Disclaimer I have changed my dataset

|  |  |
| --- | --- |
| **Price** | **MarketCap** |
| 16.39 | 6.553 |
| 21.09 | 3.058 |
| 20.82 | 2.685 |
| 79.15 | 162.88 |
| 29.41 | 6.547 |
| 160.38 | 38.37 |
| 17.62 | 3.639 |
| 70.1 | 6.273 |
| 243.47 | 44.19 |
| 13.38 | 11.67 |
| 17.85 | 3.509 |
| 769.88 | 12.986 |
| 15.1 | 2.786 |
| 10.92 | 2.357 |
| 31.88 | 3.533 |
| 245.12 | 12.587 |
| 880.55 | 49.252 |
| 16.83 | 3.667 |
| 105.83 | 6.128 |
| 24.4 | 16.55 |
| 15.24 | 2.074 |
| 19.39 | 1.914 |
| 123.41 | 10.962 |

This is my new dataset

(a)

setwd("C:/Users/maxim/OneDrive/Documents")

getwd()

Market<- read.csv("Markets.csv")

y<-Market$Price

x<-Market$MarketCap

(b)

I chose this dataset to explore the relationship between stock price and market capitalization due to its relevance in financial analysis. Understanding how these factors have a linear relationship provides me with insights into market valuation dynamics, aiding investors, and analysts in decision-making processes which interests me.

(c)

Code:

model <- lm(y~x)

summary(model)

Output:

Call:  
lm(formula = y ~ x)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-255.09 -91.19 -85.83 -1.11 707.93   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 102.571 54.762 1.873 0.0751 .  
x 1.422 1.435 0.991 0.3329   
---  
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
  
Residual standard error: 231.5 on 21 degrees of freedom  
Multiple R-squared: 0.04469, Adjusted R-squared: -0.0008059   
F-statistic: 0.9823 on 1 and 21 DF, p-value: 0.3329

Regression Parameters:

The least square parameters of the model are B0hat and B1hat

(Intercept) 102.571 =B0hat  
x 1.422 =B1hat

(d)

The intercept of 102.571$ represents the estimated stock price when market capitalization is zero and the coefficient (slope) of 1.422 indicates that on average, each unit increase in market capitalization is associated with a corresponding increase of 1.422 billion units in stock price.

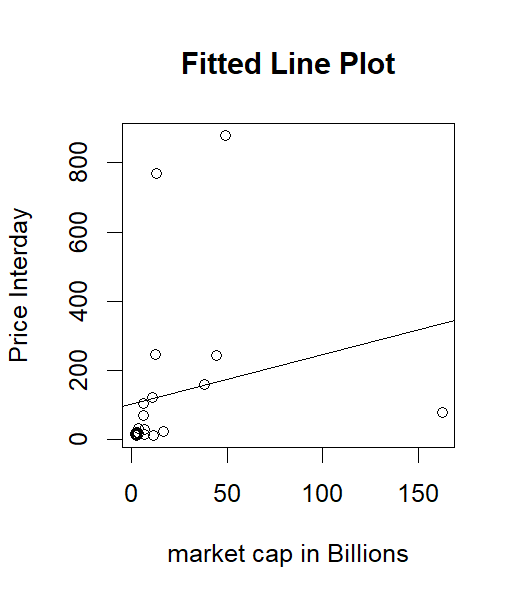
(e)

Delving into the performance of the simple linear regression model tasked with explaining the relationship between market capitalization(X) and stock price(Y), we uncover limitations that call its effectiveness into question. By scrutinizing its fit statistics, parameter significance, residual analysis, and compliance with underlying assumptions we can illuminate its strengths and weaknesses as an explanatory tool.

Code:

plot(x,y, main="Fitted Line Plot",xlab="market cap in Billions", ylab="Price Interday")

abline(model)



Firstly, the scatter plot shows a weak and unclear linear trend. Data points are scattered, and there is a slight positively increasing slope. There are several outliers, particularly at higher market caps, deviating significantly from the potential trend.

Furthermore, a meager R-squared value of 0.0469 indicates it explains only a negligible portion of price variance. It suggests only 4.469% of the variance in the dependent variable is explained by the independent variables in the model revealing other potent factors that the model fails to capture. Further casting doubt on the model's explanatory power is the lack of statistically significant parameters. Neither the intercept(p-value=0.0751) nor the slope(p-value=0.3329) coefficient reaches the conventional 5% significance level, raising concerns about the existence of a genuine linear relationship between these variables.

Code:

anova(model)

critical<-qf(0.95,1,21,)

0.9823>critical

Output:

Analysis of Variance Table  
  
Response: y  
 Df Sum Sq Mean Sq F value Pr(>F)  
x 1 52663 52663 0.9823 0.3329  
Residuals 21 1125872 53613

4.324794  
FALSE

The model's overall significance is similarly questionable, as evidenced by the F-statistic(0.9823). Furthermore, The F-statistic from the data(0.9823) falls below the 95th quantile(4.324794) critical value. While it suggests the model lacks statistically significant explanatory power at the 5% level, an exceptionally low F-statistic(0.9823) indicates that the model explains only a small portion of the variance in the data. The inability to reject the null hypothesis of the variance ratio corroborates the absence of statistically meaningful explanatory power. This points towards a weak linear relationship, but it does not entirely disprove its existence.

Code:

stdres <- rstandard(model)

yhat <- fitted(model)

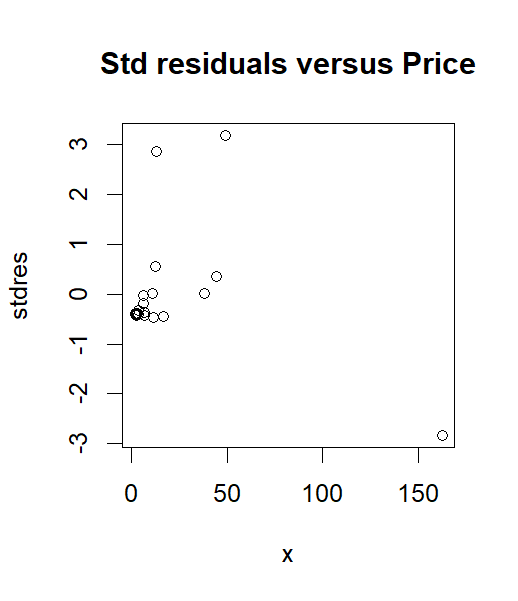
plot(x,stdres, main="Std residuals versus Price")

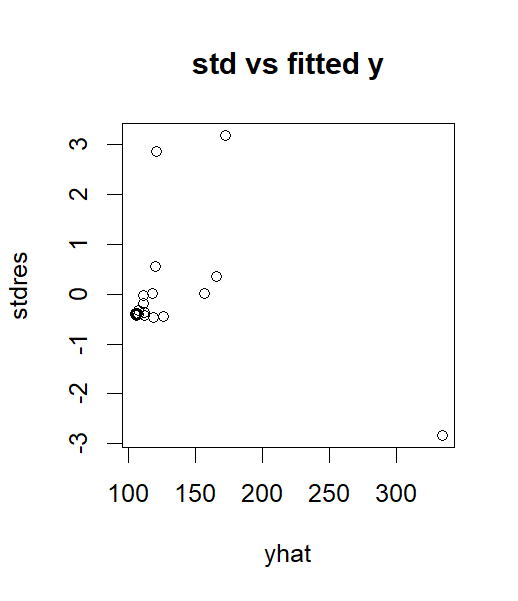
plot(yhat,stdres,main="std vs fitted y")

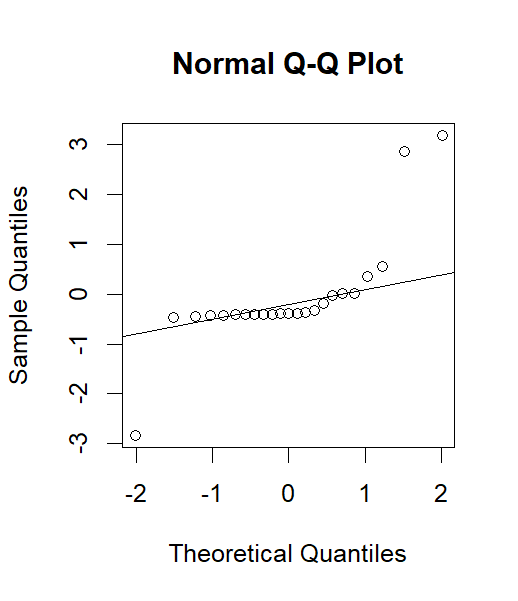
qqnorm(stdres)

qqline(stdres)

Output:







Delving deeper, the residual analysis exposes additional shortcomings. A non-random pattern in the x vs stdres plot reveals the model's inability to capture the underlying trend, suggesting a more nuanced relationship than a simple linear one. The funnel-shaped pattern in the yhat vs stdres plot with data points spreading out as yhats increase, indicative of heteroscedasticity, violates model assumption of constant variance and compromises the reliability of its results. Furthermore, scattered outliers across both plots highlight potentially influential data points that the model fails to account for. Finally, the non-normal distribution of standardized residuals, evident in the QQ plot at higher theoretical quantities, confirms yet another violation of the model's fundamental assumptions. Although lower values do seem to follow a normal distribution.

In conclusion, the cumulative evidence from various analyses demonstrates that the simple linear regression model is inadequate in explaining the relationship between market capitalization and stock price. Its low explanatory power, lack of overall significance and violation of assumptions require further exploration.