Time Series Analysis Homework08 Explaination

Hung-Ta, Wang (R13546017)

December 08, 2024

2. Recall the dataset "robot" firstly introduced in TSA HW06.

2.a. Use IMA(1, 1) to forecast five values ahead and calculate the 95% confidence intervals.

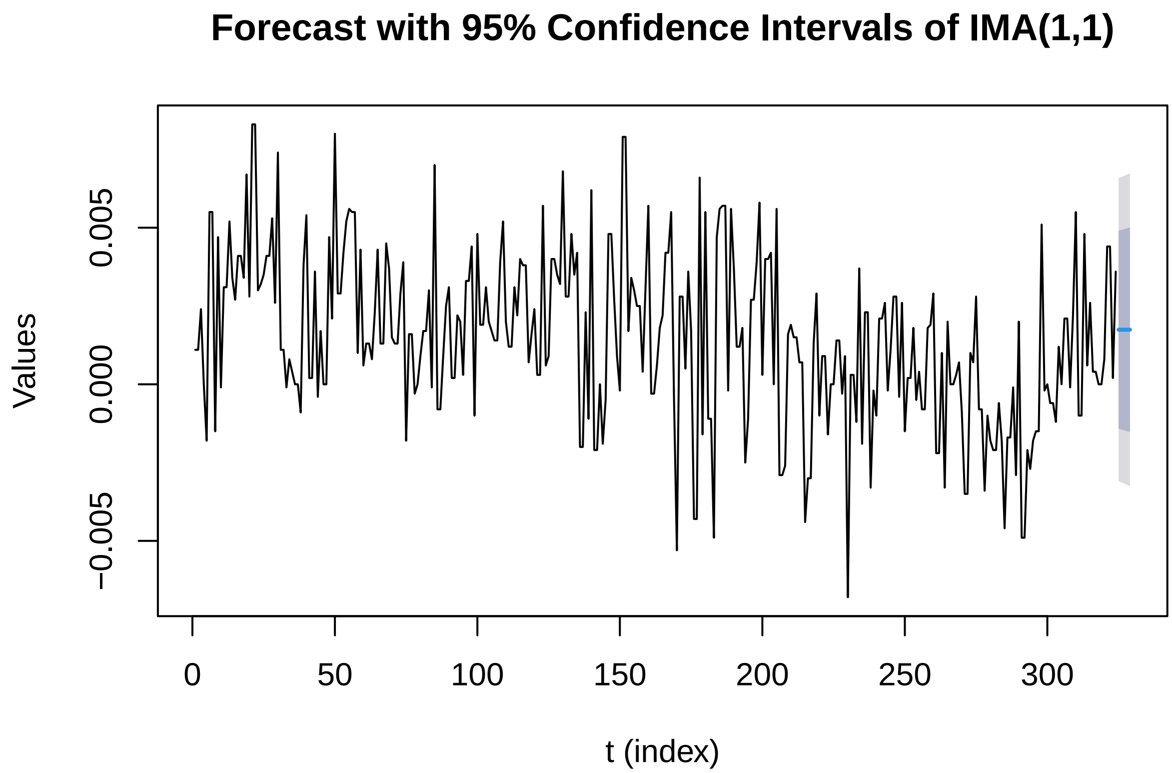
```
library(forecast)
library(TSA)
library(tseries)

robot_data <- read.csv("C:/Git_Code/Some-practice/TSA HW06.robot.csv")
robot_ts <- ts(robot_data$robot)
robot_ima <- Arima(robot_ts, order = c(0, 1, 1))
forecast_result_ima <- forecast(robot_ima, h = 5)
print(forecast_result_ima)
```

```
##     Point Forecast        Lo 80       Hi 80        Lo 95       Hi 95
## 325    0.001742672 -0.001419468 0.004904812 -0.003093404 0.006578748
## 326    0.001742672 -0.001445557 0.004930901 -0.003133304 0.006618647
## 327    0.001742672 -0.001471435 0.004956778 -0.003172880 0.006658224
## 328    0.001742672 -0.001497105 0.004982449 -0.003212140 0.006697483
## 329    0.001742672 -0.001522574 0.005007918 -0.003251091 0.006736435
```

Explanation: We can see that all forecast values are in 95% confidence intervals.

2.b. Display the actual values, the five forecasts and the 95% confidence intervals of the five forecasts, all in

```
plot(forecast_result_ima,
     main = "Forecast with 95% Confidence Intervals of IMA(1,1)",
     xlab = "t (index)", ylab = "Values")
```



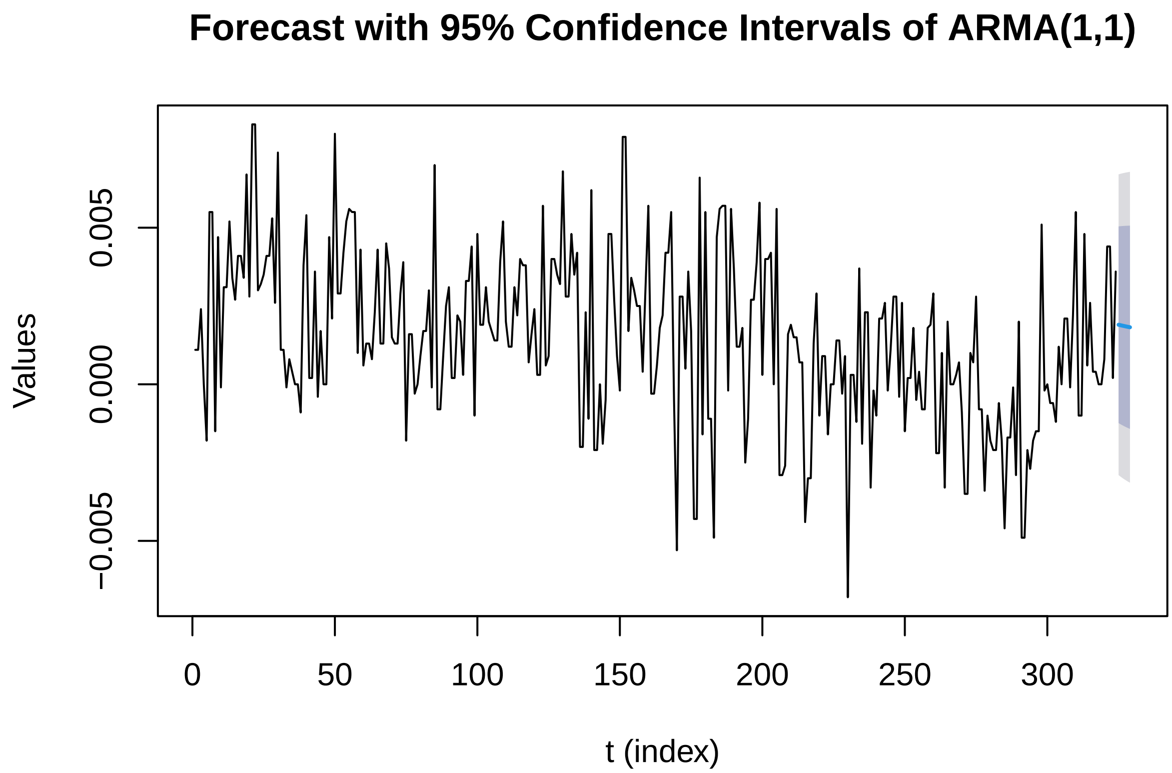
Explanation: We can see that the blue line is in the gray zone.

2.c. Use ARMA(1, 1) to forecast five values ahead and calculate the 95% confidence intervals. Compare the results with those in (a), what do you observe?

```
robot_arma <- Arima(robot_ts, order = c(1, 0, 1))
forecast_result_arma <- forecast(robot_arma, h = 5)
print(forecast_result_arma)
```

```
##     Point Forecast        Lo 80       Hi 80        Lo 95       Hi 95
## 325    0.001901348 -0.001238721 0.005041418 -0.002900973 0.006703670
## 326    0.001879444 -0.001291705 0.005050593 -0.002970410 0.006729298
## 327    0.001858695 -0.001340084 0.005057474 -0.003033415 0.006750805
## 328    0.001839041 -0.001384327 0.005062409 -0.003090675 0.006768757
## 329    0.001820424 -0.001424848 0.005065697 -0.003142792 0.006783640
```

```
plot(forecast_result_arma,
     main = "Forecast with 95% Confidence Intervals of ARMA(1,1)",
     xlab = "t (index)", ylab = "Values")
```

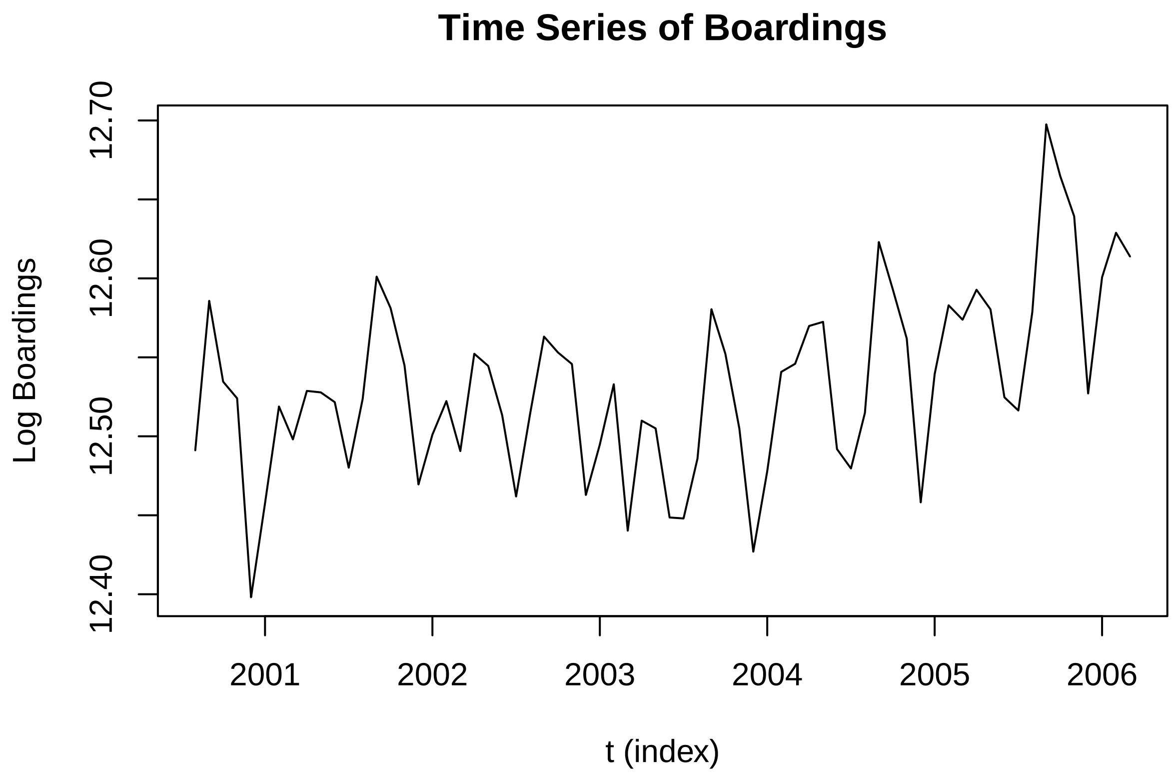


Explanation: We can see that all forecast values are in 95% confidence intervals just like the blue line is in the gray zone.

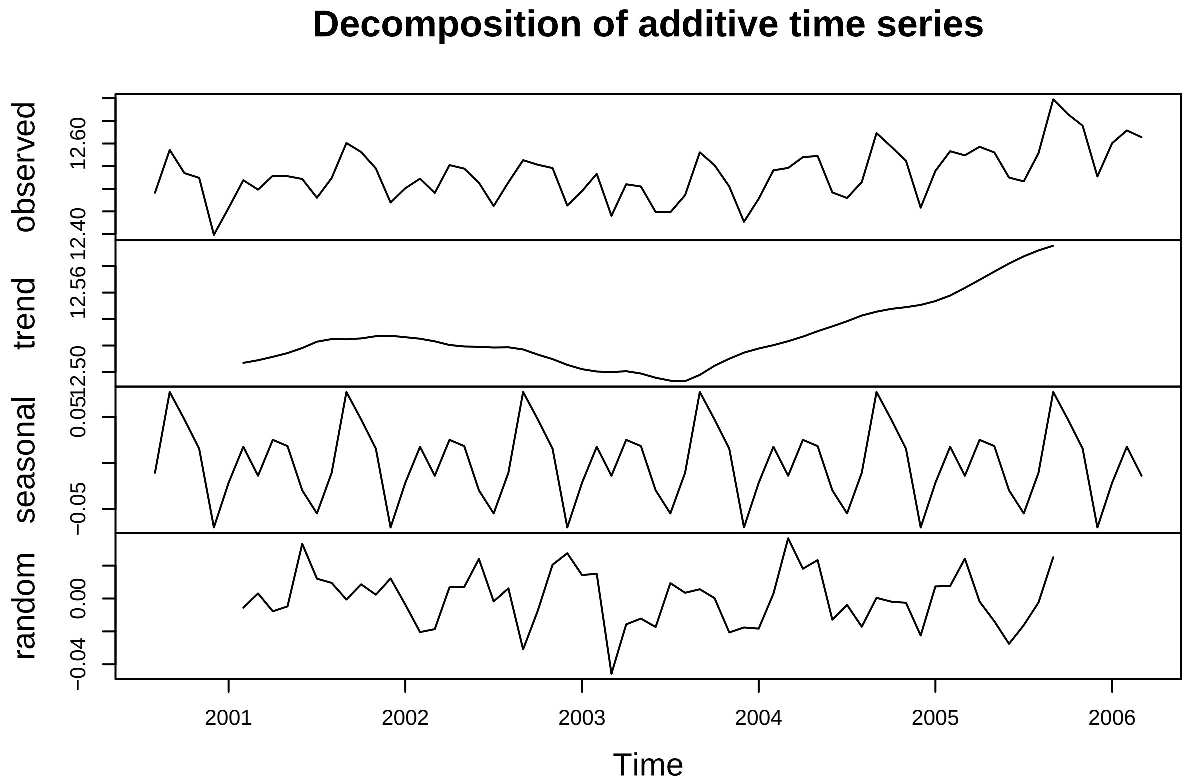
3. The dataset "boardings" contains the monthly number of passengers who boarded light rail trains and buses in Denver, Colorado, from August 2000 to March 2006.

3.a. Plot the time series and tell your observation if there exists seasonality and if the series is stationary.

```
boardings_data <- read.csv("C:/Git_Code/Some-practice/TSA HW08.boardings.csv")
boardings_ts <- ts(boardings_data$log_boardings, start = c(2000, 8), frequency = 12)
plot(boardings_ts, main = "Time Series of Boardings",
     xlab = "t (index)", ylab = "Log Boardings")
```



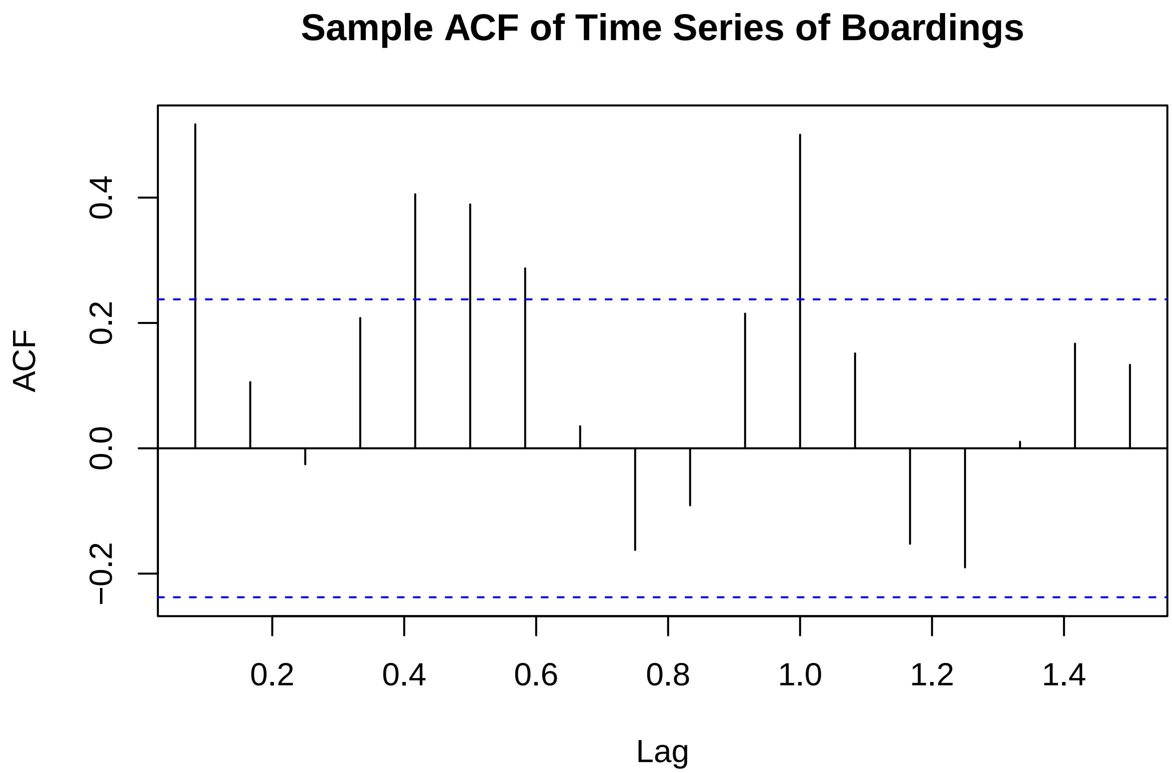

```r
decomposed_boardings_ts <- decompose(boardings_ts)
plot(decomposed_boardings_ts)
```



Explanation: We can see that the observation exists seasonality by the graph of decomposed time series. The series is not stationary by its seasonality and trend.

3.b. Plot the sample ACF and see what are the significant lags?

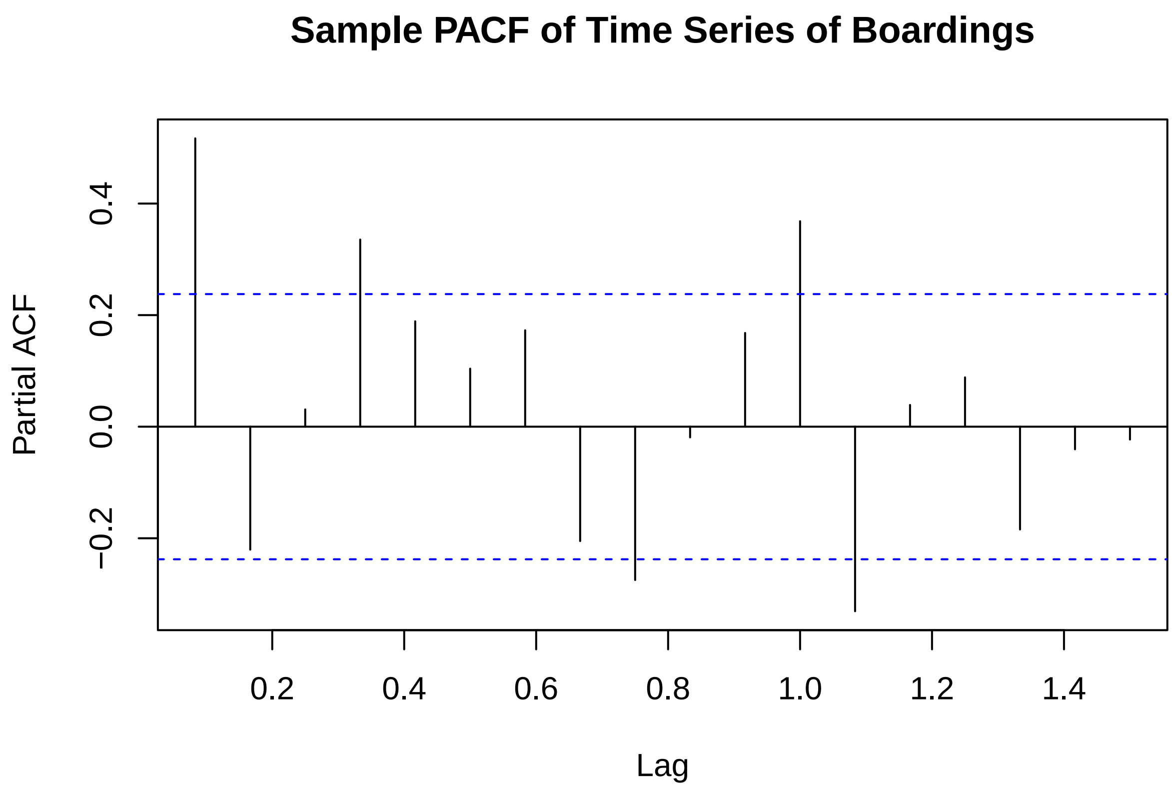
```
acf(boardings_ts, main = "Sample ACF of Time Series of Boardings")
```



Explanation: We can observe that the $1^{st}, 5^{th}, 6^{th}, 7^{th}, 12^{th}$ bar is out of the interval. These are the significant lags.

3.c. Fit the data with ARMA(0, 3) $\times$ (1,0)_12, evaluate if the estimated coefficients $\{\theta_1, \theta_2, \theta_3, \phi_{12}\}$ are significant.

```
pacf(boardings_ts, main = "Sample PACF of Time Series of Boardings")
```




```r
model <- Arima(boardings_ts, order = c(0, 0, 3), seasonal = c(1, 0, 0))
summary(model)
```

```
## Series: boardings_ts
## ARIMA(0,0,3)(1,0,0)[12] with non-zero mean
##
## Coefficients:
##          ma1     ma2     ma3    sar1     mean
##       0.7290  0.6116  0.2950  0.8776  12.5455
## s.e.  0.1186  0.1172  0.1118  0.0507   0.0354
##
## sigma^2 = 0.0007062:  log likelihood = 143.54
## AIC=-275.09   AICc=-273.71   BIC=-261.77
##
## Training set error measures:
##                       ME       RMSE        MAE        MPE      MAPE      MASE
## Training set 0.004380537 0.02557808 0.02064695 0.03447698 0.1647581 0.5748688
##                     ACF1
## Training set -0.01836489
```

```r
coefficients <- coef(model)
standard_errors <- sqrt(diag(model$var.coef))
t_values <- coefficients / standard_errors
results <- data.frame(
  Coefficient = coefficients,
```

```
  `Standard Error` = standard_errors,
  `t-value` = t_values
)
print(results)
```

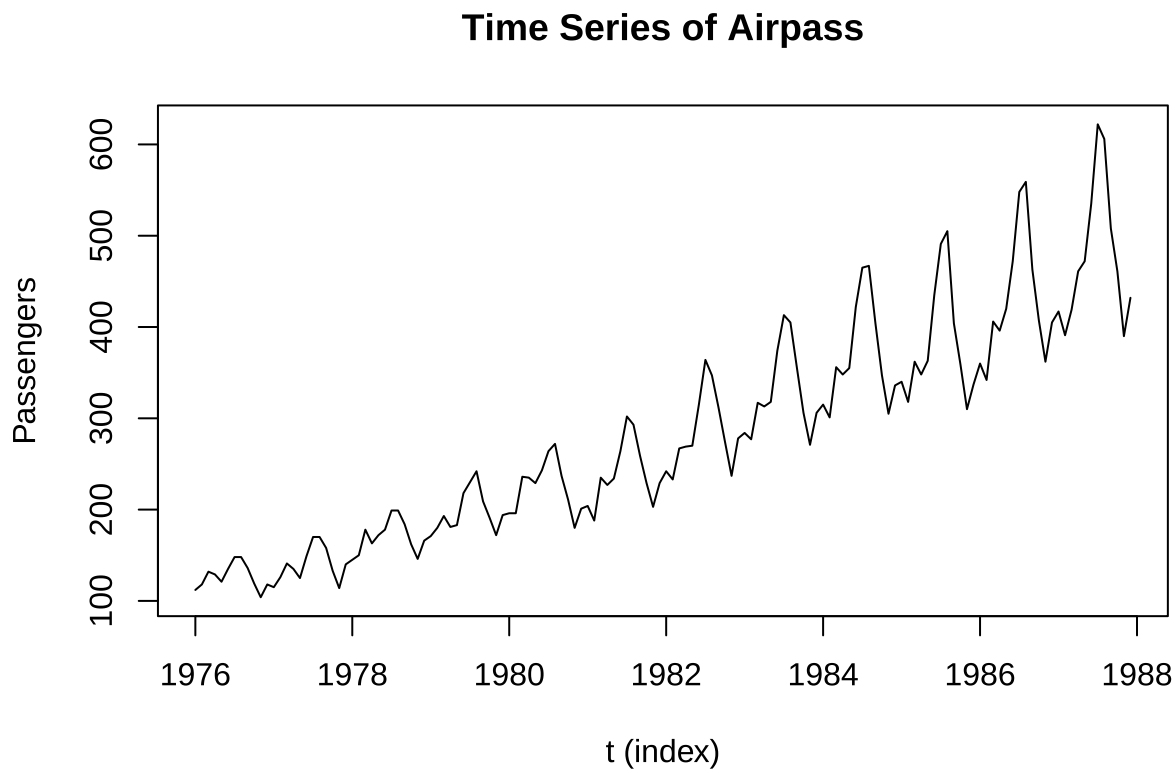
```
##           Coefficient Standard.Error    t.value
## ma1         0.7289691     0.11855662   6.148700
## ma2         0.6116224     0.11718350   5.219356
## ma3         0.2950298     0.11179703   2.638977
## sar1        0.8776131     0.05071032  17.306401
## intercept  12.5454987     0.03542996 354.092990
```

Explanation: We have degree of freedom: 69-1 = 68. If t.value is greater than 2 or less than 2, it is significant. For 95% confidence interval, we can see that t.value are all greater than 2. Thus, all coefficients are significant. However, t.value of ma1, ma2, ma3 are decreasing since ma1 is significant by observing ACF.

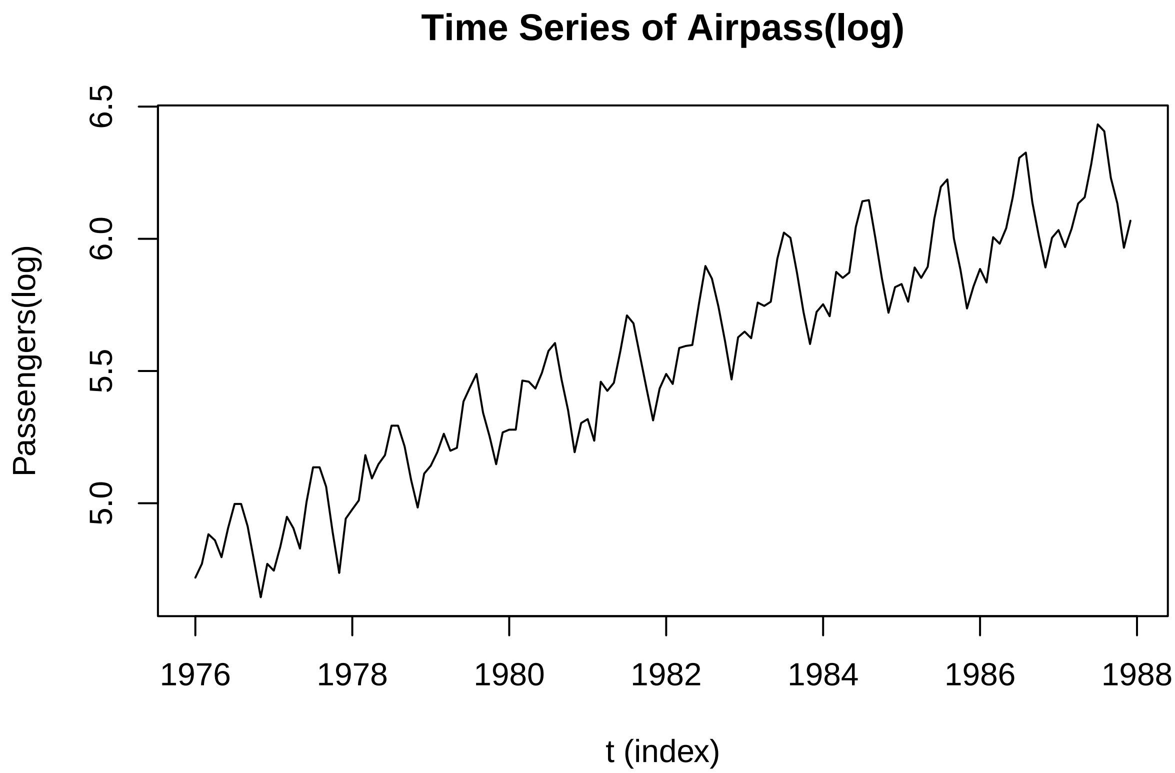
4. The monthly airline passengers, first investigated by Box and Jenkins in 1976, is considered as the classic time series dataset (see "TSA HW08.airpass.csv").

4.a. Plot the time series in its original scale and the log-transformed scale. Do you think making the log-transformation is appropriate?

```
airpass_data <- read.csv("C:/Git_Code/Some-practice/TSA HW08.airpass.csv")
airpass_ts <- ts(airpass_data$airpass, start = c(1976,1) , frequency = 12)
log_airpass_ts <- log(airpass_ts)

plot(airpass_ts,
     main = "Time Series of Airpass",
     xlab = "t (index)", ylab = "Passengers")
```



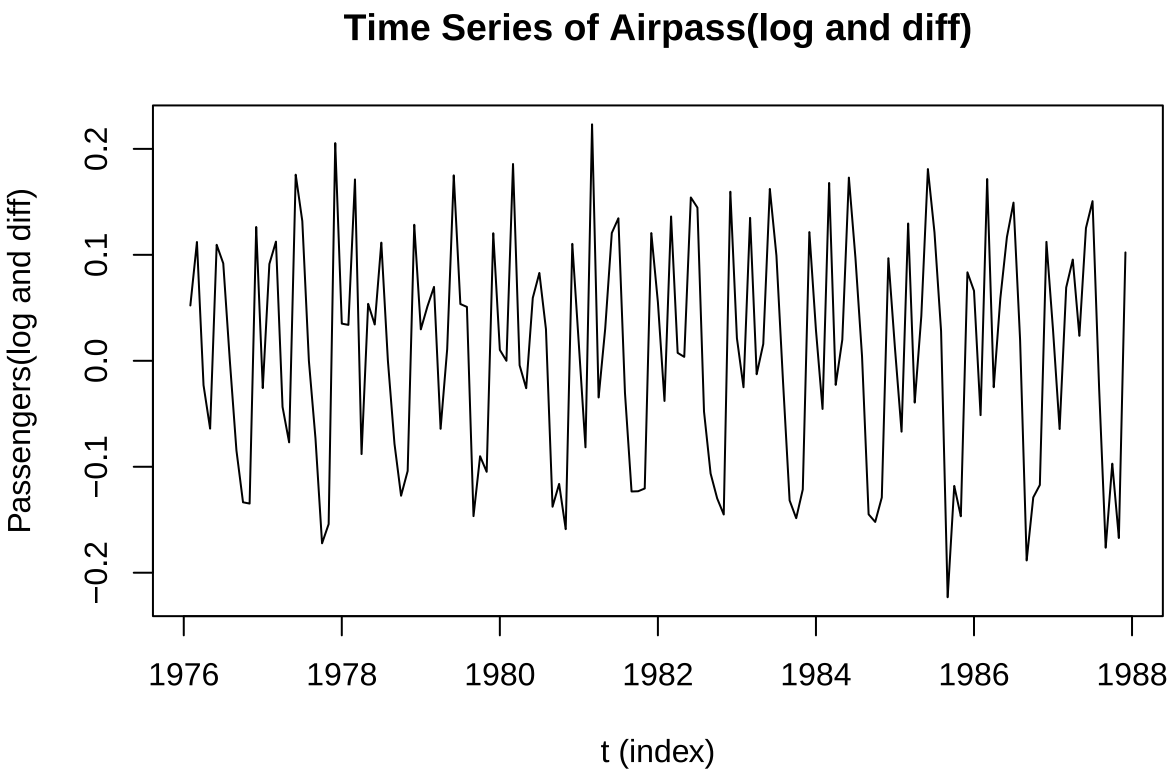

```r
plot(log_airpass_ts,
     main = "Time Series of Airpass(log)",
     xlab = "t (index)", ylab = "Passengers(log)")
```



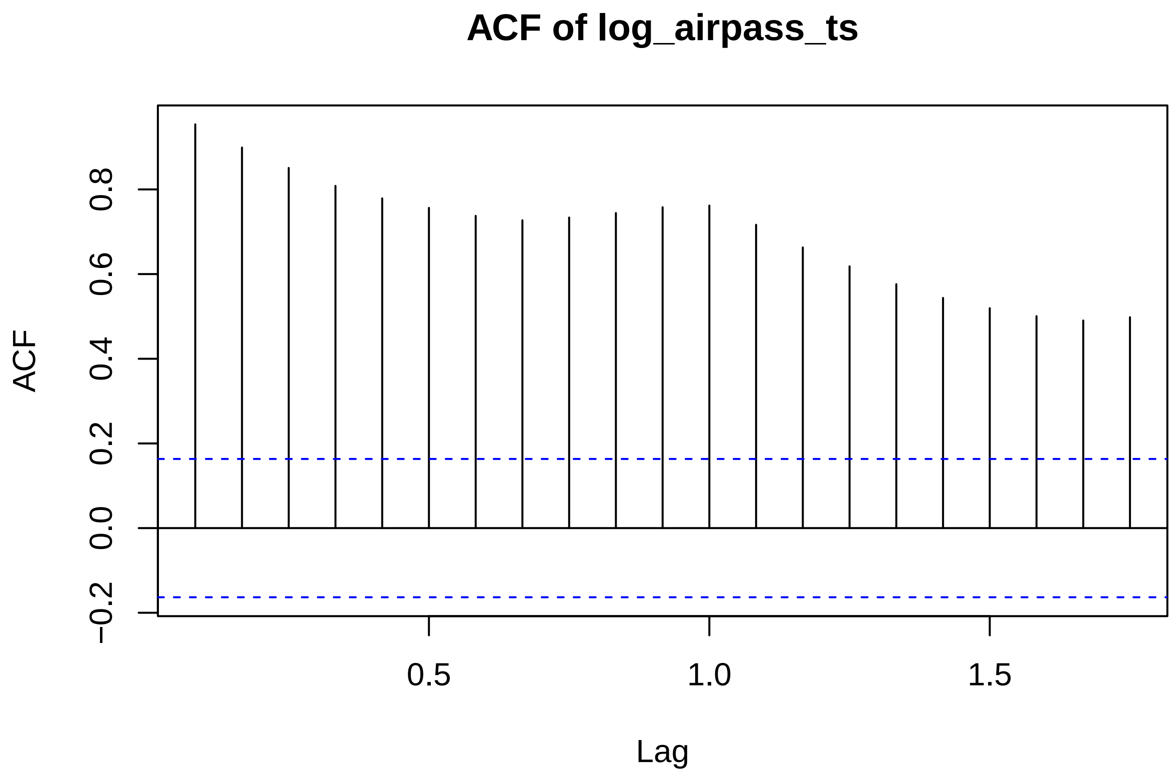
Explanation: We can see that there are trend and seasonality in the time series no matter it is log-transformed. From the aspect of making data stationary, I do not think log-transformation is an appropriate way.

4.b. Make the first-order difference over the "log-transformed" data. What do you observe?

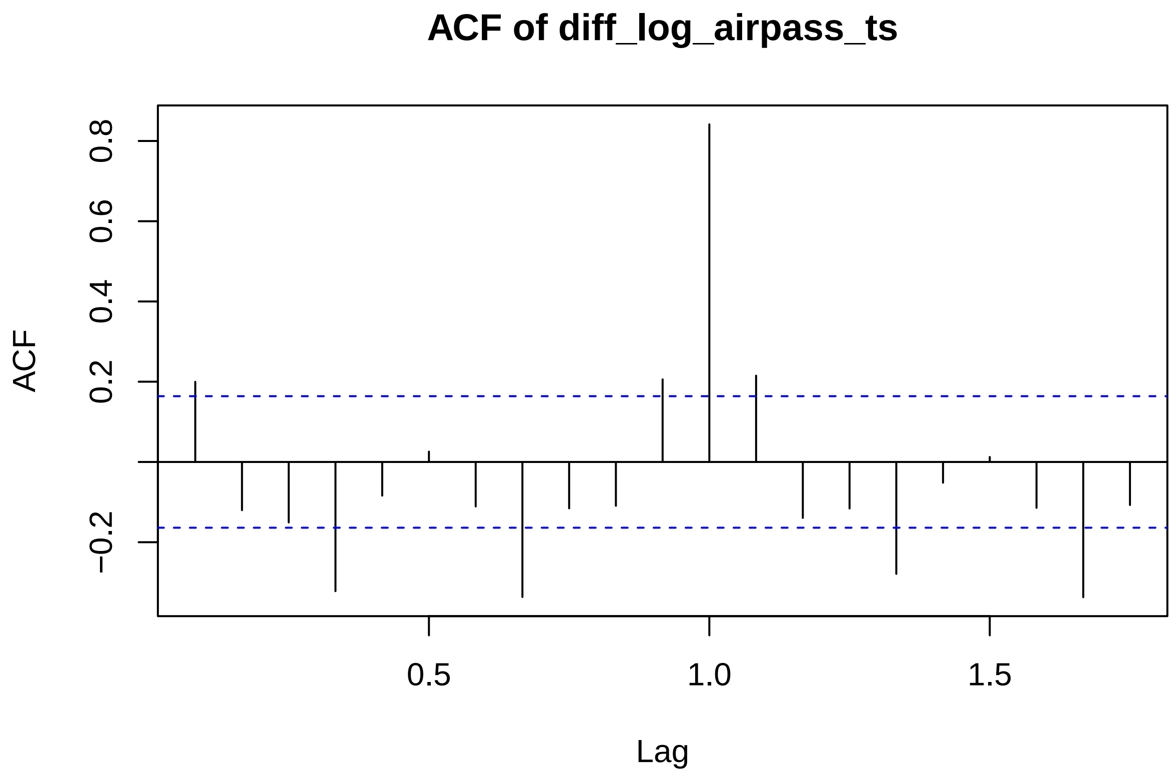
```
diff_log_airpass_ts <- diff(log_airpass_ts)
plot(diff_log_airpass_ts,
     main = "Time Series of Airpass(log and diff)",
     xlab = "t (index)", ylab = "Passengers(log and diff)")
```




```r
acf(log_airpass_ts, main = "ACF of log_airpass_ts")
```



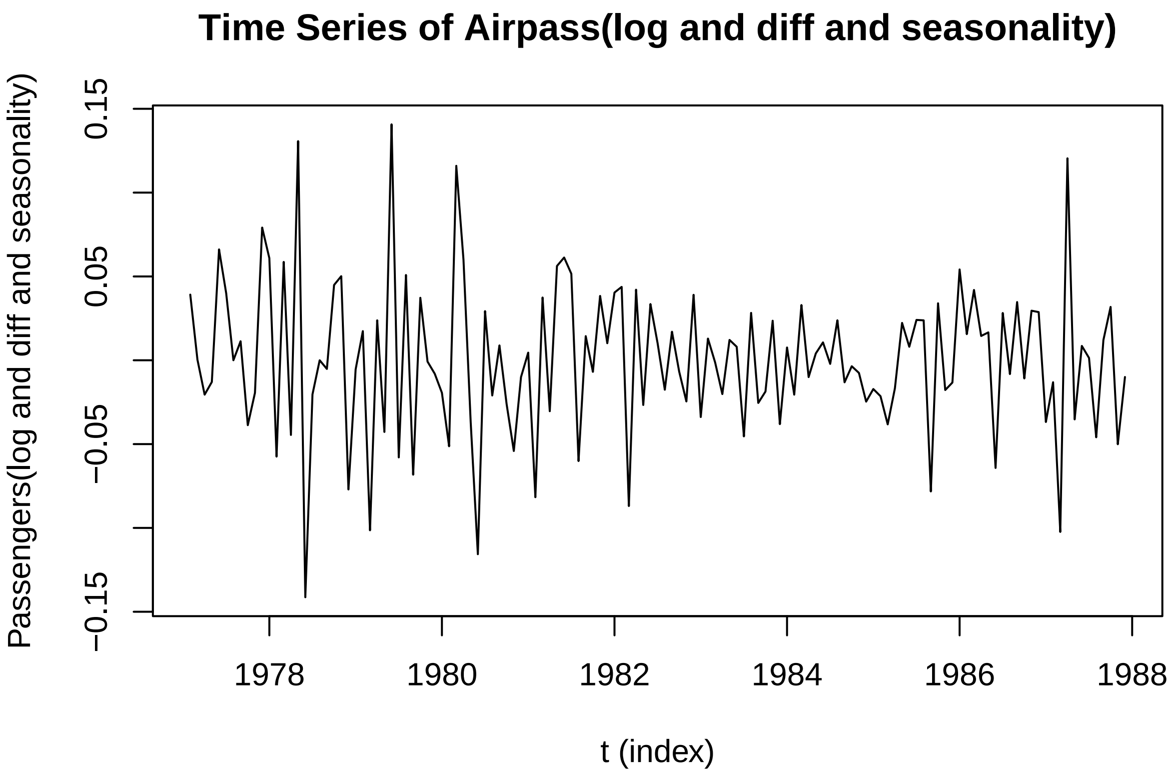

```r
acf(diff_log_airpass_ts, main = "ACF of diff_log_airpass_ts")
```



Explanation: We can see that the first-order difference over the log-transformed data may be not stationary. However, from the aspect of stationarity , it is much better than log-transformation.

4.c. Make a seasonal difference of the resulted series in (b), what do you observe?

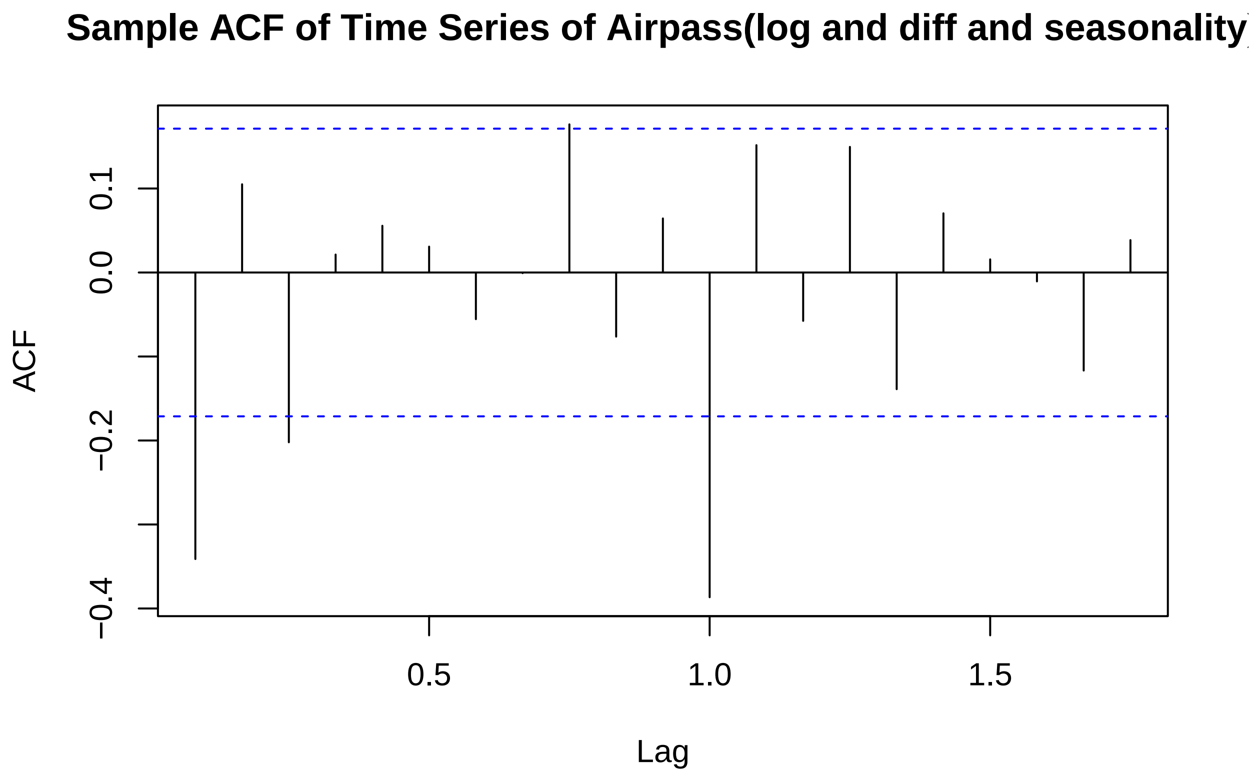
```
seasonality_diff_log_airpass_ts <- diff(diff_log_airpass_ts, lag = 12)
plot(seasonality_diff_log_airpass_ts,
     main = "Time Series of Airpass(log and diff and seasonality)",
     xlab = "t (index)", ylab = "Passengers(log and diff and seasonality)")
```



Explanation: We can see that the time series seems to be more stationary than the resulted series in (b). However, we still cannot ensure that it is stationary merely by the graph.

4.d. Plot the sample ACF of the resulted series in (c), explain what you see.

```
acf(seasonality_diff_log_airpass_ts,
    main = "Sample ACF of Time Series of Airpass(log and diff and seasonality)")
```



Explanation: We can see that there are more sample ACF in the confidence interval which means there are less ACF are significant. However, we cannot observe an obvious pattern, such as cut-off in the ACF graph. However, from the aspect of stationarity , it is a bit better than the resulted series in (b).

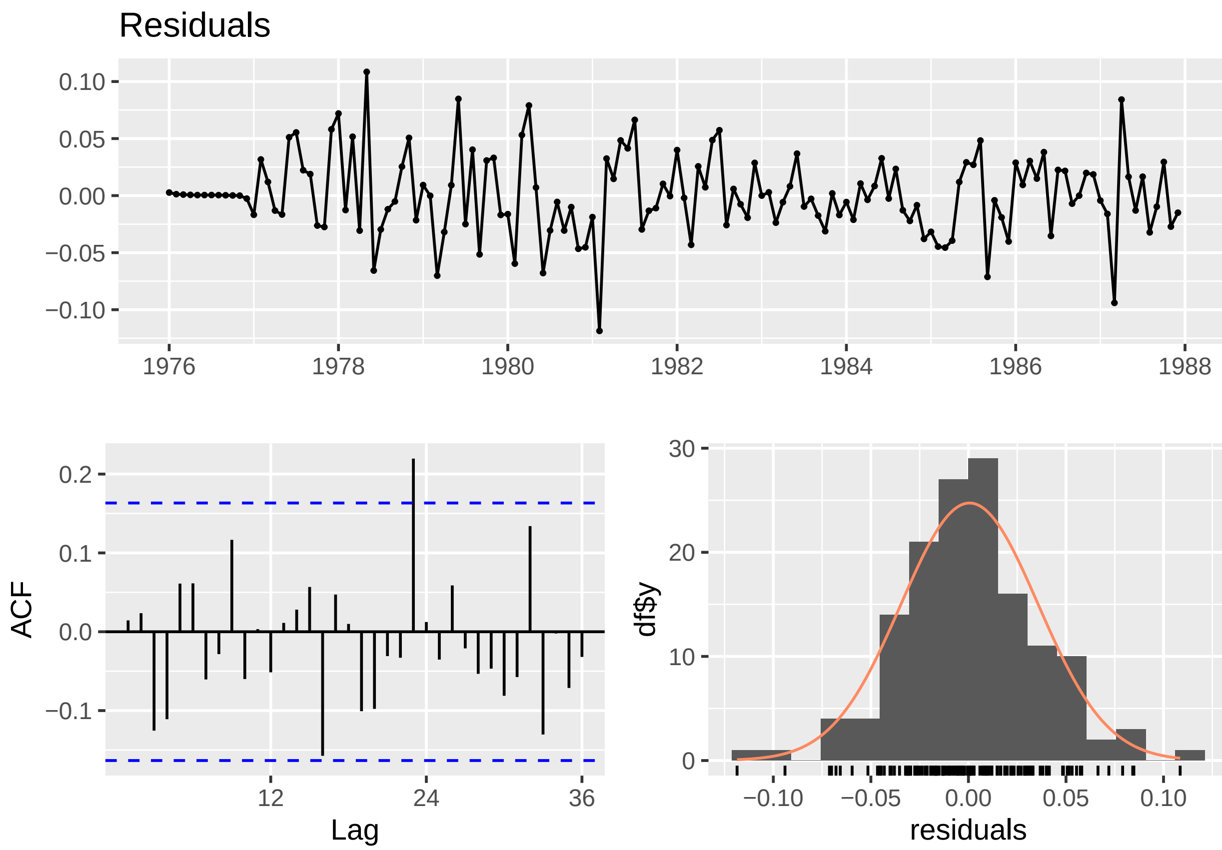
4.e. Fit an ARIMA(0,1,1) × (0,1,1)__12 model to the log-transformed series. Diagnose the residuals of this model, including the sample ACF and the normality test

```r
airpass_sarima <- Arima(log_airpass_ts, order = c(0, 1, 1), seasonal = c(0, 1, 1))
summary(airpass_sarima)
```

```
## Series: log_airpass_ts
## ARIMA(0,1,1)(0,1,1)[12]
##
## Coefficients:
##           ma1     sma1
##       -0.4018  -0.5569
## s.e.   0.0896   0.0731
##
## sigma^2 = 0.001371:  log likelihood = 244.7
## AIC=-483.4   AICc=-483.21   BIC=-474.77
##
## Training set error measures:
##                         ME       RMSE        MAE         MPE       MAPE       MASE
```

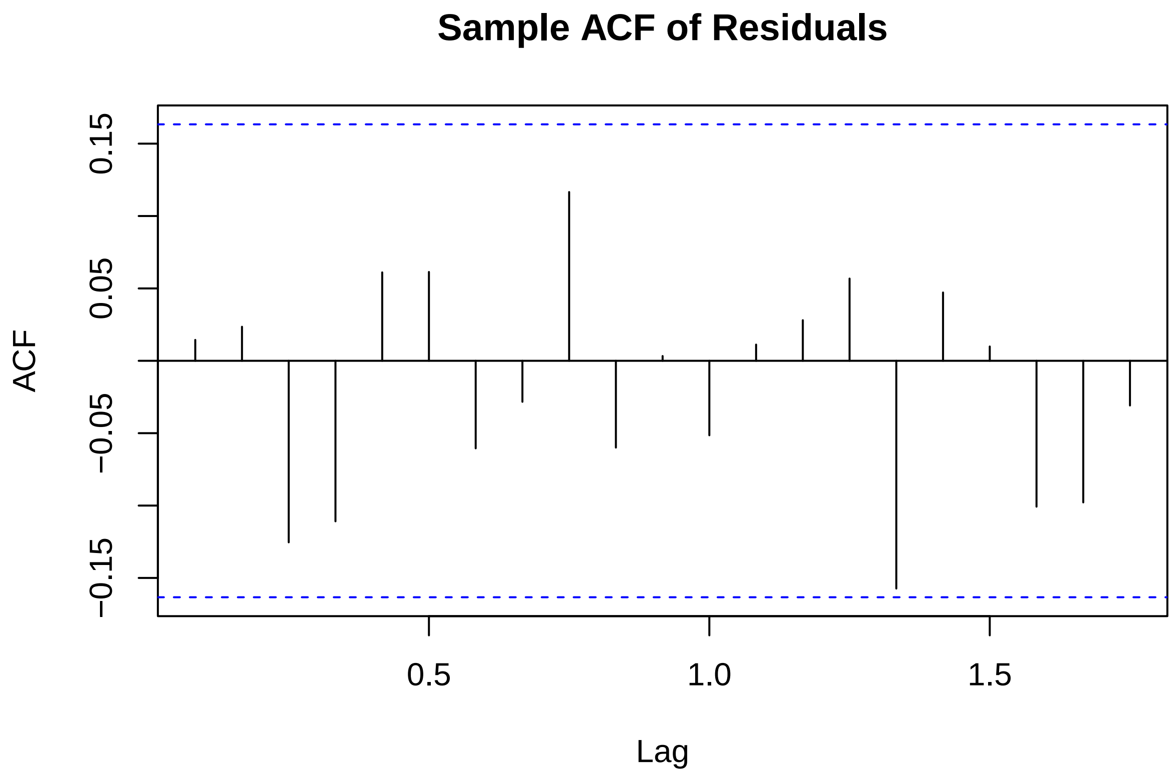
```
## Training set 0.0005730622 0.03504883 0.02626034 0.01098898 0.4752815 0.2169522
##                     ACF1
## Training set 0.01443892
```

```r
residuals <- residuals(airpass_sarima)
checkresiduals(residuals)
```




```
##
##  Ljung-Box test
##
## data:  Residuals
## Q* = 26.446, df = 24, p-value = 0.3309
##
## Model df: 0.   Total lags used: 24
```

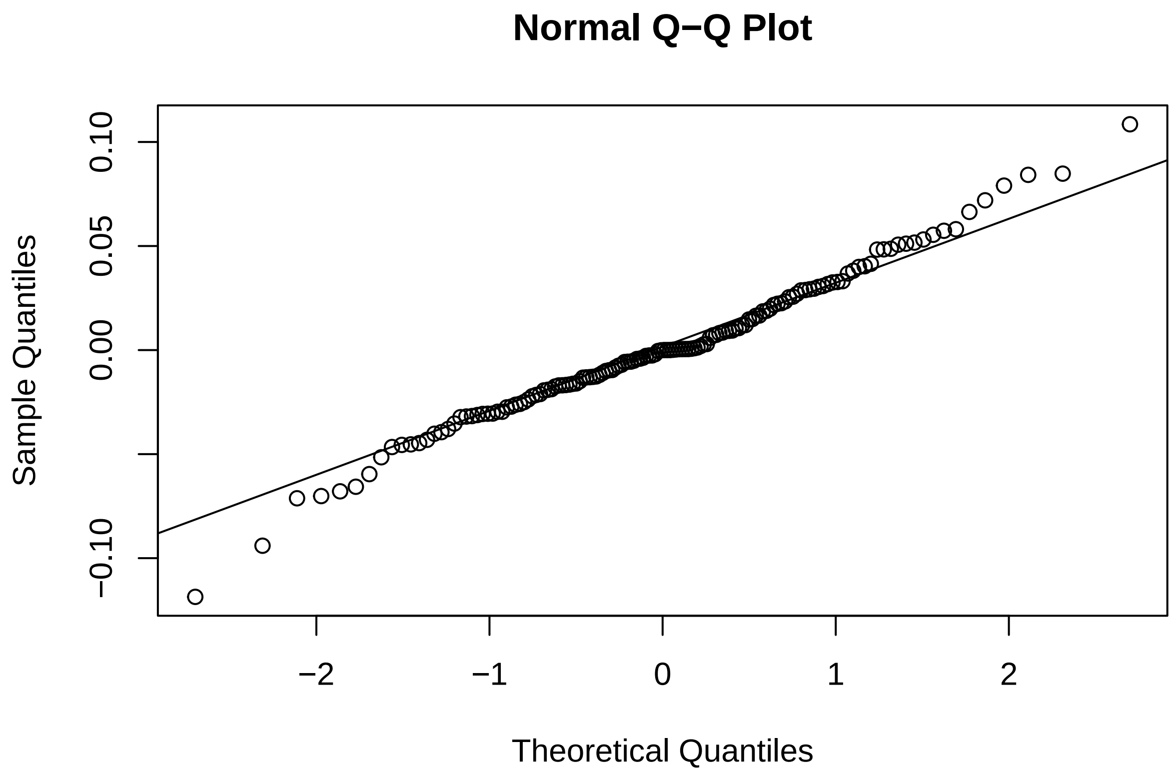
```r
acf(residuals, main = "Sample ACF of Residuals")
```




```r
adf.test(residuals)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  residuals
## Dickey-Fuller = -4.7907, Lag order = 5, p-value = 0.01
## alternative hypothesis: stationary
```

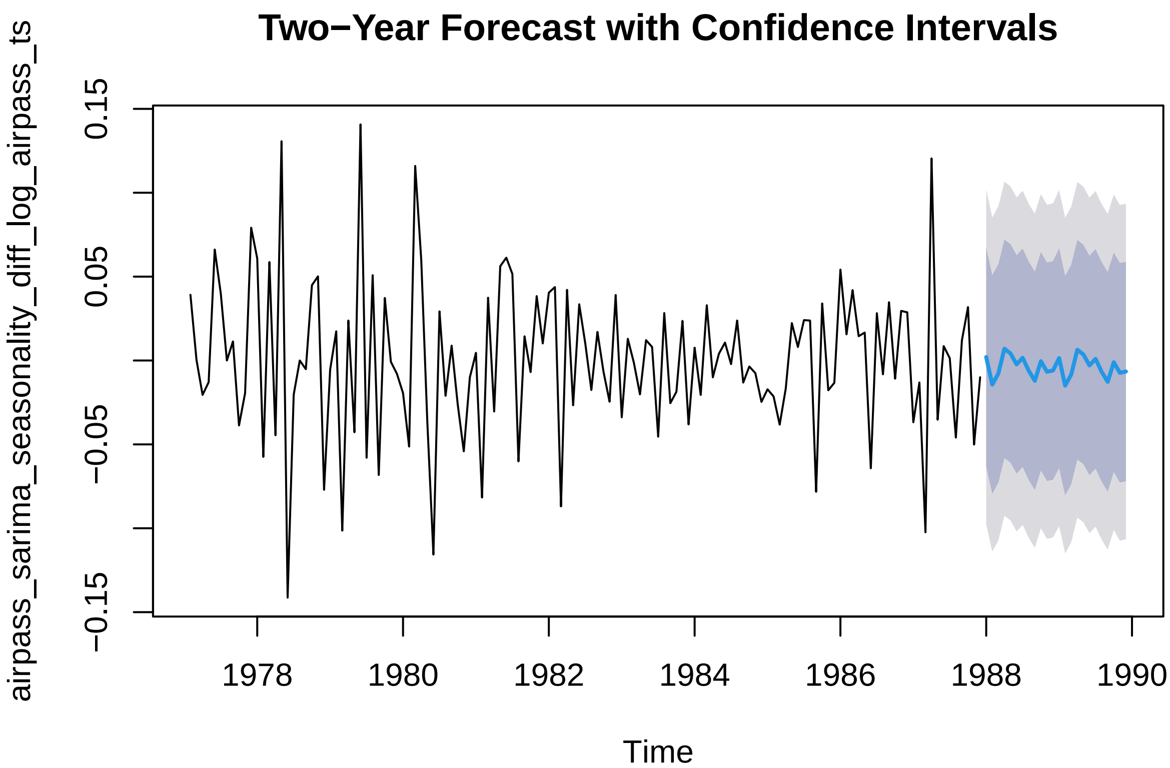
```r
qqnorm(residuals)
qqline(residuals)
```



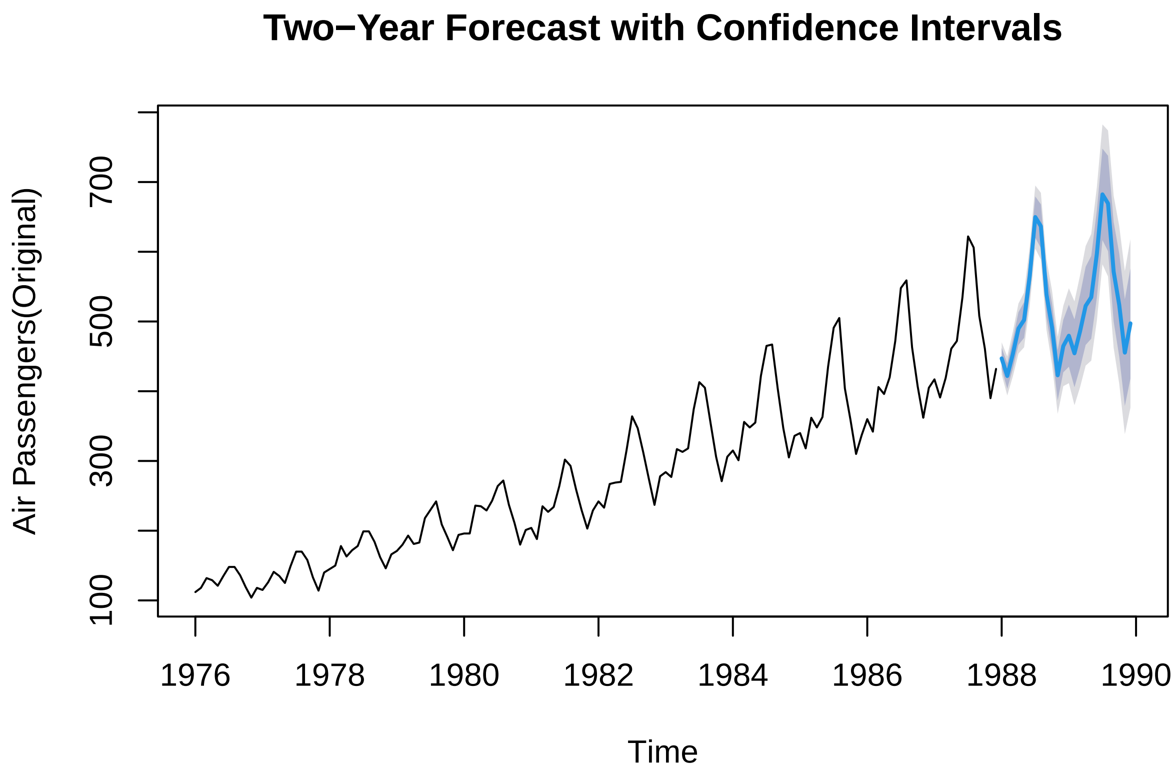
Explanation: Sample ACF shows that the ACF of the residuals are separate and there is not a obvious pattern. We guess that the ACF may be stationary. Moreover, We can see that the Ljung Box test and ADF test are all show that the residuals is a stationary process. From Normal Q Q Plot, The residuals are approximately normal, it may slightly not follow normal distribution on two side.

4.f. Make forecasts for "two" years based on the model in (e). The confidence intervals shall be included.

```
airpass_sarima_seasonality_diff_log_airpass_ts <- Arima(seasonality_diff_log_airpass_ts, order = c(0, 1
forecasts_airpass_sarima_seasonality_diff_log_airpass_ts <- forecast(airpass_sarima_seasonality_diff_log
plot(forecasts_airpass_sarima_seasonality_diff_log_airpass_ts, main = "Two-Year Forecast with Confidence
     xlab = "Time", ylab = "airpass_sarima_seasonality_diff_log_airpass_ts")
```




```r
airpass_sarima_airpass_ts <- Arima(airpass_ts, order = c(0, 1, 1), seasonal = c(0, 1, 1))
forecasts_Passengers <- forecast(airpass_sarima_airpass_ts, h = 24)
plot(forecasts_Passengers, main = "Two-Year Forecast with Confidence Intervals",
     xlab = "Time", ylab = "Air Passengers(Original)")
```




```r
print(forecasts_airpass_sarima_seasonality_diff_log_airpass_ts)
```

```
##          Point Forecast       Lo 80      Hi 80       Lo 95      Hi 95
## Jan 1988   0.0020661461 -0.06318224 0.06731453 -0.09772264 0.10185494
## Feb 1988  -0.0143375103 -0.07945094 0.05077592 -0.11391991 0.08524489
## Mar 1988  -0.0076456494 -0.07275908 0.05746778 -0.10722805 0.09193675
## Apr 1988   0.0070334928 -0.05807994 0.07214693 -0.09254891 0.10661589
## May 1988   0.0042231440 -0.06089029 0.06933658 -0.09535926 0.10380555
## Jun 1988  -0.0023040620 -0.06741749 0.06280937 -0.10188646 0.09727834
## Jul 1988   0.0015988760 -0.06351456 0.06671231 -0.09798353 0.10118128
## Aug 1988  -0.0061098467 -0.07122328 0.05900359 -0.10569225 0.09347256
## Sep 1988  -0.0120899410 -0.07720337 0.05302349 -0.11167234 0.08749246
## Oct 1988  -0.0004272932 -0.06554073 0.06468614 -0.10000970 0.09915511
## Nov 1988  -0.0066969924 -0.07181043 0.05841644 -0.10627939 0.09288541
## Dec 1988  -0.0059239167 -0.07103735 0.05918952 -0.10550632 0.09365849
## Jan 1989   0.0014426884 -0.06407476 0.06696014 -0.09875760 0.10164298
## Feb 1989  -0.0149609680 -0.08036644 0.05044451 -0.11499001 0.08506807
## Mar 1989  -0.0082691071 -0.07367458 0.05713637 -0.10829815 0.09175993
## Apr 1989   0.0064100352 -0.05899544 0.07181551 -0.09361900 0.10643907
## May 1989   0.0035996863 -0.06180579 0.06900516 -0.09642935 0.10362873
## Jun 1989  -0.0029275197 -0.06833299 0.06247795 -0.10295656 0.09710152
## Jul 1989   0.0009754183 -0.06443005 0.06638089 -0.09905362 0.10100446
## Aug 1989  -0.0067333044 -0.07213878 0.05867217 -0.10676234 0.09329573
## Sep 1989  -0.0127133987 -0.07811887 0.05269207 -0.11274244 0.08731564
## Oct 1989  -0.0010507508 -0.06645622 0.06435472 -0.10107979 0.09897829
```

```
## Nov 1989  -0.0073204501 -0.07272592 0.05808502 -0.10734949 0.09270859
## Dec 1989  -0.0065473743 -0.07195285 0.05885810 -0.10657641 0.09348166
```

```r
print(forecasts_Passengers)
```

```
##          Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## Jan 1988       447.0532 432.0243 462.0821 424.0685 470.0379
## Feb 1988       421.8774 403.6068 440.1481 393.9349 449.8200
## Mar 1988       453.5262 432.5080 474.5444 421.3816 485.6708
## Apr 1988       489.9008 466.4548 513.3468 454.0433 525.7584
## May 1988       502.1835 476.5385 527.8285 462.9629 541.4041
## Jun 1988       564.2246 536.5549 591.8943 521.9074 606.5418
## Jul 1988       649.7953 620.2392 679.3514 604.5931 694.9974
## Aug 1988       636.7146 605.3856 668.0437 588.8009 684.6284
## Sep 1988       538.9209 505.9139 571.9279 488.4411 589.4007
## Oct 1988       491.0672 456.4636 525.6708 438.1456 543.9889
## Nov 1988       422.8243 386.6945 458.9540 367.5686 478.0800
## Dec 1988       464.7526 427.1586 502.3466 407.2575 522.2476
## Jan 1989       479.5818 435.0853 524.0782 411.5303 547.6332
## Feb 1989       454.4060 405.7584 503.0535 380.0060 528.8060
## Mar 1989       486.0548 433.5835 538.5260 405.8069 566.3026
## Apr 1989       522.4294 466.3947 578.4640 436.7318 608.1269
## May 1989       534.7120 475.3275 594.0966 443.8912 625.5329
## Jun 1989       596.7532 534.1978 659.3086 501.0830 692.4234
## Jul 1989       682.3238 616.7508 747.8969 582.0385 782.6091
## Aug 1989       669.2432 600.7854 737.7010 564.5460 773.9404
## Sep 1989       571.4495 500.2236 642.6753 462.5189 680.3800
## Oct 1989       523.5958 449.7055 597.4860 410.5904 636.6011
## Nov 1989       455.3528 378.8910 531.8147 338.4145 572.2911
## Dec 1989       497.2811 418.3314 576.2309 376.5379 618.0243
```

Explanation: We can see that no matter original and seasonality diff log data are all in confidence interval.