TIME SERIES AND FORCASTING

BDA542AN

CIA - 3:

BY

Umang Sharma (21215210),5 BDA

SUBMITTED TO-

Dr Ashish Sharma

CHRIST
(DEEMED TO BE UNIVERSITY)
DELHI-NCR, INDIA

SCHOOL OF SCIENCES

2022-2023

Aim: To Develop an ARIMA model and perform forecasting on a real-world dataset

Software used: - R Studio

About the Dataset -:

Context

This dataset describes the monthly number of sales of shampoo over a 3 year period. The units are a sales count and there are 36 observations.

Content

Contain the sales of shampoo for 36 months time

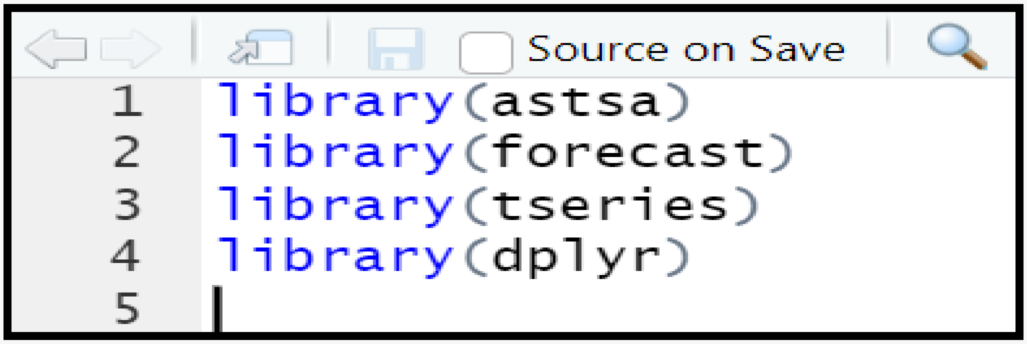
Acknowledgements

The original dataset is credited to Makridakis, Wheelwright, and Hyndman (1998).

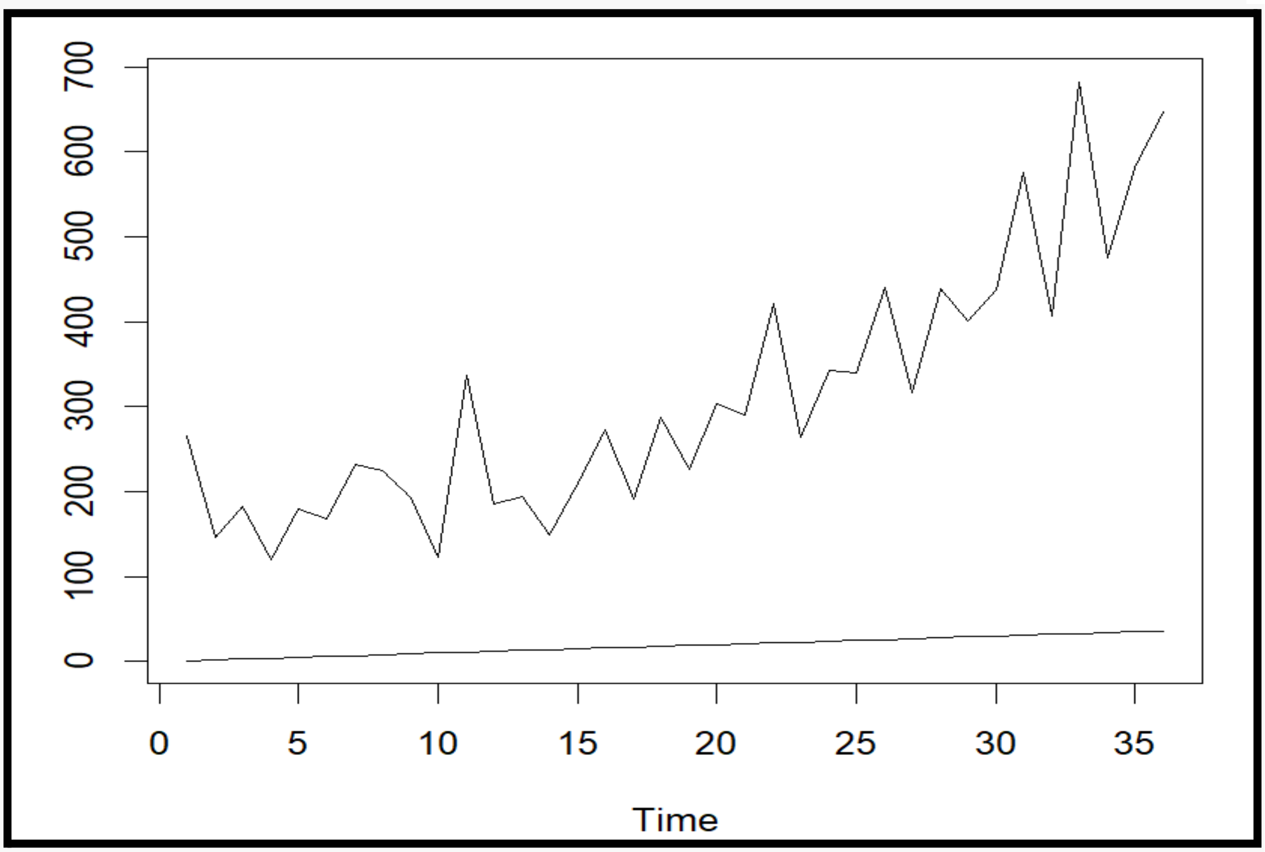
Link -: https://www.kaggle.com/datasets/redwankarimsony/shampoo-saled-dataset

Implementation:

```
#Installing required libraries
```




```
# Load your time series data

Temp_data <- read.csv("C:/Users/ASUS/Favorites/Downloads/archive (6)/shampoo_sales.csv")

View(Temp_data)

attach(Temp_data)

ts.plot(Temp_data)
```



 # Check the structure of the dataset

str(Temp_data)

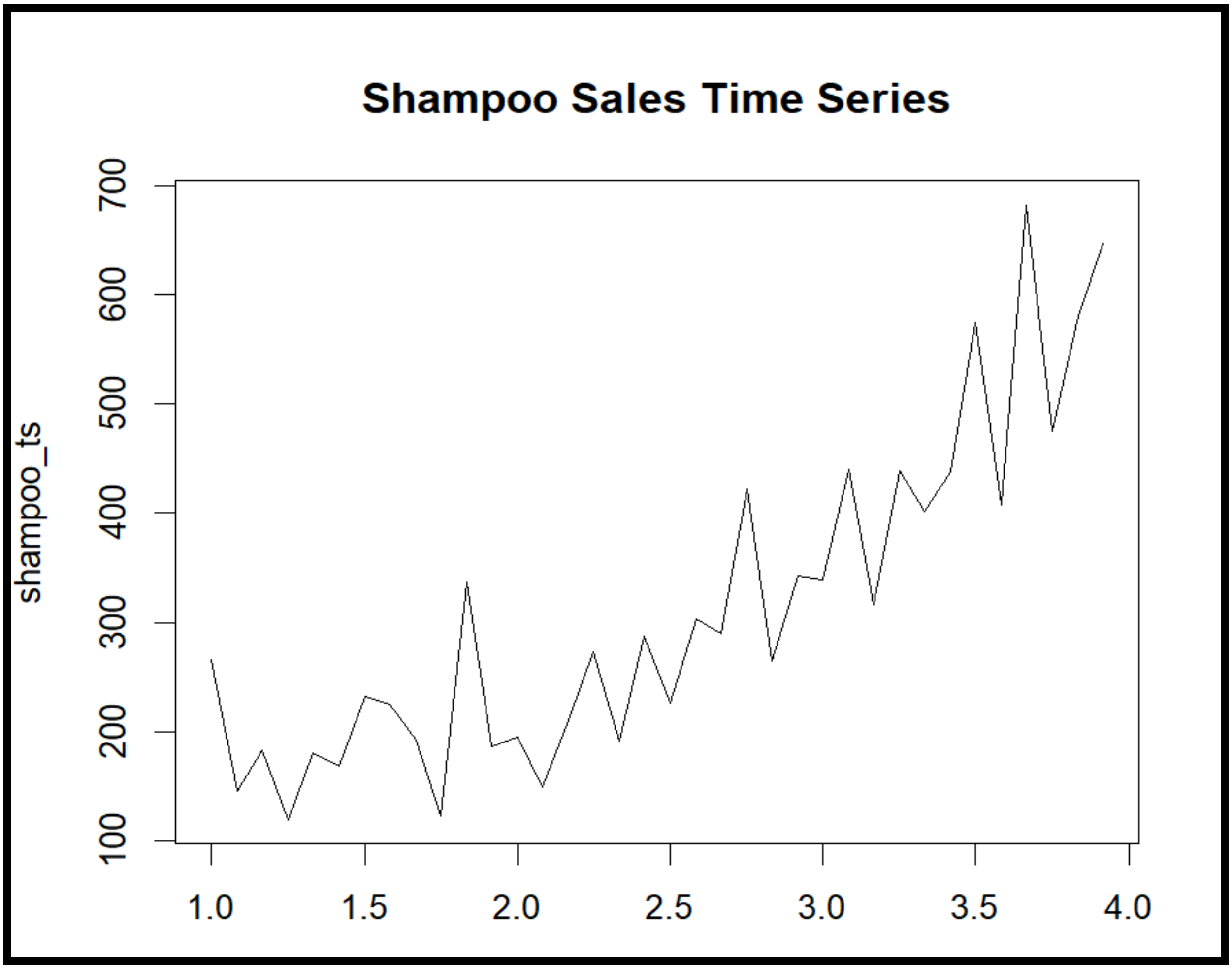
```
> str(Temp_data)
'data.frame':   36 obs. of  2 variables:
 $ Month: chr  "1-01" "1-02" "1-03" "1-04" ...
 $ Sales: num  266 146 183 119 180 ...
```

 # Create a time series object

shampoo_ts <- ts(Temp_data$Sales, frequency = 12)

 # Plot the time series

ts.plot(shampoo_ts, main = "Shampoo Sales Time Series")



 # Fitting an ARIMA model

fit <- auto.arima(shampoo_ts_new_diff, seasonal = FALSE)

fit

```
Series: shampoo_ts_new_diff
ARIMA(1,0,1) with zero mean

Coefficients:
          ar1      ma1
      -0.5617  -0.5726
s.e.   0.1814   0.1636

sigma^2 = 11769:  log likelihood = -140.05
AIC=286.1   AICc=287.36   BIC=289.5
```

ARIMA Model:

The ARIMA model fitted to the stationary time series is ARIMA(1,0,1).

Coefficient estimates:

AR(1) coefficient (ar1): -0.5617

MA(1) coefficient (ma1): -0.5726

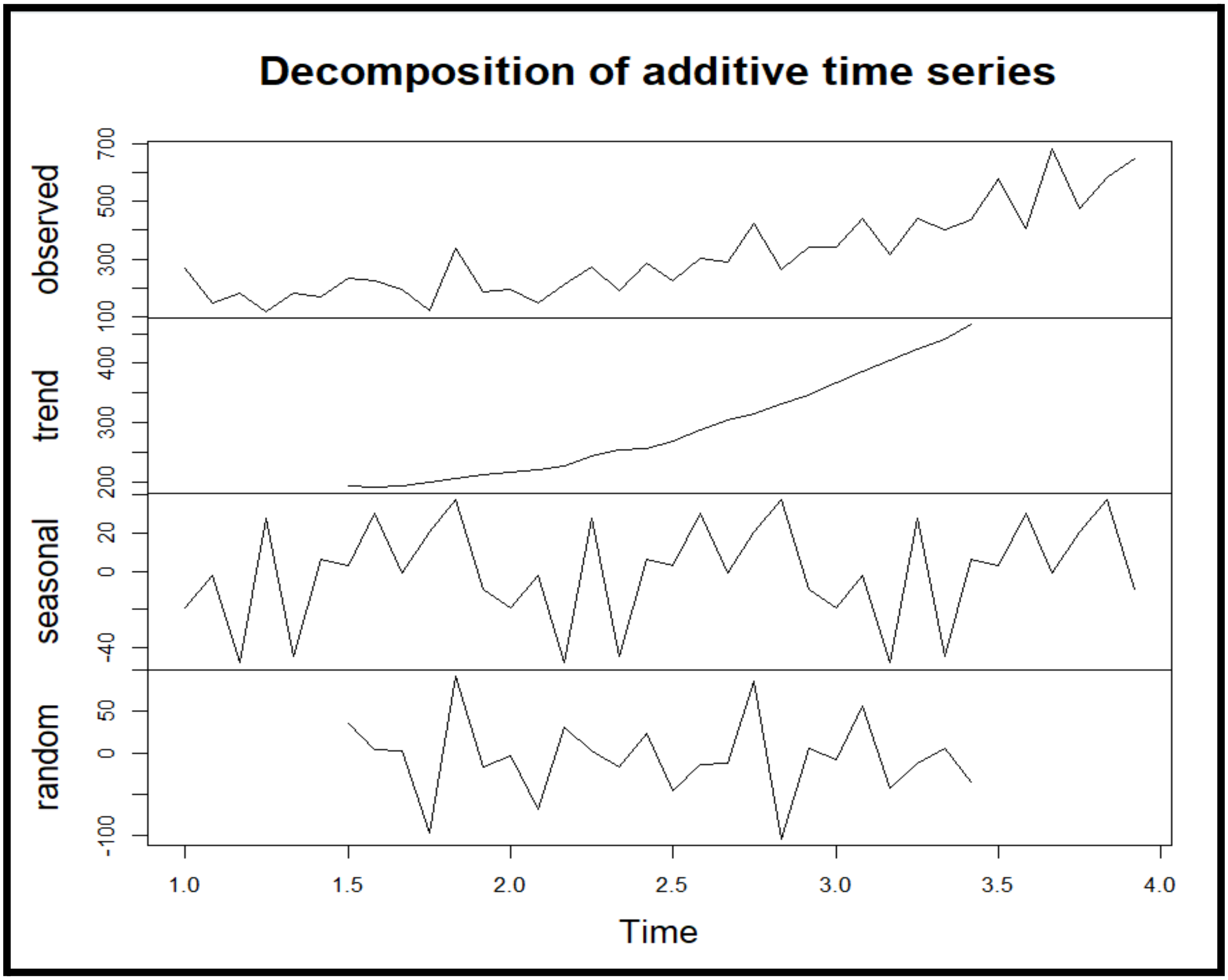
Standard errors (s.e.) are also provided for the coefficients.

The variance of the residuals (sigma^2) is 11,769, and the log-likelihood is -140.05.

```
# Decomposing original time series

ddata <- decompose(shampoo_ts, "additive")

plot(ddata)

# Removing trend and seasonality

shampoo_ts_new <- diff(shampoo_ts, lag = 12)  # Eliminate seasonality
```

ts.plot(shampoo_ts_new)

shampoo_ts_new_diff <- diff(shampoo_ts_new) # Eliminate trendts.plot(shampoo_ts_new_diff)



 # Dickey-Fuller test for Stationary condition

adf.test(shampoo_ts_new_diff)

ACF and PACF plots for stationary time series

acf(shampoo_ts_new_diff, ylim = c(-1, 1))

pacf(shampoo_ts_new_diff, ylim = c(-1, 1))

```
        Augmented Dickey-Fuller Test

data:  shampoo_ts_new_diff
Dickey-Fuller = -5.1555, Lag order = 2, p-value = 0.01
alternative hypothesis: stationary
```

Dickey-Fuller Test:

The Dickey-Fuller test was applied to the differenced time series (shampoo_ts_new_diff) to assess its stationarity.

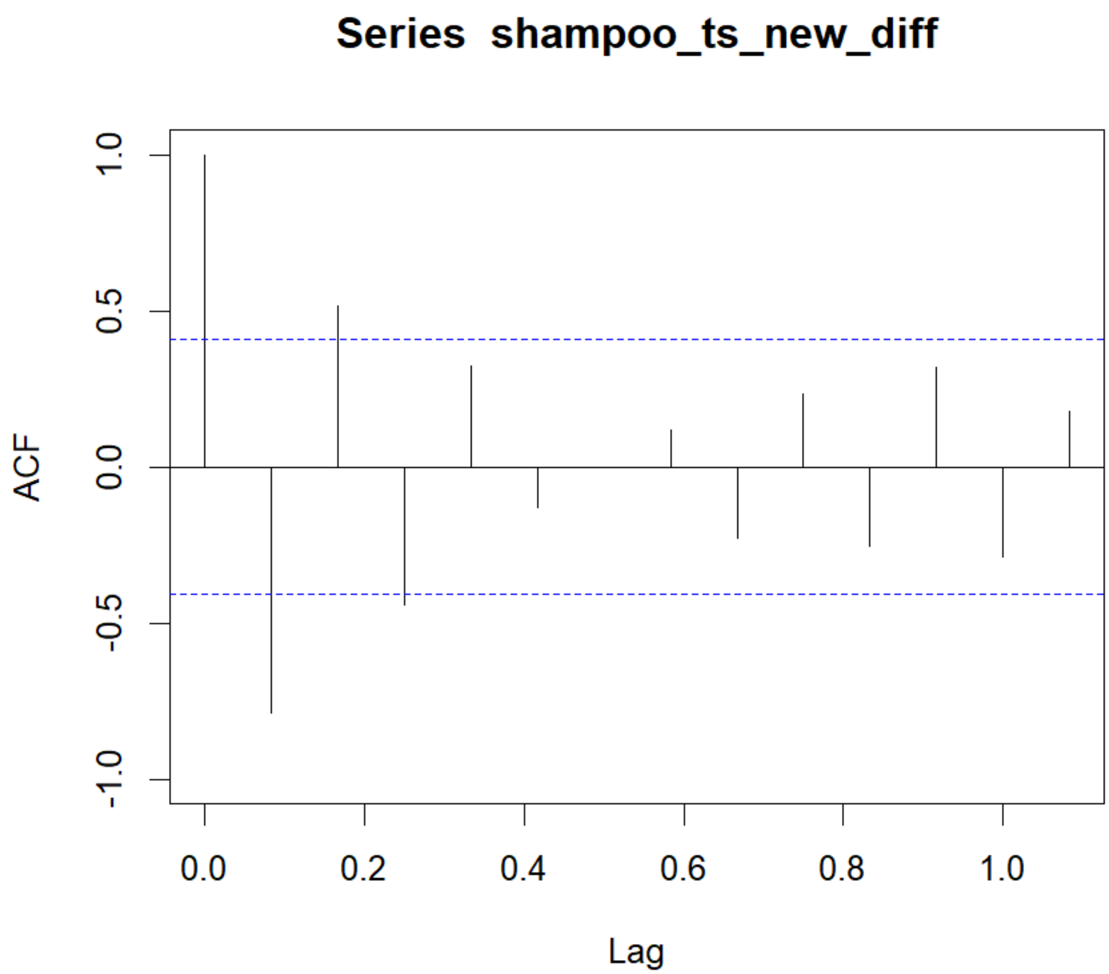
The test statistic is -5.1555, and the p-value is 0.01.

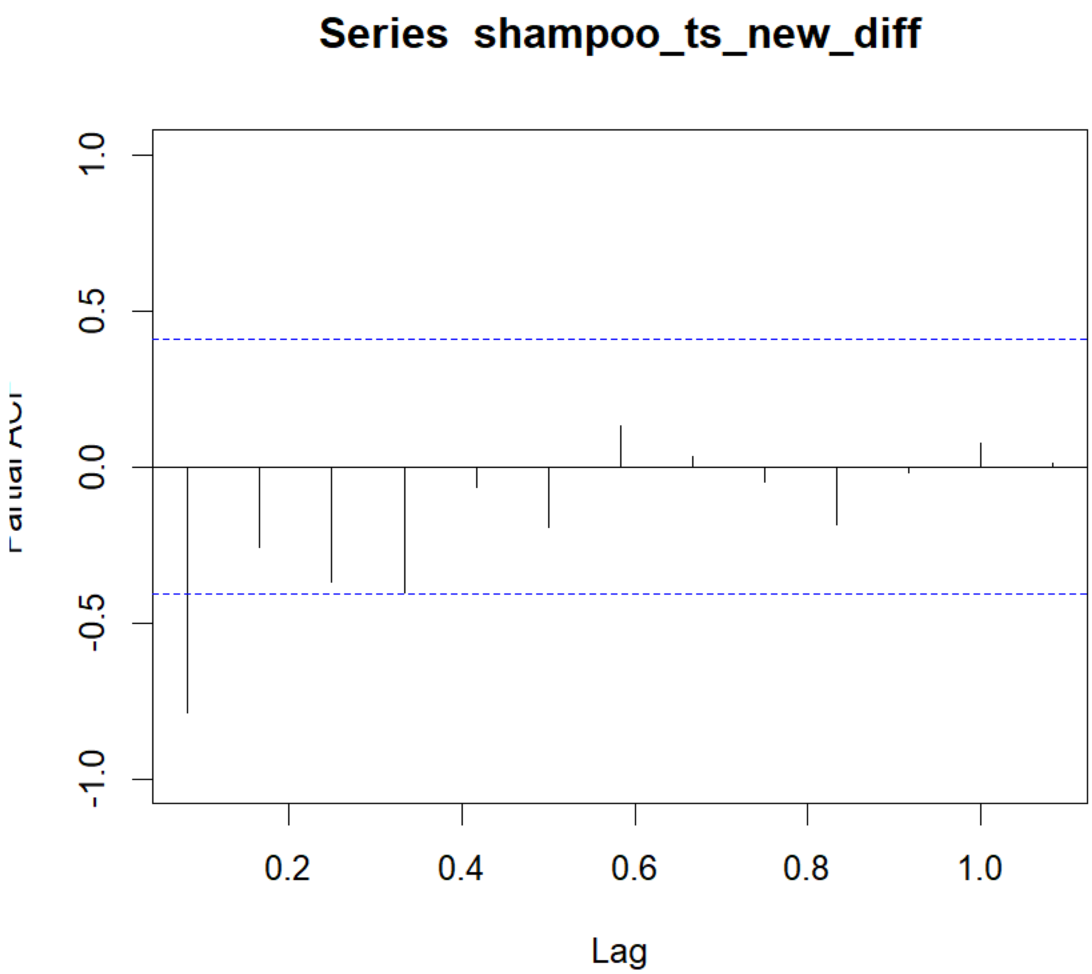
Interpretation: The p-value is less than the significance level (e.g., 0.05), indicating that the null hypothesis of non-stationarity is rejected. Therefore, the time series is stationary.

Fitting an ARIMA model

fit <- auto.arima(shampoo_ts_new_diff, seasonal = FALSE)

fit


Series shampoo_ts_new_diff
ACF
1.0
0.5
0.0
-0.5
-1.0
0.0
0.2
0.4
0.6
0.8
1.0
Lag

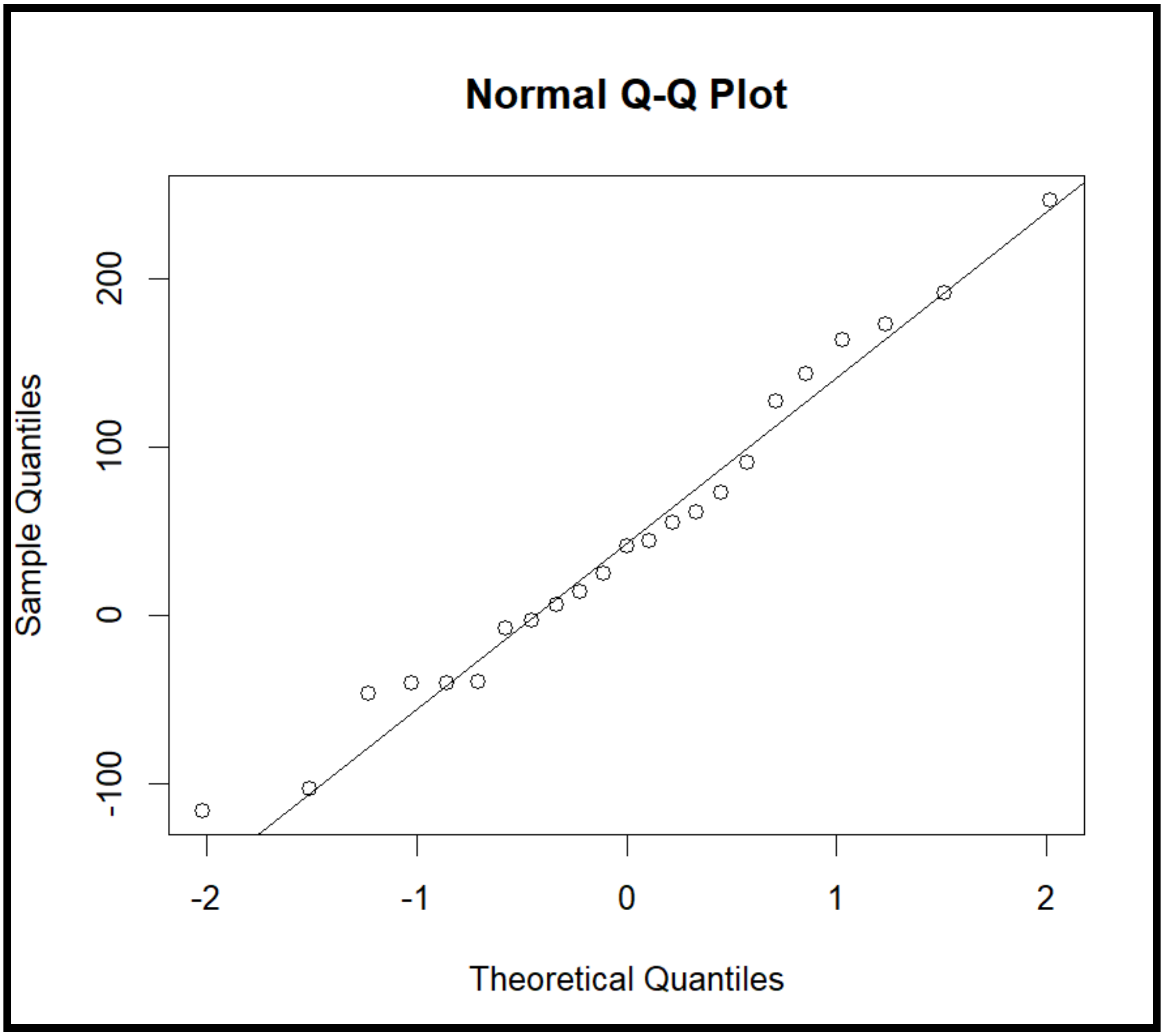



```
Series: shampoo_ts_new_diff
ARIMA(1,0,1) with zero mean

Coefficients:
          ar1      ma1
      -0.5617  -0.5726
s.e.   0.1814   0.1636

sigma^2 = 11769:  log likelihood = -140.05
AIC=286.1   AICc=287.36   BIC=289.5
```

```
# Portmanteau Ljung-Box test to check for correlation between residuals

Box.test(fit$residuals, lag = 10, fitdf = 1)
```

```
        Box-Pierce test

data:  fit$residuals
X-squared = 8.6419, df = 9, p-value = 0.471
```

```
# Normality test for residuals

qqnorm(fit$residuals)

qqline(fit$residuals)

shapiro_test_result <- shapiro.test(fit$residuals)

shapiro_test_result
```



 # Plot residuals

plot(fit$residuals, main = "Residuals from ARIMA Model", ylab = "Residuals")

 # Plot a histogram of residuals with a bell-shaped curve

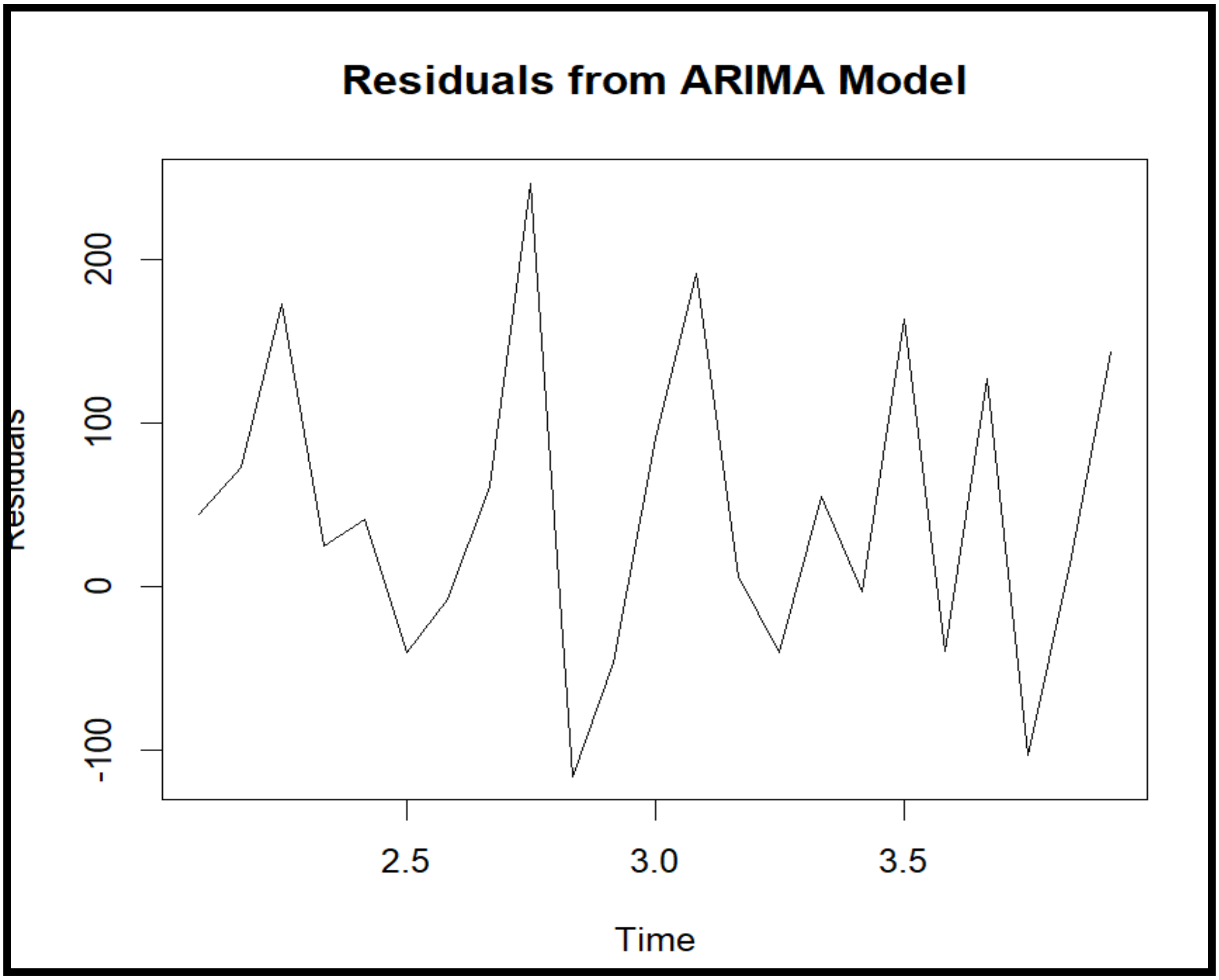
hist(fit$residuals, main = "Histogram of Residuals", xlab = "Residuals", col = "darkblue")

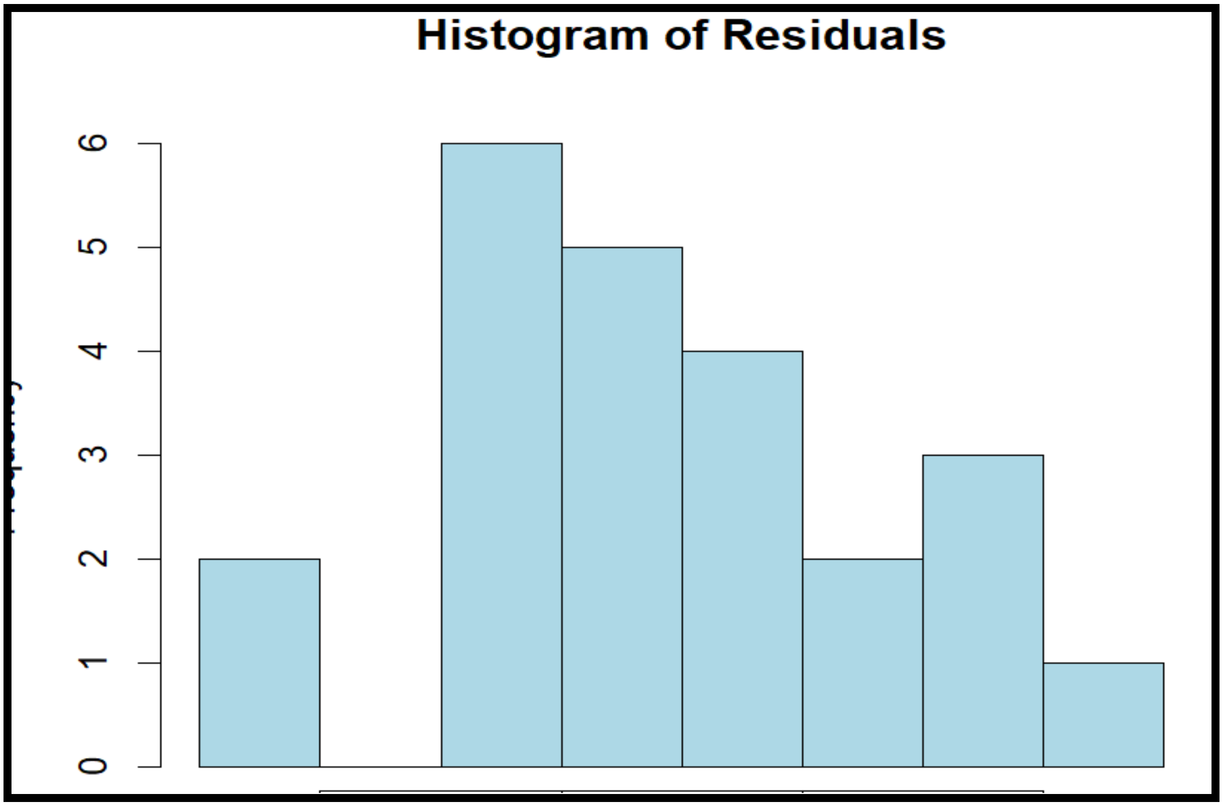
 # Add a red curve representing a normal distribution

mu <- mean(fit$residuals)

sigma <- sd(fit$residuals)

curve(dnorm(x, mean = mu, sd = sigma), col = "red", lwd = 2, add = TRUE)





 # Forecasting

forecast_result <- forecast(fit, h = 24) # Forecasting for the next 24 time periods

forecast_result

plot(forecast_result)

observations

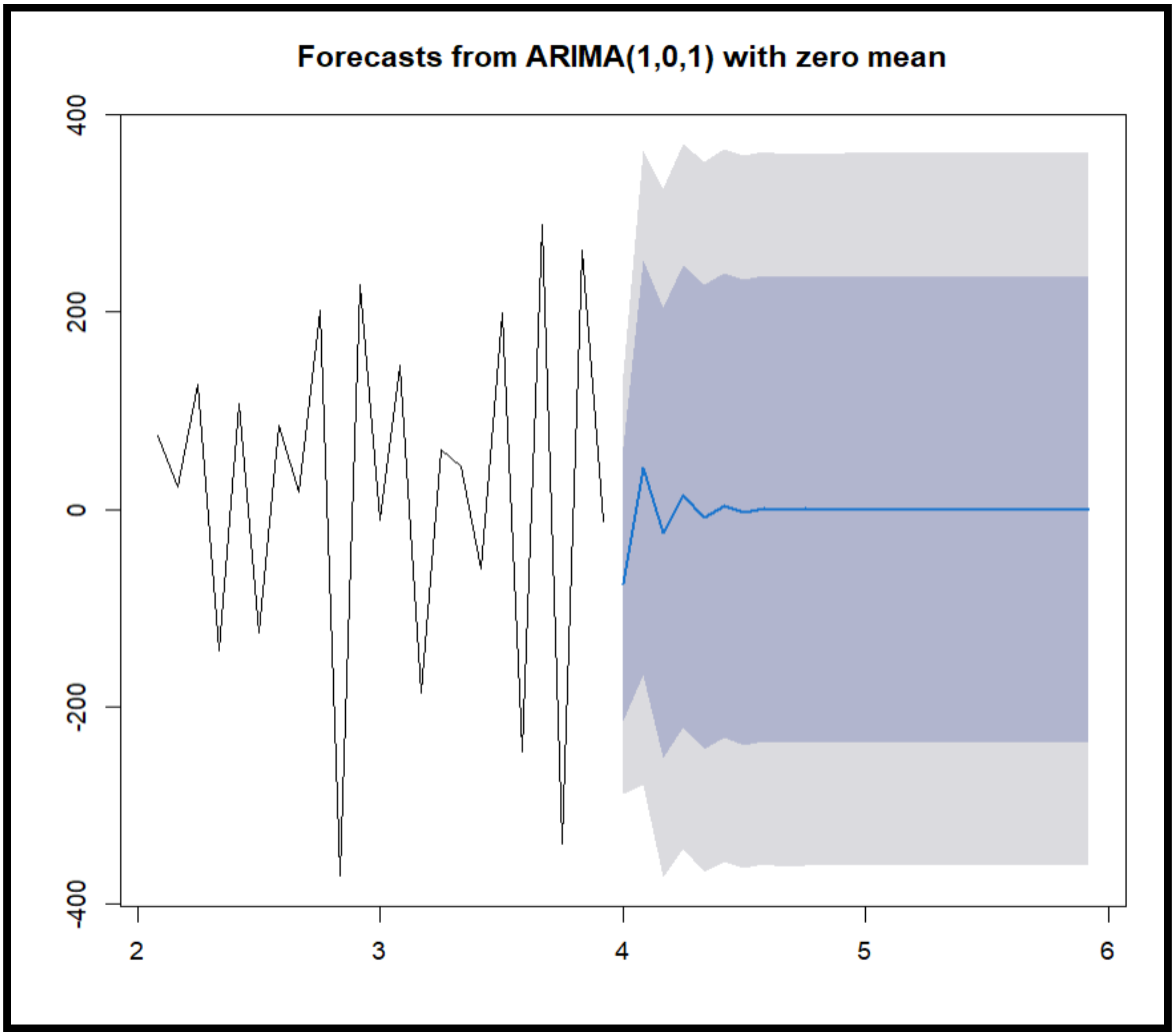
Forecasts:

The forecast values for the next 24 months (until Dec 2024) are provided.

The forecast includes point forecasts as well as lower and upper prediction intervals at different confidence levels (e.g., 80% and 95%).

Interpretation: These forecasts represent the expected values of shampoo sales for each month in the forecast period, along with the associated uncertainty intervals.

```
> forecast_result
       Point Forecast     Lo 80     Hi 80     Lo 95    Hi 95
Jan 4   -7.525241e+01 -214.2793  63.77449 -287.8757 137.3709
Feb 4    4.226761e+01 -167.9565 252.49170 -279.2424 363.7776
Mar 4   -2.374078e+01 -251.8610 204.37946 -372.6206 325.1390
Apr 4    1.333467e+01 -220.1470 246.81636 -343.7448 370.4141
May 4   -7.489789e+00 -242.6376 227.65798 -367.1173 352.1377
Jun 4    4.206848e+00 -231.4641 239.87779 -356.2208 364.6344
Jul 4   -2.362894e+00 -238.1986 233.47286 -363.0426 358.3168
Aug 4    1.327185e+00 -234.5605 237.21491 -359.4320 362.0863
Sep 4   -7.454504e-01 -236.6496 235.15867 -361.5297 360.0388
Oct 4    4.187030e-01 -235.4906 236.32799 -360.3734 361.2108
Nov 4   -2.351762e-01 -236.1461 235.67574 -361.0298 360.5594
Dec 4    1.320932e-01 -235.7793 236.04353 -360.6633 360.9275
Jan 5   -7.419385e-02 -235.9858 235.83740 -360.8698 360.7215
Feb 5    4.167304e-02 -235.8700 235.95332 -360.7541 360.8374
Mar 5   -2.340682e-02 -235.9351 235.88826 -360.8192 360.7723
Apr 5    1.314709e-02 -235.8985 235.92482 -360.7826 360.8089
May 5   -7.384431e-03 -235.9191 235.90429 -360.8031 360.7884
Jun 5    4.147671e-03 -235.9075 235.91582 -360.7916 360.7999
Jul 5   -2.329655e-03 -235.9140 235.90934 -360.7981 360.7934
Aug 5    1.308516e-03 -235.9104 235.91298 -360.7945 360.7971
Sep 5   -7.349642e-04 -235.9124 235.91094 -360.7965 360.7950
Oct 5    4.128131e-04 -235.9113 235.91208 -360.7954 360.7962
Nov 5   -2.318680e-04 -235.9119 235.91144 -360.7960 360.7955
Dec 5    1.302351e-04 -235.9115 235.91180 -360.7956 360.7959
> plot(forecast_result)
```



Observations -:

Overall, the ARIMA model you've fitted to the stationary time series is expected to provide reasonable forecasts for shampoo sales. The Dickey-Fuller test confirmed that the time series is stationary, which is a prerequisite for ARIMA modeling. The coefficients of the ARIMA model describe the relationship between the current and past values of the time series. The provided forecasts offer insights into future sales trends and associated uncertainty.

This analysis allows you to make informed decisions and predictions regarding shampoo sales based on the available data and the ARIMA model.