PROCESSING TASKS BY R

Import Library and create dataset

```
#import library
library(readxl)
library(ggplot2)
library(tidyverse)  #for dataframe manipulation
library(forcats) #for handling factors
library(scales)
library(readxl)
library(dplyr)
#TASK 1
#data collection and input
data <- read_excel('K204141924.xlsx')
```

Output:

	Date	Equity	NAGR	EBT	ROAA	Investment
1	2012-03-31	82273	13.47	1774	0.44	0.0017268743
2	2012-06-30	80213	10.71	1837	0.65	0.0019744961
3	2012-09-30	80390	-29.31	1224	0.40	0.0006600495
4	2012-12-31	85924	-19.76	5792	2.28	0.0066500695
5	2013-03-31	83042	-23.77	275	0.10	0.0031259333
6	2013-06-30	83947	-8.05	4383	1.53	0.0133659739
7	2013-09-30	89667	58.53	2315	0.69	0.0062237623
8	2013-12-31	93873	61.60	4546	1.14	0.0102129692
9	2014-03-31	91615	91.91	3297	0.95	0.0066681809
10	2014-06-30	95904	20.45	4941	1.12	0.0049773173

The dataset consists of information on 4 feature variables: Equity, NAGR, EBT, ROAA, and a target variable called Investment. The data is collected over a span of 44 quarters, including 32 quarters from 2012 to 2019 (pre-COVID) and 12 quarters from 2020 to 2022 (post-COVID). Each quarter in the dataset records the values of the feature variables and the corresponding value of the Investment variable.

Descriptive statistics

Descriptive statistics

```
# Chuyển đổi cột "Date" sang định dạng ngày-tháng-năm
data$Date <- as.Date(data$Date, format = "%d/%m/%Y")
# Lọc dữ liệu chỉ lấy từ năm 2020 về trước và đặt tên tập dữ liệu là "before"
before <- data[data$Date <= as.Date("2019-12-31"), ]
# Lọc dữ liệu chỉ lấy sau năm 2020 và đặt tên tập dữ liệu là "after"
after <- data[data$Date > as.Date("2019-12-31") & data$Date <= as.Date("2022-12-31"), ]
# Tính toán số liệu mô tả cho toàn bộ kỳ
summary_all <- summary(data[, c("Equity", "NAGR", "EBT", "ROAA", "Investment")])
# Tính toán số liệu mô tả cho toàn bộ kỳ
```

```
summary_before <- summary(before[, c("Equity", "NAGR", "EBT", "ROAA", "Investment")])
# Tính toán số liệu mô tả cho toàn bộ kỳ
summary_after <- summary(after[, c("Equity", "NAGR", "EBT", "ROAA", "Investment")])
# In kết quả
print("Toàn bộ kỳ:")
print(summary_all)
print("Kỳ trước:")
print(summary_before)
print("Kỳ sau:")
print(summary_after)
```

Output:

```
[1] "Toàn bộ kỳ:"
> print(summary_all)
     Equity             NAGR              EBT             ROAA           Investment
 Min.   :  80213   Min.   :-29.310   Min.   :  275   Min.   :-0.0400   Min.   :0.0006601
 1st Qu.: 106410   1st Qu.:  7.117   1st Qu.: 4907   1st Qu.: 0.0300   1st Qu.:0.0063129
 Median : 810595   Median : 34.245   Median : 9588   Median : 0.2750   Median :0.0133212
 Mean   :1381662   Mean   : 35.721   Mean   :14520   Mean   : 0.7186   Mean   :0.0167584
 3rd Qu.:2124482   3rd Qu.: 59.583   3rd Qu.:17860   3rd Qu.: 1.1775   3rd Qu.:0.0194428
 Max.   :4570105   Max.   :114.160   Max.   :88789   Max.   : 3.0100   Max.   :0.0915729
> print("Kỳ trước:")
[1] "Kỳ trước:"
> print(summary_before)
     Equity             NAGR              EBT             ROAA           Investment
 Min.   :  80213   Min.   :-29.310   Min.   :  275   Min.   :0.0300   Min.   :0.0006601
 1st Qu.:  93309   1st Qu.:  6.697   1st Qu.: 4505   1st Qu.:0.2400   1st Qu.:0.0063129
 Median : 293587   Median : 27.120   Median : 7950   Median :0.6700   Median :0.0113767
 Mean   : 625706   Mean   : 32.733   Mean   :12337   Mean   :0.9828   Mean   :0.0180497
 3rd Qu.:1005907   3rd Qu.: 50.542   3rd Qu.:17335   3rd Qu.:1.6075   3rd Qu.:0.0225147
 Max.   :2053324   Max.   :114.160   Max.   :67278   Max.   :3.0100   Max.   :0.0915729
> print("Kỳ sau:")
[1] "Kỳ sau:"
> print(summary_after)
     Equity             NAGR              EBT             ROAA            Investment
 Min.   :2071687   Min.   : -8.71   Min.   :  763   Min.   :-0.04000   Min.   :0.001225
 1st Qu.:2651258   1st Qu.: 16.93   1st Qu.: 6877   1st Qu.: 0.00750   1st Qu.:0.008660
 Median :3746486   Median : 48.53   Median :11222   Median : 0.01000   Median :0.014131
 Mean   :3397544   Mean   : 43.69   Mean   :20343   Mean   : 0.01417   Mean   :0.013315
 3rd Qu.:3866358   3rd Qu.: 72.50   3rd Qu.:23256   3rd Qu.: 0.02250   3rd Qu.:0.015754
 Max.   :4570105   Max.   :104.35   Max.   :88789   Max.   : 0.08000   Max.   :0.028754
>
```

Based on the results of the descriptive statistics for the entire period, previous period, and subsequent period, we can draw the following observations:

• Equity:

The value of Equity increases gradually from the previous period to the entire period, indicating the growth of equity during this time.

Equity ranges from a minimum of 80,213 to a maximum of 4,570,105 over the entire period.

• NAGR (Net Income Growth Rate):

The average Net Income Growth Rate (NAGR) over the entire period is approximately 35.7%.

NAGR ranges from -29.31% to 114.16%, indicating significant fluctuations in income growth.

• EBT (Earnings Before Tax):

Earnings Before Tax (EBT) increases from 275 to 88,789 over the entire period.

On average, EBT is 14,520.

• ROAA (Return on Average Assets):

The average Return on Average Assets (ROAA) over the entire period is 0.7186, with a range from 0.03 to 3.01.

ROAA reflects the profitability of assets for the business.

• Investment:

The average value of Investment is 0.0168.

Investment ranges from 0.0006601 to 0.0915729 over the entire period.

These observations provide insights into the trends and variations in Equity, Net Income Growth Rate, Earnings Before Tax, Return on Average Assets, and Investment over the analyzed periods.

Possible impact of the Covid-19 pandemic

Equity (Shareholders' Equity): In the period after Covid, the average value of Equity decreased compared to the pre-Covid period, indicating that companies may face difficulties in recovering capital and assets after the pandemic.

NAGR (Net Annual Growth Rate): The growth rate of income from business operations decreased in the post-Covid period, indicating that the production and business capabilities of companies may be affected by supply disruptions and reduced demaand.

EBT (Earnings Before Tax): Pre-tax earnings increased in the post-Covid period, possibly due to economic stimulus measures and financial support from the government to mitigate the impact of the pandemic.

ROAA (Return on Average Assets): The return on average assets decreased in the post-Covid period, indicating that companies may face difficulties in generating profits from their investment assets.

Investment: The average value of the Investment variable decreased in the post-Covid period, indicating that companies may reduce investment and consumption in the context of economic recession and uncertainty.

Overall, the Covid-19 pandemic has had a negative impact on the financial indicators and business operations of companies, causing disruptions and hampering economic recovery.

Visualization

```
# Vẽ biểu đồ whisker plot cho cột Investment
boxplot(before$Investment, main = "Whisker plot for (before) Investment", ylab = "Investment")
```

Output:

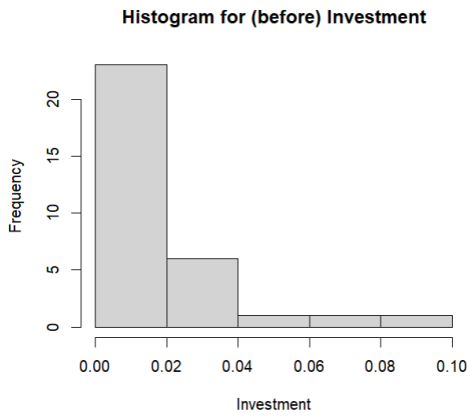
The median value of the "Investment" variable is approximately 0.0114. The data has a first quartile (Q1) value of approximately 0.0063 and a third quartile (Q3) value of approximately 0.0225.

There are some outliers in the data, with one outlier value approximately around 0.09, which is significantly higher than the other values. The range between the lowest and highest values in the data is approximately from 0.00 to 0.05.

Overall, the boxplot indicates a significant variability in the distribution of the "Investment" variable in this data, with a notable outlier. The majority of values are concentrated within the range from Q1 to Q3, while the outlier lies further outside this range.

```
# Vẽ biểu đồ histogram cho cột Investment
hist(before$Investment, main = "Histogram for (before) Investment", xlab = "Investment", ylab = "Frequency")
```

Output:

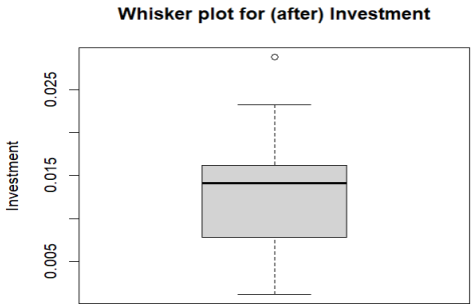


The histogram shows an increasing frequency from the value range of 0.00 to 0.01 and a decreasing frequency from the range of 0.02 to 0.04.

This indicates that the distribution of the "Investment" variable in the data is variable and uneven. There is a relatively high concentration of values within the range of 0.00 to 0.02, while there are fewer values in the range of 0.02 to 0.1.

```
# Vẽ biểu đồ whisker plot cho cột Investment
boxplot(after$Investment, main = "Whisker plot for (after) Investment", ylab = "Investment")
```

Output:



The range of the variable "Investment" in this dataset is from 0.001 to 0.029, with a minimum value of 0.001 and a maximum value of 0.029.

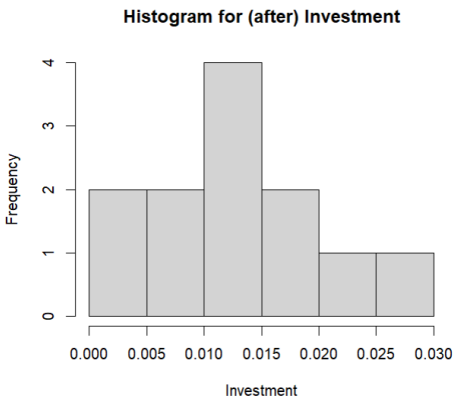
The median of the "Investment" variable falls within the range of values from 0.014 to 0.015.

The box of the boxplot is relatively narrow, indicating that the data has a moderate level of dispersion. The range between the first quartile (Q1) and the third quartile (Q3) falls within the range of 0.008 to 0.016.

The boxplot has one outlier point that lies outside the whiskers of the boxplot.

```
# Vẽ biểu đồ histogram cho cột Investment
hist(after$Investment, main = "Histogram for (after) Investment", xlab = "Investment", ylab = "Frequency")
```

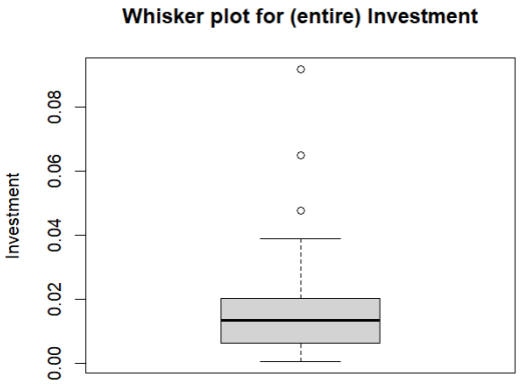
Output:



The histogram shows an increasing frequency from the range of values 0.001 to 0.015, followed by a decreasing frequency from the range of 0.015 to 0.030. This indicates that the distribution of the "Investment" variable in this dataset is variable and uneven. There is a relatively high concentration of values within the range of 0.001 to 0.015, while there are fewer values in the range of 0.020 to 0.030.

```
#Vẽ biểu đồ whisker plot cho cột Investment
boxplot(data$Investment, main = "Whisker plot for (entire) Investment", ylab = "Investment")
```

Output:



The boxplot indicates that the range of the Investment variable is between approximately 0.001 and 0.1.

The median of the Investment variable falls between the values of approximately 0.01 and 0.02. This suggests that the average value of the Investment variable is concentrated around this range.

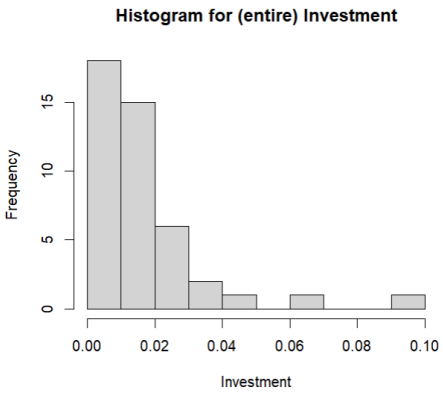
The box of the boxplot is relatively narrow, indicating that the data has a moderate dispersion. The interquartile range (IQR), which represents the range between the first quartile (Q1) and the third quartile (Q3), is between approximately 0.005 and 0.027.

The boxplot also shows the presence of some outliers, which are data points located outside the whiskers of the boxplot. This suggests the existence of some unusual values in the Investment variable.

In summary, the Investment variable has a relatively concentrated distribution with several notable outliers.

```
# Vẽ biểu đồ histogram cho cột Investment
hist(data$Investment, main = "Histogram for (entire) Investment", xlab = "Investment", ylab = "Frequency")
```

Output:



Based on the histogram, we can observe that the frequency increases gradually from the range of values from 0.00 to 0.02, and then decreases from the range of 0.02 to 0.05. There are no values in the ranges from 0.05 to 0.06 and 0.07 to 0.09.

This indicates that the distribution of the Investment variable is variable and uneven. There is a relatively high concentration of values in the range from 0.00 to 0.02, while there are fewer values in the range from 0.02 to 0.10.

Buiding model

Model 1

```
# Lọc dữ liệu chỉ với các biến và target của đề tài
df <- select(data, Equity, NAGR, EBT, ROAA, Investment)
# Thực hiện hồi quy bội với tất cả các biến riêng lẻ
model1 <- lm(Investment ~ ., data = df)
# Xem kết quả của mô hình hồi quy
summary(model1)
```

Output:

```
Residuals:
      Min        1Q    Median        3Q       Max
-0.018271 -0.011046 -0.002922  0.003168  0.075691

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.789e-02  6.366e-03   2.810   0.0077 **
Equity      -2.718e-09  2.997e-09  -0.907   0.3701
NAGR         6.719e-05  7.829e-05   0.858   0.3961
EBT          2.800e-07  1.849e-07   1.514   0.1380
ROAA        -5.345e-03  4.745e-03  -1.126   0.2669
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01761 on 39 degrees of freedom
Multiple R-squared:  0.07107,   Adjusted R-squared:  -0.02421
F-statistic: 0.7459 on 4 and 39 DF,  p-value: 0.5667
```

The estimation results of the regression model (Model 1) can be interpreted as follows:

Residuals: The residuals represent the difference between the actual values and the predicted values of the dependent variable (Investment). The residual values range from -0.018271 to 0.075691.

Coefficients:

- Intercept: The intercept coefficient has a value of 1.789e-02 with a standard deviation of 6.366e-03. This represents the predicted value of the dependent variable (Investment) when all other independent variables are zero.
- Equity: The estimated coefficient for Equity is -2.718e-09 with a standard deviation of 2.997e-09. However, this coefficient is not statistically significant (p-value > 0.05), indicating that there is no significant relationship between Equity and Investment.

- NAGR: The estimated coefficient for NAGR is 6.719e-05 with a standard deviation of 7.829e-05. Similarly, this coefficient is not statistically significant (p-value > 0.05), indicating no significant relationship between NAGR and Investment.
- EBT: The estimated coefficient for EBT is 2.800e-07 with a standard deviation of 1.849e-07. This coefficient is also not statistically significant (p-value > 0.05), indicating no significant relationship between EBT and Investment.
- ROAA: The estimated coefficient for ROAA is -5.345e-03 with a standard deviation of 4.745e-03. Again, this coefficient is not statistically significant (p-value > 0.05), indicating no significant relationship between ROAA and Investment.

Significance: The p-value of the F-statistic is 0.5667, exceeding the significance threshold of 0.1. This indicates that the model does not achieve statistical significance, meaning there is insufficient evidence to accept the model as meaningful in explaining the dependent variable (Investment).

R-squared: The Multiple R-squared value is 0.07107, indicating that only 7.107% of the variation in the dependent variable (Investment) is explained by the independent variables in the model. However, when adjusted for the number of independent variables in the model, the Adjusted R-squared value is -0.02421, indicating that the model poorly explains the variability of the dependent variable.

Overall, the estimation results of the regression model (Model 1) indicate that there is no significant relationship between the independent variables (Equity, NAGR, EBT, ROAA) and the dependent variable (Investment). The model does not achieve statistical significance and poorly explains the variability of the dependent variable.

Model 2

```
# Creating the Covid-19 dummy variable
data$CovidDummy <- ifelse(data$Date >= as.Date("2019-12-31"), 1, 0)
# Fitting Model 2
model2 <- lm(Investment ~ Equity + NAGR + EBT + ROAA + CovidDummy + Equity:CovidDummy + NAGR:CovidDummy + EBT:CovidDummy + ROAA:CovidDummy, data = data)
# Printing the summary of Model 2
summary(model2)
```

Output:

```
Residuals:
      Min        1Q    Median        3Q       Max
-0.026262 -0.006891 -0.002706  0.003128  0.070928

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)        7.756e-03  7.403e-03   1.048    0.302
Equity             1.090e-08  6.629e-09   1.644    0.109
NAGR               6.522e-05  8.749e-05   0.746    0.461
EBT                3.978e-07  2.801e-07   1.420    0.165
ROAA              -2.773e-03  5.425e-03  -0.511    0.613
CovidDummy         4.939e-03  2.479e-02   0.199    0.843
Equity:CovidDummy -1.067e-08  1.012e-08  -1.055    0.299
NAGR:CovidDummy   -9.731e-05  1.769e-04  -0.550    0.586
EBT:CovidDummy    -3.013e-07  3.673e-07  -0.820    0.418
ROAA:CovidDummy   -2.852e-02  7.884e-02  -0.362    0.720

Residual standard error: 0.01672 on 34 degrees of freedom
Multiple R-squared:  0.2701,    Adjusted R-squared:  0.07694
F-statistic: 1.398 on 9 and 34 DF,  p-value: 0.2276
```

Residuals: The residuals represent the difference between the actual values and the predicted values of the dependent variable (Investment). The residuals range from -0.026262 to 0.070928.

Coefficients:

- Intercept: The estimated coefficient for the intercept is 7.756e-03 with a standard deviation of 7.403e-03. This represents the predicted value of the dependent variable (Investment) when all independent variables are zero.
- Equity: The estimated coefficient for Equity is 1.090e-08 with a standard deviation of 6.629e-09. However, this coefficient is not statistically significant (p-value > 0.05), indicating that there is no significant relationship between Equity and Investment.
- NAGR: The estimated coefficient for NAGR is 6.522e-05 with a standard deviation of 8.749e-05. Similarly, this coefficient is not statistically significant (p-value > 0.05), suggesting no significant relationship between NAGR and Investment.
- EBT: The estimated coefficient for EBT is 3.978e-07 with a standard deviation of 2.801e-07. Again, this coefficient is not statistically significant (p-value > 0.05), indicating no significant relationship between EBT and Investment.
- ROAA: The estimated coefficient for ROAA is -2.773e-03 with a standard deviation of 5.425e-03. This coefficient is also not statistically significant (p-value > 0.05), suggesting no significant relationship between ROAA and Investment.
- CovidDummy: The estimated coefficient for the CovidDummy variable is 4.939e-03 with a standard deviation of 2.479e-02. This coefficient is not statistically significant (p-value > 0.05), indicating no significant relationship between the CovidDummy variable and Investment.
- Equity:CovidDummy: The estimated coefficient for the interaction between Equity and CovidDummy is -1.067e-08 with a standard deviation of 1.012e-08. This coefficient is not statistically significant (p-value > 0.05), suggesting no significant interaction effect between Equity and CovidDummy on Investment.

- NAGR:CovidDummy: The estimated coefficient for the interaction between NAGR and CovidDummy is -9.731e-05 with a standard deviation of 1.769e-04. This coefficient is also not statistically significant (p-value > 0.05), indicating no significant interaction effect between NAGR and CovidDummy on Investment.
- EBT:CovidDummy: The estimated coefficient for the interaction between EBT and CovidDummy is -3.013e-07 with a standard deviation of 3.673e-07. This coefficient is not statistically significant (p-value > 0.05), suggesting no significant interaction effect between EBT and CovidDummy on Investment.
- ROAA:CovidDummy: The estimated coefficient for the interaction between ROAA and CovidDummy is -2.852e-02 with a standard deviation of 7.884e-02. This coefficient is also not statistically significant (p-value > 0.05), indicating no significant interaction effect between ROAA and CovidDummy on Investment.

R-squared: The Multiple R-squared value is 0.2701, indicating that approximately 27.01% of the variation in the dependent variable (Investment) is explained by the independent variables in the model. However, when adjusting for the number of independent variables and interactions in the model, the Adjusted R-squared value is 0.07694, suggesting that the model does not explain the variation in the dependent variable well.

F-statistic: The F-statistic value is 1.398 with 9 and 34 degrees of freedom, and the p-value is 0.2276. Since this p-value is greater than the usual significance level (0.05), we do not have enough evidence to reject the null hypothesis of no relationship between the independent variables and the dependent variable in the model.

Overall, the model does not show significant correlations between the independent variables (Equity, NAGR, EBT, ROAA, and CovidDummy) and the dependent variable (Investment). The low R-squared value and low Adjusted R-squared value also indicate that the model is not suitable for explaining the variation in the dependent variable.,

Predict

```
# Predicting the value of AssignedTopic for all quarters using Model 1
predicted_values <- predict(model1, newdata = data)


# Printing the predicted values
print(predicted_values)
```

Output:

```
          1           2           3           4           5           6           7           8           9
0.016714570 0.015430019 0.013905236 0.005763194 0.015607868 0.010169184 0.018537234 0.016951742 0.019660009
         10          11          12          13          14          15          16          17          18
0.014399015 0.013054604 0.011003507 0.016737537 0.013875406 0.019088419 0.010457714 0.013889538 0.020930845
         19          20          21          22          23          24          25          26          27
0.022148957 0.022266815 0.017913657 0.023562408 0.016540595 0.015882094 0.019837438 0.020194005 0.020552633
         28          29          30          31          32          33          34          35          36
0.018052316 0.016107391 0.017044667 0.018163436 0.016396570 0.018813620 0.018507033 0.013041287 0.011751421
         37          38          39          40          41          42          43          44
0.009253658 0.013337929 0.015960659 0.017376373 0.014075840 0.024998370 0.016277317 0.033138788
```

The predicted values of the variable "AssignedTopic" range from 0.005 to 0.033 for different quarters in the dataset.

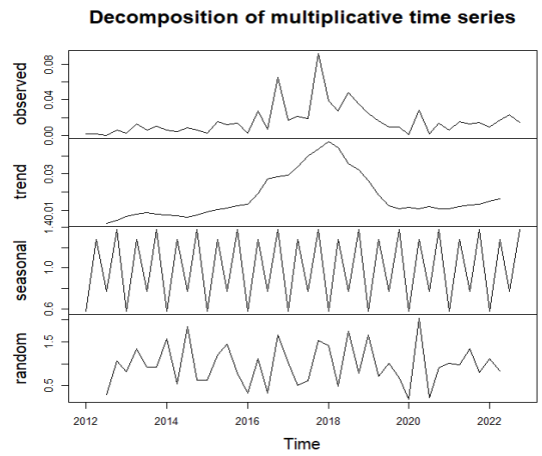
The predicted values do not exhibit a continuous increase or decrease across quarters but rather show uneven fluctuations.

It can be observed that the predicted values do not show a significant upward or downward trend across quarters.

ARIMA

```
#Tạo đối tượng phân rã dữ liệu
time_series <- ts(data$Investment, start = c(2012, 1), end = c(2022,4), frequency = 4)
# Phân rã chuỗi thời gian
decompose_results <- decompose(time_series, type = "multiplicative")
# Vẽ biểu đồ kết quả
plot(decompose_results))
```

Output:



From observing the chart, the Investment value of the company shows a generally increasing trend over time. This suggests that the company is likely increasing its investment in business activities or assets. However, in the year 2020, the Investment value of the company experienced a sharp decline, possibly due to the impact of the Covid-19 pandemic. However, it has gradually increased again since then. In addition, it can be seen that the Investment value is stable, cyclical.

```
# Kiểm định tính dừng của chuỗi thời gian Investment
result_ <- adf.test(data$Investment)
print(result_)
```

Output:

```
        Augmented Dickey-Fuller Test

data:  data$Investment
Dickey-Fuller = -1.3653, Lag order = 3, p-value = 0.825
alternative hypothesis: stationary
```

The Augmented Dickey-Fuller test result for the time series data$Investment indicates a p-value of 0.825, which exceeds the usual threshold of 0.05. Therefore, there is insufficient evidence to reject the null hypothesis that the series is non-stationary. Instead, the null hypothesis that the series is stationary cannot be rejected.

Hence, in this case, there is no need to perform transformations to achieve a stationary time series, and the previously determined ARIMA model with order (0, 1, 0) is appropriate.

```
# Chia tập dữ liệu
train <- data[1:40,]
test <- data[41:44,]
model_3 <- arima(train$Investment, order = c(0, 1, 0))
model_3
```

Output:

```
Call:
arima(x = train$Investment, order = c(0, 1, 0))


sigma^2 estimated as 0.0004637:  log likelihood = 94.35,  aic = -186.69
>
```

The data was converted to first-order differencing using the d=1 parameter in the ARIMA model.

The value of sigma^2 (estimated variance) is 0.0004637, which indicates the volatility of the data series.

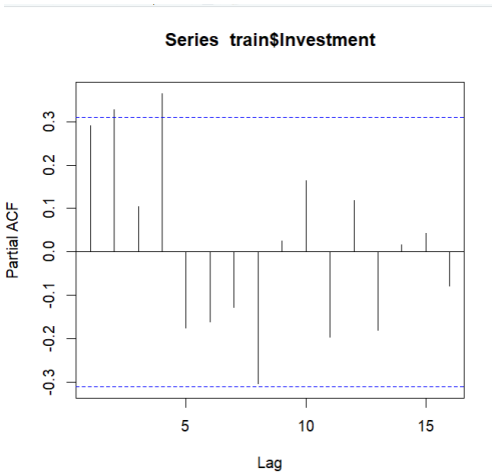
The log likelihood value is 94.35, which indicates how well the model fits the data.

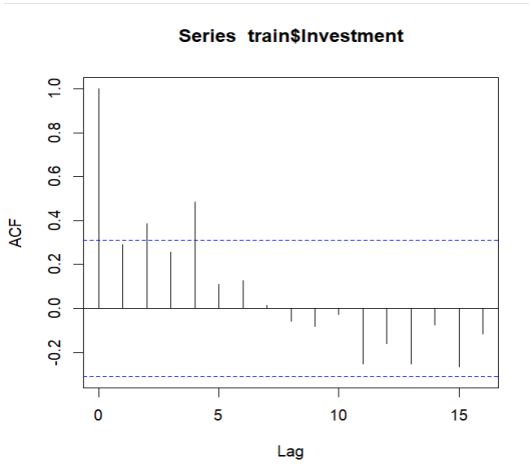
The AIC (Akaike Information Criterion) value is -186.69, indicating the fit and simplicity of the model. The smaller the AIC value, the better the model.

The ARIMA model (0, 1, 0) has no autoregressive or moving average components, using only simple differences. If the data has a more complex structure, the ARIMA model (0, 1, 0) may not be able to correctly predict the future trend.

For more accurate prediction results, consider using other ARIMA models or further analyzing other factors such as autoregressive components and moving averages in the data.

```
acf(train$Investment)
pacf(train$Investment)
model_4 <- arima(train$Investment, order = c(2, 1, 0))
model_4
```





Based on PACF and ACF can select ARIMA model(2,1,0)

The estimated sigma^2 (variance) is 0.0002915. The log likelihood is 103.15, and the AIC (Akaike Information Criterion) is -200.29.

Overall, the ARIMA(2, 1, 0) model provides a better fit to the data compared to the ARIMA(0, 1, 0) model, as indicated by a higher log likelihood and a lower AIC value. The inclusion of the autoregressive terms allows the model to capture more complex patterns and dependencies in the data, leading to improved forecasting performance.

```
# Dự đoán 4 quý đầu năm 2022
forecast <- predict(model_4, n.ahead = 4)
# Hiển thị dự báo
print(forecast$pred)
# So sánh dự báo với dữ liệu thực tế
real_data <- data$Investment[41:44]  # Dữ liệu thực tế cho 4 quý trong năm 2022
comparison <- data.frame(Real_Data = real_data, Forecast = forecast$pred)
print(comparison)Output:
```

```
    Real_Data   Forecast
1 0.009482318 0.01416337
2 0.017051661 0.01414637
3 0.023251382 0.01426154
4 0.014868137 0.01418497
```

It appears that the forecasted values are relatively close to the actual values.