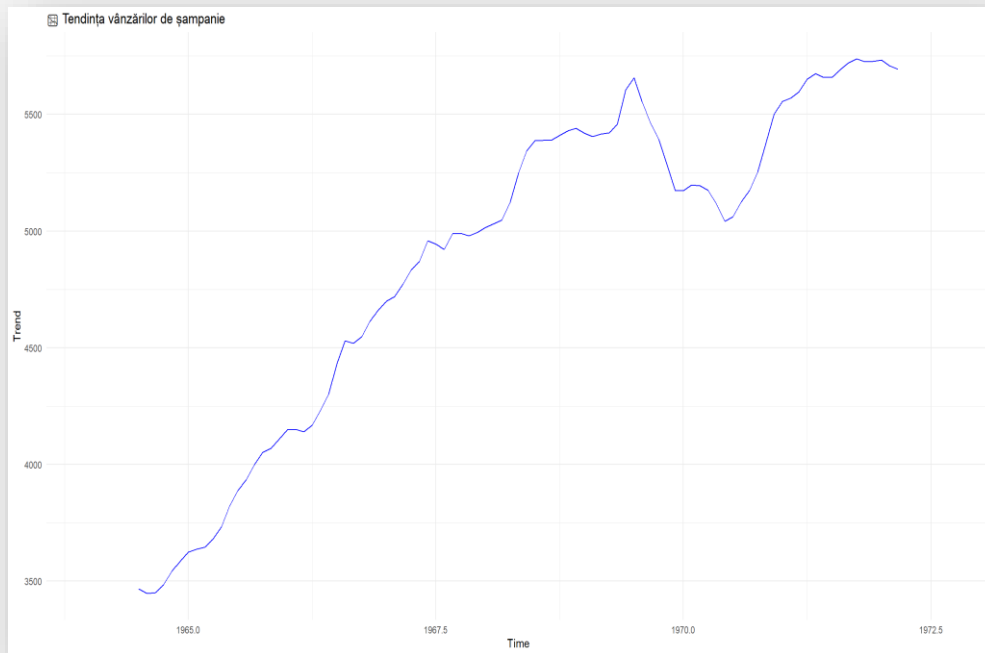
# 8 Graficul predicției
forecast_plot <- autoplot(forecast_values) +
  ggtitle("Predicția vânzărilor cu ARIMA") +
  theme_minimal()

# 9 Afișarea graficelor separat
print(trend_plot)
print(seasonal_plot)
print(residuals_plot)
print(forecast_plot)

```

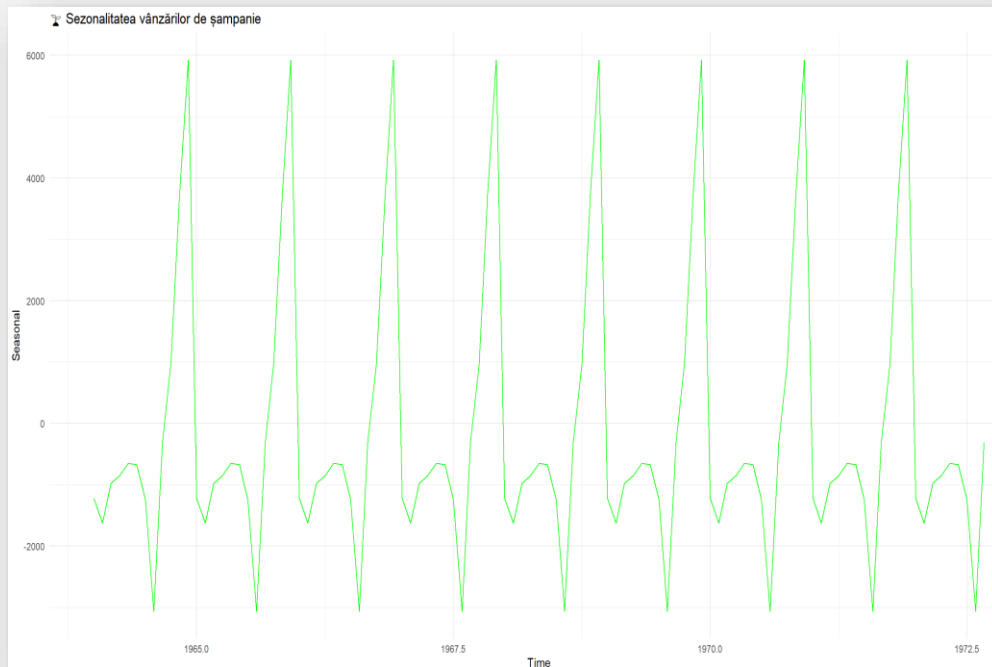
Tendința vânzărilor de șampanie

Graficul arată o **creștere generală** a vânzărilor de șampanie, cu unele **fluctuații**. Creșterea poate fi influențată de **factori economici** și **perioade festive**, iar scăderile temporare pot avea **cauze externe**. Per ansamblu, **tendința este pozitivă**.

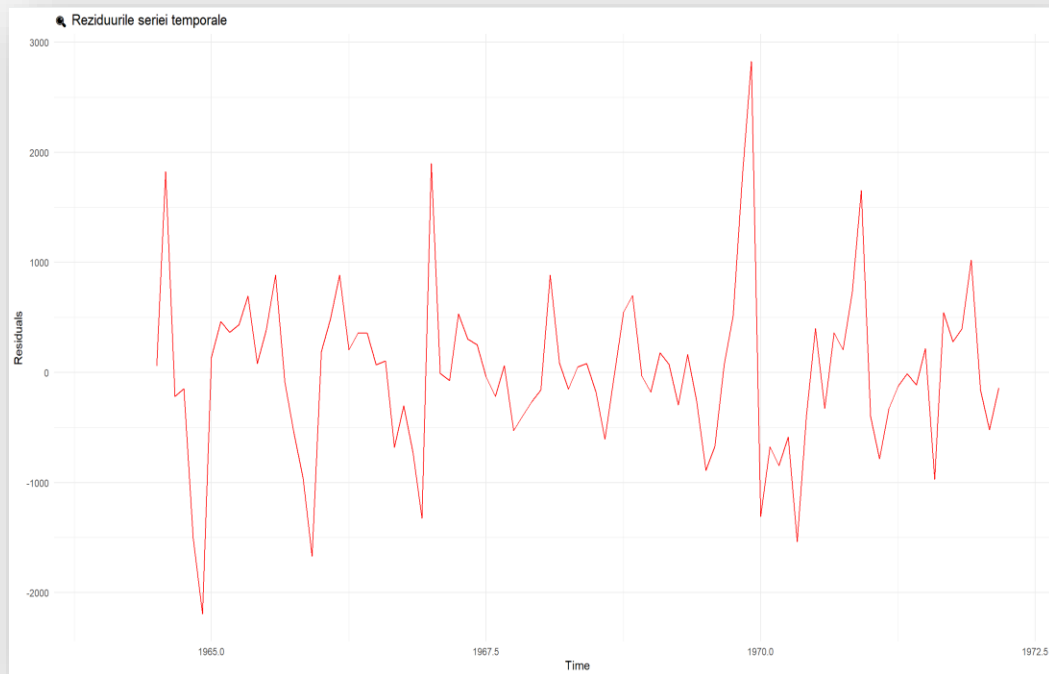


Sezionalitatea vânzărilor de șampanie

Graficul arată un model clar de **sezonalitate** în vânzările de șampanie, cu **vârfuri** distincte care apar anual. Aceste creșteri semnificative indică o **cerere mai mare** în anumite perioade ale anului, cel mai probabil în preajma sărbătorilor. După fiecare vârf, vânzările scad, reflectând un **ciclu repetitiv** de consum.



Reziduurile seriei temporale

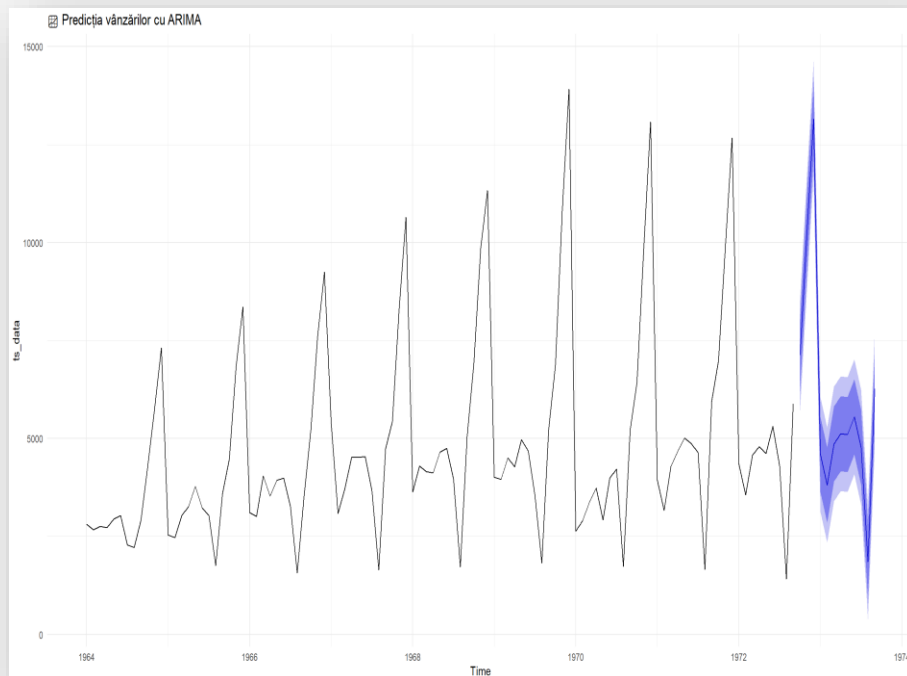


Reziduurile variază aleatoriu în jurul valorii zero, ceea ce indică faptul că modelul surprinde bine tendința și sezonalitatea. Totuși, **există unele fluctuații mari**, semn că pot exista factori neașteptați care influențează vânzările.

Predicția vânzărilor folosind modelul ARIMA

$$y_t = \mu + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j \epsilon_{t-j} + \epsilon_t$$

- **Linia neagră subțire** → vânzările estimate.
- **Benzile albastre** → intervale de încredere:
 - **Banda închisă (80%)** → predicție mai sigură.
 - **Banda deschisă (95%)** → incertitudine mai mare.



SARIMAX

Formula matematică SARIMAX

Modelul SARIMAX este definit de următoarea ecuație:

$$\Phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^DY_t = c + \Theta_q(B)\Theta_Q(B^s)\epsilon_t + \beta X_t$$

unde:

- Y_t = seria de timp de modelat
- X_t = variabila exogenă (PIB, de exemplu)
- B = operatorul de decalare (lag), astfel încât $BY_t = Y_{t-1}$
- $(1-B)^d$ = diferențierea pentru a face seria staționară
- $(1-B^s)^D$ = diferențierea sezonieră
- $\Phi_p(B)$ = termenul autoregresiv (AR)
- $\Theta_q(B)$ = termenul medie mobilă (MA)
- $\Phi_P(B^s)$ = termenul autoregresiv sezonier (SAR)
- $\Theta_Q(B^s)$ = termenul medie mobilă sezonieră (SMA)
- ϵ_t = eroarea aleatorie (white noise)
- β = coeficientul variabilei exogene

```
# Încărcarea pachetelor necesare
library(ggplot2)
library(forecast)
library(tseries)
library(TSA)

# 1 Încărcarea datelor
data <- economics

# 2 Crearea unei serii temporale pentru șomaj (unemploy)
ts_data <- ts(data$unemploy, start = c(1967, 7), frequency = 12)

# 3 Adăugarea PIB-ului ca variabilă exogenă (exemplu cu valori simulate)
set.seed(123)
data$gdp <- rnorm(nrow(data), mean = 15000, sd = 2000)

# 4 Descompunerea seriei temporale
decomposed <- decompose(ts_data)

# 5 Grafic pentru componentele seriei
plot(decomposed)

# 6 Crearea modelului SARIMAX cu PIB ca variabilă exogenă
train_size <- round(0.8 * length(ts_data))
train_data <- ts_data[1:train_size]
test_data <- ts_data[(train_size + 1):length(ts_data)]

# Variabila exogenă (PIB)
train_gdp <- data$gdp[1:train_size]
test_gdp <- data$gdp[(train_size + 1):length(ts_data)]

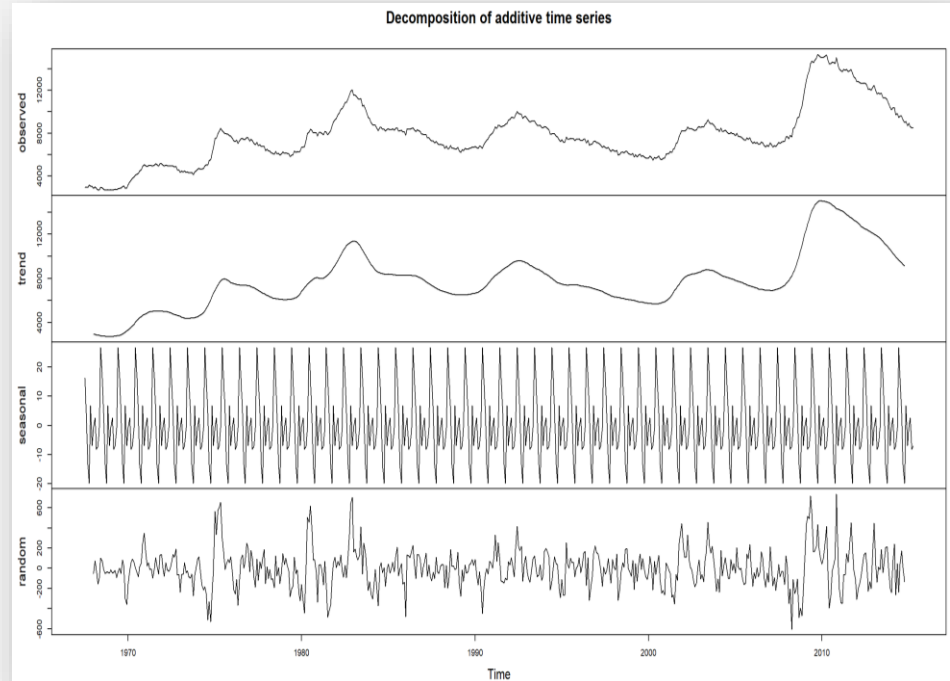
# Modelarea cu ARIMA și variabilă exogenă
sarimax_model <- auto.arima(train_data, xreg = train_gdp, seasonal = TRUE)

# Predicția pe setul de testare
forecast_values <- forecast(sarimax_model, xreg = test_gdp, h = length(test_data))

# 7 Graficul predicției
autoplot(forecast_values) +
  ggtitle("Predicția ratei șomajului cu SARIMAX") +
  theme_minimal()
```


Descompunerea seriei temporale

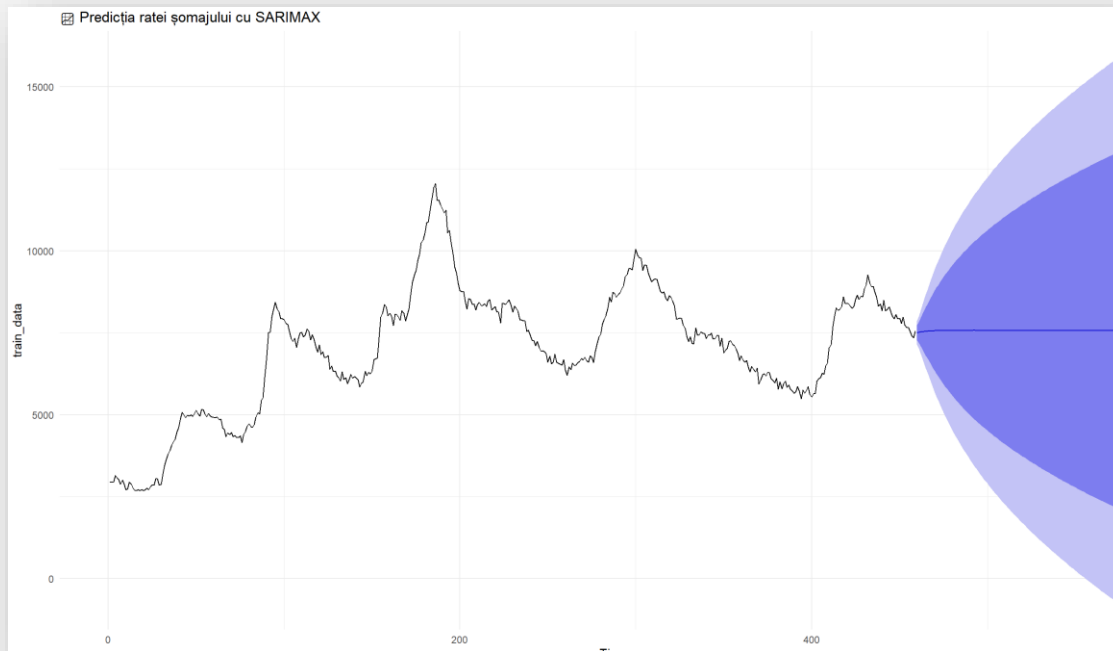
- **Tendința** arată evoluția generală a numărului de șomeri, evidențiind perioade de creștere și scădere.
- **Sezonalitatea** reflectă fluctuațiile regulate ale șomajului, influențate de factori economici recurenți.
- **Reziduurile** includ variațiile neexplicate, cauzate de evenimente economice imprevizibile.



SARIMAX

Aceste intervale reflectă incertitudinea prognozelor, cu intervalul de **95%** având o probabilitate **mai mare** de a include valoarea reală a șomajului.

Prognoza șomajului utilizând modelul SARIMAX cu PIB ca variabilă externă



Concluzii



Analiza seriilor de timp este o metodă esențială pentru înțelegerea și prognozarea evoluției unui fenomen în timp. Identificarea tendințelor, sezonality și fluctuațiilor ajută la luarea unor decizii informate și la optimizarea resurselor. Prin utilizarea unor modele avansate, se pot îmbunătăți prognozele, luând în considerare și factorii externi care influențează datele. Această abordare este aplicabilă într-o varietate de domenii, de la economie și transport, până la sănătate și tehnologie, contribuind la o gestionare mai eficientă a proceselor și la creșterea competitivității.

