Shubham Dhadage

Prn:22060641063

LINEAR MODEL ASSIGNMENT

## Introduction

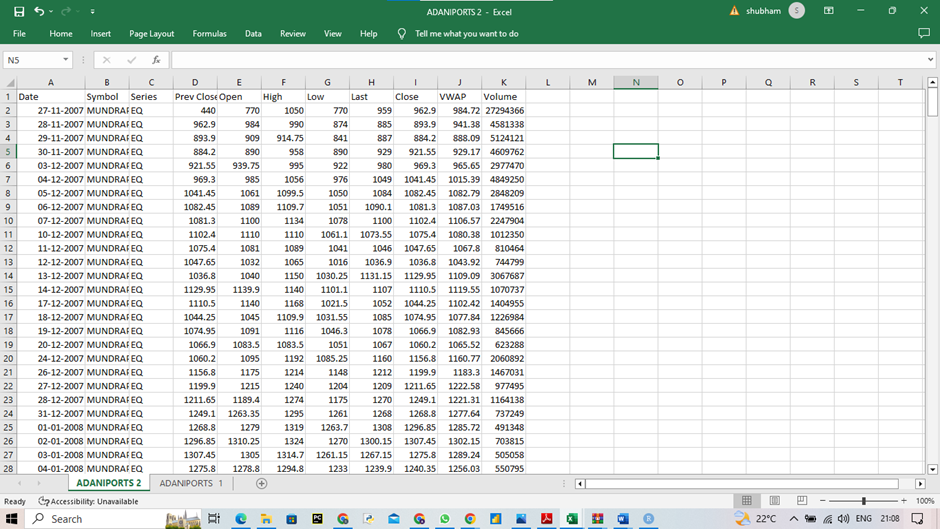
The closing price is an important metric in the stock market because it represents the final traded price of a security for the day. It is the last price at which the stock was traded before the market closes for the day. The closing price is used by investors, traders, and analysts to assess the performance of a stock for that day, and to make decisions about buying, selling or holding the stock.It is also used as a benchmark for calculating daily returns, price changes, and other technical indicators, such as moving averages and trading ranges.Overall, the closing price is an important factor in determining the daily performance and value of a stock, and it plays a significant role in the decision-making process of market participants.

## **Problem Statement**

**Investors and analysts focus on intraday fluctuations, others prioritize the closing price as the key indicator of a stock's performance for the day. This disparity in opinion creates confusion and uncertainty among market participants and can lead to suboptimal investment decisions. Therefore, there is a need to determine the significance of the closing price and its impact on the valuation and trading of stocks, and to establish a standardized method for evaluating stock performance**

## Data Collection

The data has been collected from **kaggle.** For the regression analysis, The sample size is 350.

Source of data : Kaggle

## **Methodology:**

## **Simple linear regression**

Simple linear regression is a statistical technique used to model the relationship between two variables, typically denoted as x and y. It assumes that there is a linear relationship between the two variables, meaning that a change in x will result in a proportional change in y.

The goal of simple linear regression is to find the line of best fit, which is a straight line that minimizes the sum of the squared differences between the predicted y values and the actual y values. This line can be expressed as **y = mx + b**, where m is the slope of the line and b is the y-intercept.

## **Multiple linear regression**

Multiple linear regression is a statistical technique used to model the relationship between **two or more independent variables** (x1, x2, ..., xn) and a **dependent variable (y).** It extends the concept of simple linear regression by considering multiple independent variables that may have an impact on the dependent variable.

The goal of multiple linear regression is to find the best fitting linear equation that describes the relationship between the independent variables and the dependent variable. This equation can be expressed as:

**y = b0 + b1x1 + b2x2 + ... + bnxn + e**

where **y is the dependent variable**,

**b0 is the intercept**,

**b1** x1,**, b2** x2,**, ..., bn are the coefficients of the independent variables**,

**e is the error term.**

The coefficients of the independent variables, b1, b2, ..., bn, represent the change in y for each unit change in the corresponding independent variable, holding all other independent variables constant. The error term, e, represents the difference between the predicted y value and the actual y value for each observation.

Multiple linear regression is used in many fields, such as economics, social sciences, and business, to analyze the relationship between multiple independent variables and a dependent variable. It can also be used to make predictions about future outcomes based on past data.

## **Characteristics of multiple linear regression: -**

1. Multiple independent variables: Unlike simple linear regression, which has only one independent variable, multiple linear regression has two or more independent variables.
2. Linear relationship: Multiple linear regression assumes that there is a linear relationship between the independent variables and the dependent variable. This means that the relationship between the variables can be represented by a straight line.
3. Assumptions: Multiple linear regression has several assumptions that must be met, including normality, homoscedasticity, independence, and linearity.
4. Coefficients: The coefficients in multiple linear regression represent the change in the dependent variable associated with a one-unit change in the corresponding independent variable, holding all other independent variables constant.
5. Model fit: Multiple linear regression aims to find the best fitting model that describes the relationship between the independent variables and the dependent variable. The quality of the fit is usually assessed using measures such as R-squared, adjusted R-squared, and the root mean square error.

## **Hypothesis:**

**Ho: there is relationship between closing price and open price**

**H1: there is no relationship**



We have assumed alpha to be 0.05. The p-value 0.04 is less than alpha, therefore we accept **Ho**

i.e there is relation ship between closing price and open price. Therefore closing price is important

## Correlation:-

We use correlation to check the correlation between two variables. If the is no correlation between variables we cannot to regression analysis.



There is high correlation between variables so we can do regression analysis.

## **Interpretation of the Regression Statistics Table:**



1. The multiple R represents the correlation between the predicted values of y generated and the actual value of y for each unit, here the r is 0.996011 we can say that the previous closing price and open price is highly correlated with each other.
2. R2= 0.9920 means that 99.2% of the variation in Y is explained by the predictor X i.e. 99.2% of the variability accounted for by the regression model.
3. The value of standard error here refers to the estimated standard deviation of the error term e i.e It is sometimes called the standard error of the regression.

## **Interpretation of the ANOVA Table:**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |
|  | ANOVA |  |  |  |  |  |  |
|  |  | *df* | *SS* | *MS* | *F* | *Significance F* |  |
|  | Regression | 5 | 21082164 | 4216433 | 8548.587838 | 0 |  |
|  | Residual | 343 | 169178.4 | 493.2315 |  |  |  |
|  | Total | 348 | 21251342 |  |  |  |  |

The ANOVA table shows the two parts like mean squared and sum of squares i.e. residual sum of squares and regression sum of squares

H0: b=0

H1: b≠0 (i.e. this hypothesis is related to regression model.)

Here the f-statistics is 8548.58 and the corresponding p-value in the table is less than 0.05 which shows that there is relationship between prev closing price and opening price.

Interpretation of the Coefficient table



The regression output depicts the regression parameters and the associated output.

Fitting the regression model

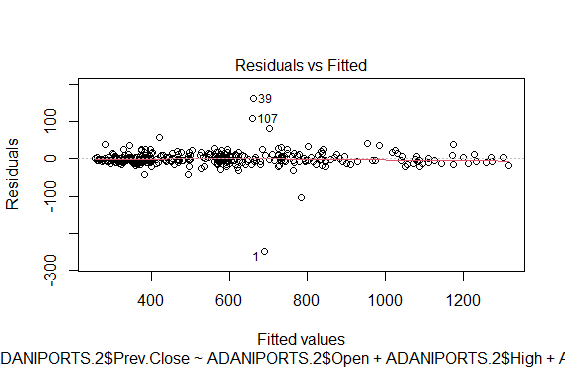
From Fig 3, we have =6.10774and =1.2373,b2= -0.5177,……..

Hence, the fitted regression equation is

Standard error of the regression coefficients

The standard errors of and are SE ()= 3.0295 ,

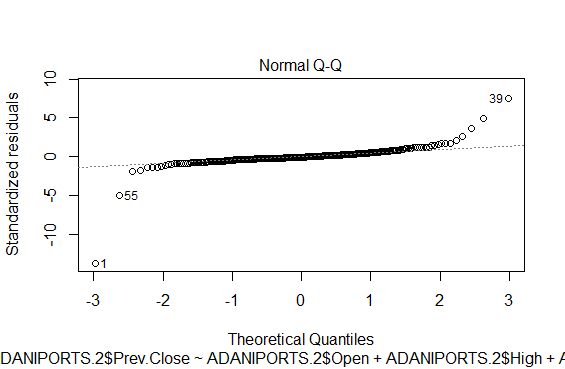
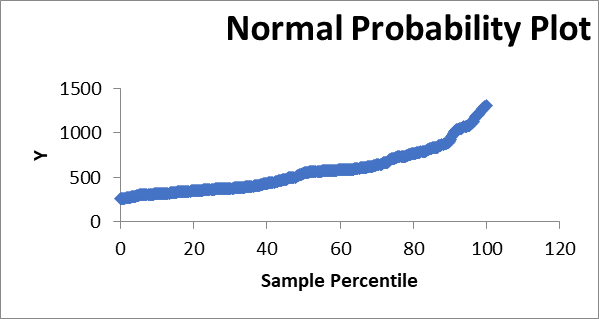
SE ()= 0.0898,(b)=0.09932.(b)=0.1112,(b)=0.31606 and (b)=0.347434

**Interpretation of the Residual plot**

It is a scatter plot of residuals on the y-axis and the fitted values on the x-axis

As we can see in this plot there are many residuals cluttered to the left side of the graph and we can see only 3 outliers.

**Interpretation of the Normal Probability Plot**



This normal qq plot is the normality plot through which we can see that majorly the points i.e. not all residuals fall on the normal line but most of them fall on line therefore we can say that the points are not that normally distributed.

**Conclusion:**

From the regression analysis we have R square is 0.99203917. since the value is close to 1 , we can say that there is high positive correlation between the opening and closing price of adaniports Stocks.

From the ANOVA table we conclude that the regression model is a good fit and there is a significant relationship between the independent variable and the dependent variable.

From the analysis of intercept, we can conclude that the line of regression passes through the origin. And from the analysis of the slope, we conclude that closing price affects the opening price of stock.

From the residual plot, we see that the residuals are randomly scattered . Hence, the regression model is well fitted.

There for the conclusion is that the closing price is important for opening price of stock.