المندوبية السامية للتخطيط

ⵜⴰⵏⵎⵓⵍⵜ ⵜⴰⵏⵎⵇⵇⵔⴰⵏⵜ ⵏ ⵓⵙⵙⵓⴷⵙ

HAUT-COMMISSARIAT AU PLAN

INSEA

INSEA
Institut National des Statistiques et d'Economie Appliquée

Projet : Séries Chronologiques

Réalisé par : Lagzouli Riyad DS

Mohamed Nabil Ouaja DS

Description des Données :

La série chronologique représente le nombre trimestriel de faillites d'entreprises à Paris sur une période de 27 ans, allant du premier trimestre de 1995 (1995-T1) au quatrième trimestre de 2021 (2021-T4).

Caractéristiques principales :

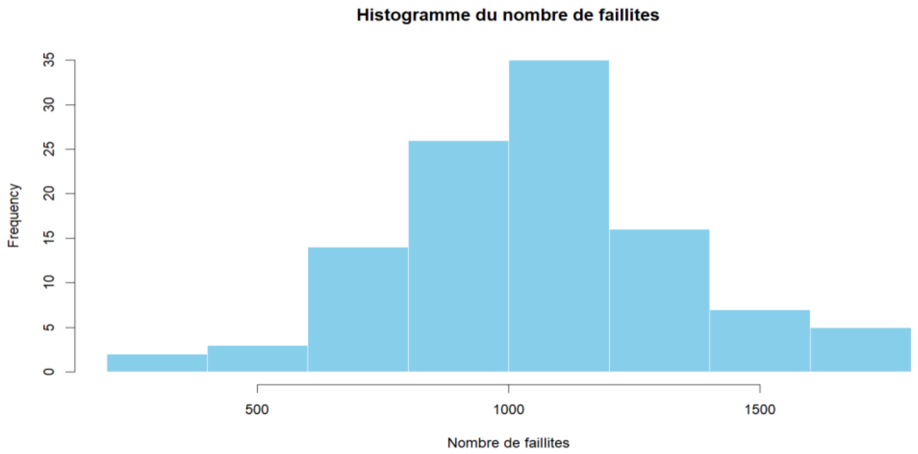
Fréquence : Données trimestrielles (4 observations par an).

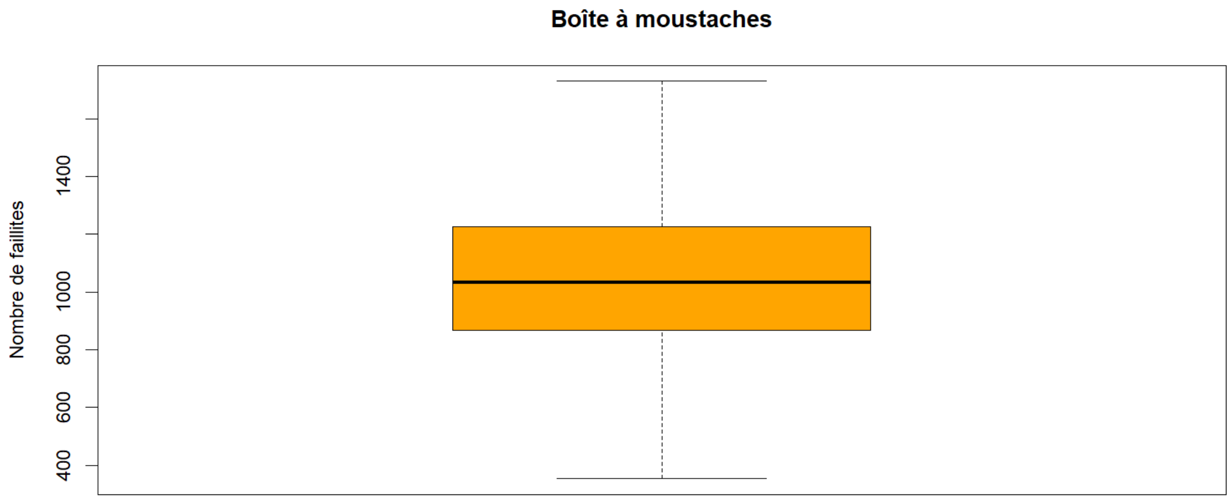
Nombre total d'observations : 108 (27 ans × 4 trimestres).

Unité de mesure : Nombre d'entreprises en faillite par trimestre.

Annee	T1	T2	T3	T4
1995	1731	1338	1190	1643
1996	1690	1542	1230	1598
1997	1696	1670	1294	1540
1998	1435	1168	1007	1385
1999	1451	1251	894	1333
2000	1222	886	814	1176
2001	635	1164	828	1439
2002	1161	1326	935	1318
2003	1499	1099	907	1367
2004	1301	1168	924	1274
2005	1357	1235	855	1189
2006	964	900	699	1063
2007	1195	1039	743	1153
2008	1050	1035	767	1347
2009	1136	1136	955	1243
2010	1162	1033	836	1006
2011	1039	884	761	1009
2012	946	1001	738	1041
2013	1078	1028	819	1188
2014	1106	918	861	992
2015	1138	1019	731	1011
2016	1087	1119	668	1087
2017	998	883	711	959
2018	880	841	634	1016
2019	869	868	640	907
2020	700	354	565	708
2021	618	491	393	563

```
> summary(data$Nombre_de_faillites)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  354.0   868.8  1034.0  1050.9  1224.0  1731.0 
> cat("Écart-type :", sd(data$Nombre_de_faillites), "\n")
Écart-type : 288.4105 
> cat("Variance :",var(data$Nombre_de_faillites), "\n")
Variance : 83180.62 
```





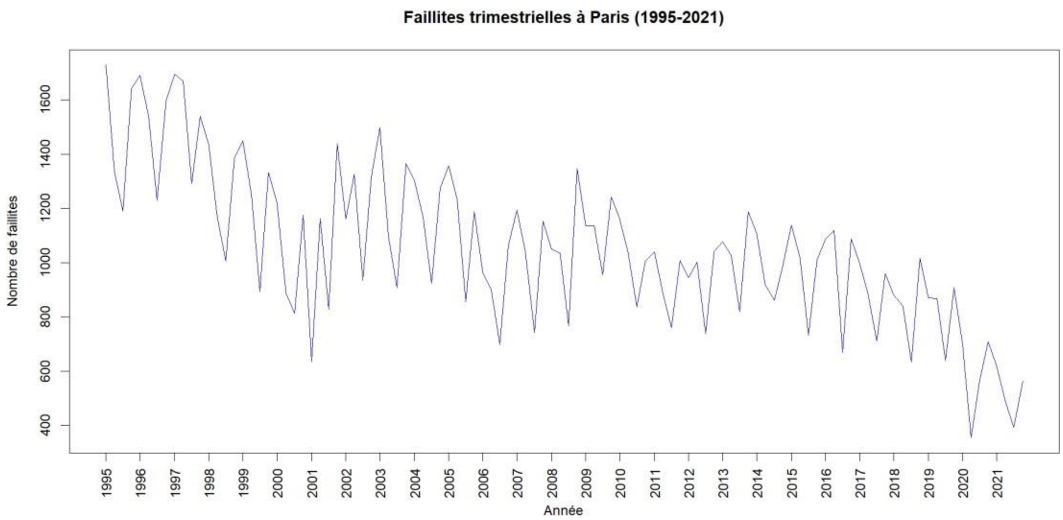
Récupération du fichier des données :

```
chemin <- "C:/Users/USER/OneDrive/Documents/S2/P2/Series chronologiques/defaillances_paris.txt"
data <- read.table(chemin, header = TRUE)
```

Conversion en série temporelle et représentation graphique de la série :

```
ts_data <- ts(data$Nombre_de_faillites,
              start = c(1995, 1),
              frequency = 4)
```

```
plot(ts_data,
     main = "Faillites trimestrielles à Paris (1995-2021)",
     ylab = "Nombre de faillites",
     xlab = "Année",
     col = "darkblue",
     xaxt = "n")


axis(1, at = seq(1995, 2021, by = 1), labels = seq(1995, 2021, by = 1), las = 2)
```



Interprétation :

La série montre une tendance décroissante et des pics qui révèlent une forte saisonnalité trimestrielle et une non-stationnarité confirmant la présence d'une tendance (justifiant l'utilisation d'un modèle SARIMA).

Division des données en un échantillon d'apprentissage (1995-2011) et un échantillon de validation (2012-2021) :

```
train <- window(ts_data, end = c(2011, 4))
test <- window(ts_data, start = c(2012, 1))
```

Test de la stationnarité de la série :

```
adf.test(train)

kpss.test(train)

        Augmented Dickey-Fuller Test

data:  train
Dickey-Fuller = -2.3566, Lag order = 4, p-value = 0.4303
alternative hypothesis: stationary
        KPSS Test for Level Stationarity

data:  train
KPSS Level = 1.0997, Truncation lag parameter = 3, p-value = 0.01
```

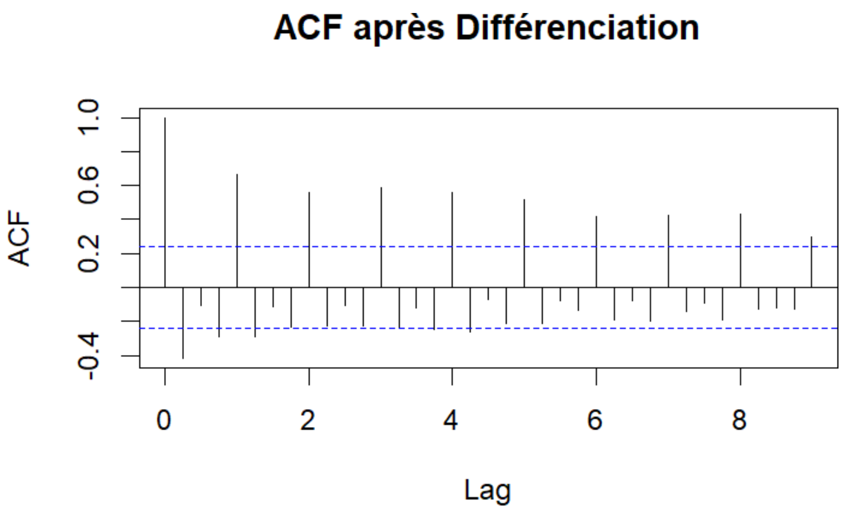
On a p-value>0.05 (Test ADF) et p-value<0.05 (Test KPSS), donc on doit faire une différenciation pour rendre la série stationnaire et par la suite on reteste la stationnarité.

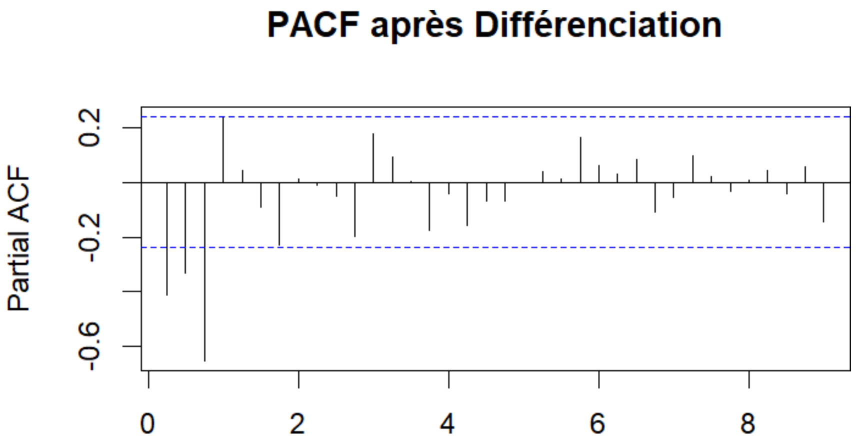
```
train_diff <- diff(train, differences = 1)
adf.test(train_diff)
kpss.test(train_diff)
        Augmented Dickey-Fuller Test

data:  train_diff
Dickey-Fuller = -3.7934, Lag order = 4, p-value = 0.0243
alternative hypothesis: stationary

        KPSS Test for Level Stationarity

data:  train_diff
KPSS Level = 0.1294, Truncation lag parameter = 3, p-value = 0.1
```

On trouve p-value<0.05 (Test ADF) et p-value>0.05 (Test KPSS) d'où la série après différenciation (d=1) est **stationnaire** (les deux tests concordent).

Traçage et Analyse des corrélogrammes simple et partiel de la série :

```
acf(train_diff, lag.max = 36, main = "ACF après Différenciation")
pacf(train_diff, lag.max = 36, main = "PACF après Différenciation")
```





L'ACF montre des pics à des multiples de 4 (saisonnalité trimestrielle), et la PACF diminue rapidement. Cela indique la présence de composantes saisonnières, ce qui guide le choix du modèle SARIMA.

```
grid <- expand.grid(
  p = 0:3,       # AR non-saisonnier (0 à 3)
  d = 1,         # Différenciation régulière fixée
  q = 0:1,       # MA non-saisonnier (0 ou 1)
  P = 0,         # SAR saisonnier fixé à 0
  D = 1,         # Différenciation saisonnière fixée
  Q = 0:1        # SMA saisonnier (0 ou 1)
)
```

La méthode de Box et Jenkins :

Identification du processus et de son ordre et estimation des paramètres du modèle :

Les combinaisons possibles pour identifier l'ordre des composantes du processus SARIMA:

```
  p d q P D Q
1 0 1 0 0 1 0
2 1 1 0 0 1 0
3 2 1 0 0 1 0
4 3 1 0 0 1 0
5 0 1 1 0 1 0
6 1 1 1 0 1 0
```

```
> evaluate_sarima(train, grid)
                 Model      AIC      BIC
1  (0,1,0)(0,1,0)[4] 862.4761 864.6956
2  (1,1,0)(0,1,0)[4] 851.2062 855.6452
3  (2,1,0)(0,1,0)[4] 851.2762 857.9347
4  (3,1,0)(0,1,0)[4] 852.1333 861.0113
5  (0,1,1)(0,1,0)[4] 848.7772 853.2162
6  (1,1,1)(0,1,0)[4] 844.3442 851.0027
7  (2,1,1)(0,1,0)[4] 844.5407 853.4187
8  (3,1,1)(0,1,0)[4] 847.6864 858.7840
9  (0,1,0)(0,1,1)[4] 834.5290 838.9680
10 (1,1,0)(0,1,1)[4] 820.0360 826.6945
11 (2,1,0)(0,1,1)[4] 820.6028 829.4809
12 (3,1,0)(0,1,1)[4] 822.1374 833.2349
13 (0,1,1)(0,1,1)[4] 818.7522 825.4107
14 (1,1,1)(0,1,1)[4] 820.1800 829.0581
15 (2,1,1)(0,1,1)[4] 821.4169 832.5144
16 (3,1,1)(0,1,1)[4] 822.9960 836.3130
```

Le plus petit AIC ou BIC

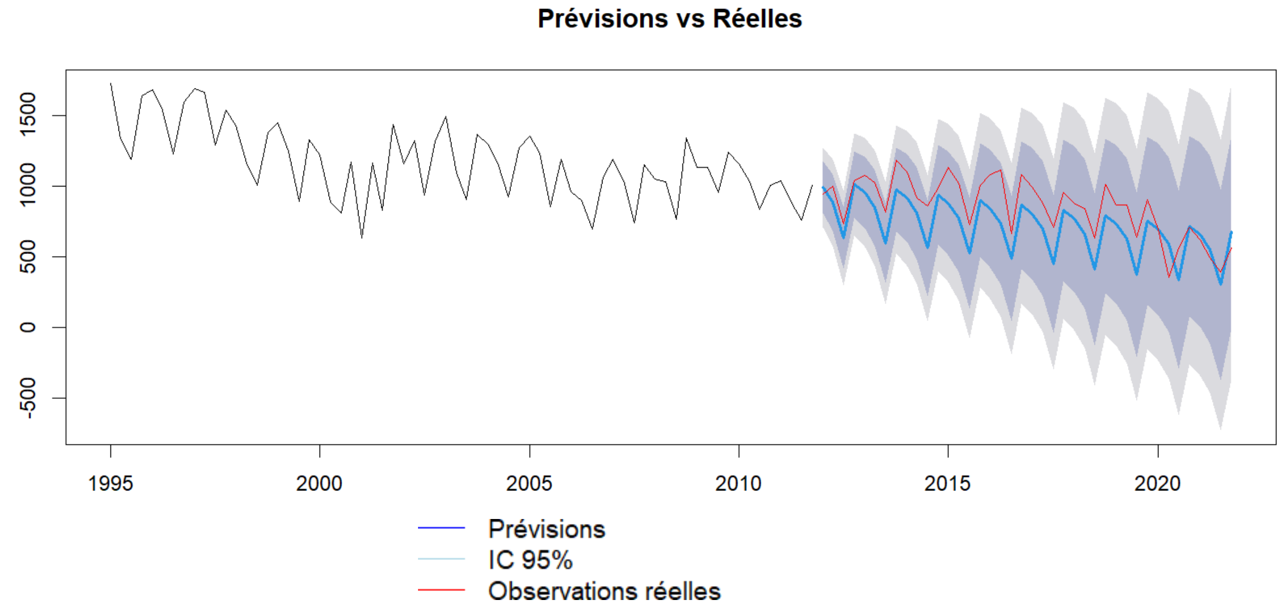
```
model <- arima(train,
               order = c(0, 1, 1),
               seasonal = list(order = c(0, 1, 1), period = 4))

coeftest(model)
```

```
z test of coefficients:

     Estimate Std. Error z value  Pr(>|z|)
ma1  -0.54157    0.10713 -5.0554 4.295e-07 ***
sma1 -1.00000    0.11700 -8.5466 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> residus <- residuals(model)
> box_test <- Box.test(residus, type = "Box-Pierce")
> cat("=== Test de Box-Pierce ===\n")
=== Test de Box-Pierce ===
> cat("Statistique X-squared:", box_test$statistic, "\n")
Statistique X-squared: 0.1657632
> cat("p-value:", box_test$p.value, "\n")
p-value: 0.683905
```

Nous avons choisi le modèle SARIMA (0,1,1) (0,1,1) car il représente le plus petit AIC et dont les coefficients sont significatifs (***). De plus, le test de Box-Pierce appliqué aux résidus donne p-value>0.05, ce qui indique l 'absence d 'autocorrelation des résidus (bruit blanc) et confirme la bonne spécification du modèle.

Prévision et représentation graphique de la série ainsi que la prévision :

```r
# Préparation des prévisions
forecast_values <- forecast(model_simple, h = length(test))

plot(forecast_values, main = "Prévisions vs Réelles")
lines(test, col = "red")
legend("topleft",
       col = c("black", "blue", "red"), lty = 1)
```



Evaluation des prévisions obtenues et interprétation des résultats :

```r
# Calcul des mesures d'erreur
mape=mean(abs(1-forecast_values$mean/test ))*100
mape
rmse=sqrt(mean((test - forecast_values$mean)^2))
rmse
```

```
> mape=mean(abs(1-forecast_values$mean/test ))*100
> mape
[1] 20.48444
> rmse=sqrt(mean((test - forecast_values$mean)^2))
> rmse
[1] 185.7862
```

On a les prévisions s'écartent de 20,48 % par rapport aux valeurs réelles d'où on peut accepter la prévision.

Test de normalité des erreurs :

```
# Test de normalité des erreurs
forecast_errors <- test - forecast_values$mean
shapiro.test(forecast_errors)


        Shapiro-Wilk normality test

data:  forecast_errors
W = 0.93677, p-value = 0.02695
```

On a p-value<0.05 d'où le test de normalité n'est pas vérifié (les résidus ne sont pas des bruits blancs gaussiens).

Conclusion :

L'analyse menée sur la série chronologique des faillites d'entreprises à Paris entre 1995 et 2021 a permis de mettre en évidence des caractéristiques essentielles telles qu'une tendance globale décroissante, une forte saisonnalité trimestrielle et une non-stationnarité initiale. Grâce à la différenciation, la série a pu être rendue stationnaire, permettant ainsi l'application rigoureuse de la méthode de Box et Jenkins.

Le modèle SARIMA (0,1,1) (0,1,1) retenu a démontré sa pertinence à travers des critères d'ajustement favorables (AIC minimal) et une absence d'autocorrélation des résidus, indiquant une bonne spécification du modèle. Les prévisions générées, bien que présentant un écart de 20,48 % par rapport aux données réelles, restent acceptables.