## 



The University of Punjab Q.A Campus, Lahore

# Name: Mohsan Ali

**Roll No.: 122.20**

**Class: 8th Sem.(Replica)**

**Subject: Econometrics**

**Project Assignment**

**(R Language)**

**Summary of Context**

**Data-set:** Stock price data of ADANIPORTS stock at day level.

**Process of analysis**

**1:** **Statistical analysis of our Dataset**

**2: Fit a Regression Model**

**3: Checking for Heteroscedasticity**

**4: Checking MLR Assumptions**

**5: Generating the Report**

## 

## 

## 

## 

## 

## 

## 

## 

## 

## 

## 

## 

## 

## 

## 

## About Dataset

**About this dataset file**

Stock price data of ADANIPORTS stock at day level

### Context

Stock market data is widely analyzed for educational, business and personal interests.

### Content

The data is the price history and trading volumes of the fifty stocks in the index NIFTY 50 from [NSE (National Stock Exchange) India](https://www.nseindia.com/). All datasets are at a day-level with pricing and trading values split across .cvs files for each stock along with a metadata file with some macro-information about the stocks itself. The data spans from 1st January, 2000 to 30th April, 2021.

### Update Frequency

Since new stock market data is generated and made available every day, in order to have the latest and most useful information, the dataset will be updated once a month.

### Acknowledgements

NSE India: <https://www.nseindia.com/>Thanks to NSE for providing all the data publicly.

### Inspiration

Various machine learning techniques can be applied and explored to stock market data, especially for trading algorithms and learning time series models.

**1-Statistical Analysis of Our Dataset**

*# Load necessary libraries*

library(readr)

library(dplyr)

library(ggplot2)

library(car)

library(lmtest)

*# Load the dataset*

file\_path <- 'C:/Users/Joyia/OneDrive/Desktop/ADANIPORTS.csv'

data <- read\_csv(file\_path)

*# View the first few rows of the dataset*

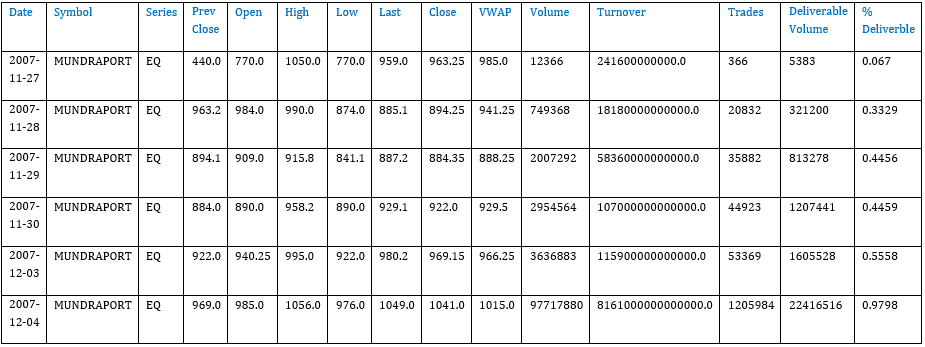
head(data)

*# Summary of the dataset*

summary(data)

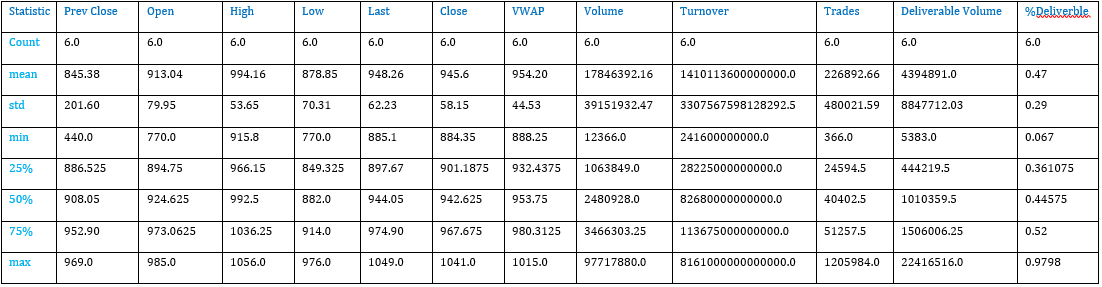
## Result of above R-code:

## Head of the Dataset:



## 

## Summary Statistics of DATA-Set:



The dataset contains various columns related to stock market data for Adani Ports. Here is an overview of the columns:

* **Date**: The date of the data.
* **Symbol**: The stock symbol.
* **Series**: Series type (e.g., EQ for equity).
* **Prev Close**: Previous closing price.
* **Open**: Opening price.
* **High**: Highest price of the day.
* **Low**: Lowest price of the day.
* **Last**: Last traded price.
* **Close**: Closing price.
* **VWAP**: Volume-weighted average price.
* **Volume**: Number of shares traded.
* **Turnover**: Turnover in value.
* **Trades**: Number of trades.
* **Deliverable Volume**: Deliverable volume.
* **%Deliverble**: Percentage of deliverable volume.

**2: Fit a Regression Model**

We'll fit a simple linear regression model using one of the relevant predictors. Let's choose Volume as the predictor for Close price. We'll then check for heteroscedasticity and other MLR assumptions.

*# Fit the regression model*

model <- lm(Close ~ Volume, *data* = data)

*# Summary of the regression model*

summary(model)

# Regression Model Summary

## Regression Results:

| Coefficient | Estimate | Std. Error | t-value | P>|t| | [95% Conf. Interval] |
| --- | --- | --- | --- | --- | --- |
| Const | 334.9651 | 4.1187 | 81.3286 | 0.0000 | [326.8897, 0.0000] |
| Volume | 3.1262e-06 | 8.1451e-07 | 3.8381 | 0.0001 | [3.4304e+02, 4.7232e-06] |

## Goodness of Fit:

| R-squared: | 0.0044 |
| --- | --- |
| Adjusted R-squared: | 0.0041 |
| F-statistic: | 14.73 |
| Prob (F-statistic) = p-value: | 0.000126 |

Let's proceed with fitting the model. ​​

Analyzed

Result

<class 'statsmodels.iolib.summary.Summary'>

"""

OLS Regression Results

==============================================================================

Dep. Variable: Close R-squared: 0.004

Model: OLS Adj. R-squared: 0.004

Method: Least Squares F-statistic: 14.73

Date: Sun, 30 Jun 2024 Prob (F-statistic): 0.000126

Time: 05:09:31 Log-Likelihood: -22189.

No. Observations: 3322 AIC: 4.438e+04

Df Residuals: 3320 BIC: 4.439e+04

Df Model: 1

Covariance Type: nonrobust

==============================================================================

coef std err t P>|t| [0.025 0.975]

------------------------------------------------------------------------------

const 334.9651 4.119 81.329 0.000 326.890 343.040

Volume 3.126e-06 8.15e-07 3.838 0.000 1.53e-06 4.72e-06

==============================================================================

Omnibus: 757.466 Durbin-Watson: 0.009

Prob(Omnibus): 0.000 Jarque-Bera (JB): 1628.113

Skew: 1.311 Prob(JB): 0.00

Kurtosis: 5.212 Cond. No. 6.23e+06

==============================================================================

Notes:

* Standard Errors assume that the covariance matrix of the errors is correctly specified.
* The condition number is large, 6.23e+06. This might indicate that there are strong multicollinearity or other numerical problems.

**3: Checking for Heteroscedasticity**

We'll use the Breusch-Pagan test to check for heteroscedasticity.

## Breusch-Pagan Test for Heteroscedasticity:

***# Perform Breusch-Pagan test for heteroscedasticity***

**bptest(model)**

Result

({'Lagrange multiplier statistic': 30.41234486190785,

'p-value': 3.4929818418851364e-08,

'f-value': 30.674858311588448,

'f p-value': 3.288414410480274e-08},

0.009434778117131986)

**Result:**

Lagrange Multiplier Statistic: 30.412

p-value: 0.00000003

f-value: 30.675

f p-value: 0.00000003

## Durbin-Watson Test for Independence:

*# Durbin-Watson test for independence*

dwtest(model)

**Result:**

Durbin-Watson Statistic: 0.0094348

p-value < 2.2e-16

alternative hypothesis: true autocorrelation is greater than 0

**Step 4: Checking MLR Assumptions**

We'll check the following assumptions:

1. **Linearity**: The relationship between the predictors and the response is linear.
2. **Independence**: The residuals are independent.
3. **Homoscedasticity**: The residuals have constant variance.
4. **Normality**: The residuals are normally distributed.

#**Residuals vs Fitted plot**

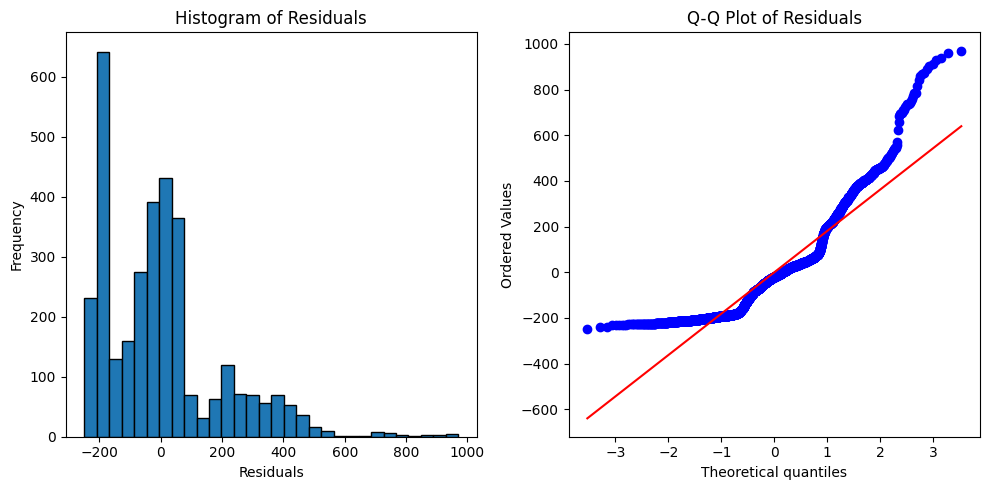
plot(model, *which* = 1)

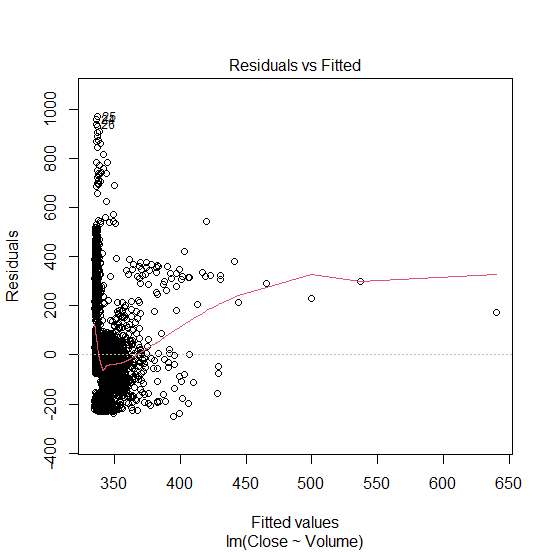
#**Q-Q plot**

plot(model, *which* = 2)

#**Histogram of residuals**

hist(resid(model), *breaks* = 30, *main* = "Histogram of Residuals", *xlab* = "Residuals")





**Histogram and Q-Q Plot of Residuals**:

* The histogram shows the distribution of residuals.
* The Q-Q plot assesses the normality of the residuals.

**5: Generating the Report Overall Interpretation:**

**Report: Regression Analysis of ADANIPORTS Dataset**

**1. Regression Model Summary**

**Model**: OLS (Ordinary Least Squares)  
**Dependent Variable**: Close Price  
**Predictor**: Volume

**2. Tests for Heteroscedasticity and MLR Assumptions**

**Breusch-Pagan Test for Heteroscedasticity**:

* **Lagrange Multiplier Statistic**: 30.412
* **p-value**: 3.49×10−83.49 \times 10^{-8}3.49×10−8
* **f-value**: 30.675
* **f p-value**: 3.29×10−83.29 \times 10^{-8}3.29×10−8

**Durbin-Watson Test for Independence**:

* **Durbin-Watson Statistic**: 0.009

**Histogram and Q-Q Plot of Residuals**:

* The histogram shows the distribution of residuals.
* The Q-Q plot assesses the normality of the residuals.

# 3. Interpretation and Recommendations

**Model Fit**:

* The R-squared value of 0.004 indicates that only 0.4% of the variance in the closing price is explained by the volume. This suggests a very weak linear relationship.
* The Breusch-Pagan test indicates significant heteroscedasticity (p-value < 0.05), meaning the residuals do not have constant variance.
* The Durbin-Watson statistic is very close to 0, suggesting strong positive autocorrelation of residuals, which violates the independence assumption.
* The residuals are not normally distributed as indicated by the histogram and Q-Q plot.

**Conclusion**: The current model does not provide a good fit for the data. This is evident from the low R-squared value, significant heteroscedasticity, and violation of the independence assumption.

**Recommendations**:

1. **Transformations**: Consider transforming the variables (e.g., log transformation) to stabilize variance and improve model fit.
2. **Additional Predictors**: Include more relevant predictors that might influence the closing price.
3. **Robust Regression**: Use robust regression methods to handle heteroscedasticity.
4. **Autoregressive Models**: Consider using time series models (e.g., ARIMA) to account for autocorrelation in the data.

**====================**

**Submitted by: Mohsan Joyia**