

Mô hình hồi quy tuyến tính bội (Nhóm 26)

- Nguyễn Đình Tuấn Long 20216941
- Nguyễn Văn Tuấn 20216965
- Vũ Văn Huy 20216931

Mô tả bài toán và bộ dữ liệu:

Dữ liệu thu thập từ Yahoo Finance cung cấp các chỉ số quan trọng liên quan đến cổ phiếu, tạo ra một bộ dữ liệu lý tưởng để áp dụng mô hình hồi quy tuyến tính. Trong những biến như Giá mở cửa (Open), Giá cao nhất (High), Giá thấp nhất (Low), và Khối lượng giao dịch (Volume), chúng ta có thể tận dụng mối quan hệ số liên quan để dự đoán Giá đóng cửa (Close).

Các thông số này có thể giúp xây dựng một mô hình hồi quy tuyến tính, trong đó các biến độc lập như Giá mở cửa, Giá cao nhất, Giá thấp nhất, và Khối lượng giao dịch sẽ được sử dụng để dự đoán biến phụ thuộc, tức là Giá đóng cửa. Mục tiêu là tìm ra một mô hình có thể mô tả mối quan hệ tuyến tính giữa các yếu tố này để dự đoán giá cổ phiếu trong các phiên giao dịch tiếp theo.

Sự linh hoạt và khả năng dự đoán của mô hình hồi quy tuyến tính có thể giúp đưa ra quyết định đầu tư thông minh, dựa trên hiểu biết chặt chẽ về diễn biến giá cổ phiếu từ các thông số quan trọng.

	Date	Open	High	Low	Close	Volume
1	2022-12-30	83.120	84.0500	82.4700	84.00	62401194
2	2022-12-29	82.870	84.5500	82.5500	84.18	54995895
3	2022-12-28	82.800	83.4800	81.6900	81.82	58228575
4	2022-12-27	84.970	85.3500	83.0000	83.04	57284035
5	2022-12-23	83.250	85.7800	82.9344	85.25	57433655
6	2022-12-22	85.520	85.6800	82.2500	83.79	81431314
7	2022-12-21	86.175	87.2250	85.2100	86.77	59267156
8	2022-12-20	85.330	86.6100	84.3300	85.19	74348301
9	2022-12-19	87.510	87.6300	84.5100	84.92	83531460
10	2022-12-16	88.265	89.3500	86.7300	87.86	146144064
11	2022-12-15	89.890	89.9700	87.4700	88.45	84802877
12	2022-12-14	92.502	93.4581	89.8700	91.58	70298006
13	2022-12-13	95.230	96.2500	90.5200	92.49	100212038
14	2022-12-12	89.210	90.5800	87.8700	90.55	61999762

Dữ liệu gồm 756 hàng, 6 cột.

Phương pháp thực hiện (Methodology):

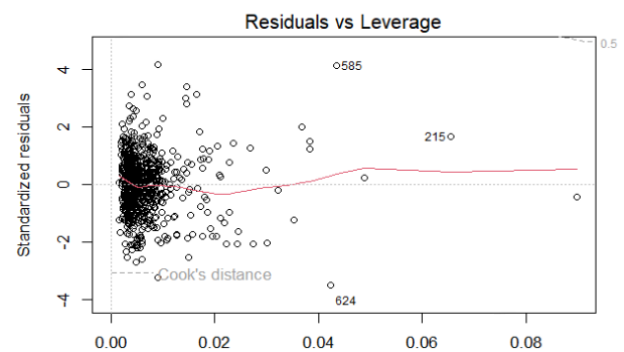
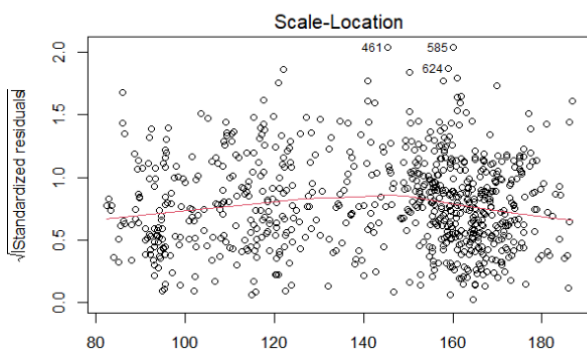
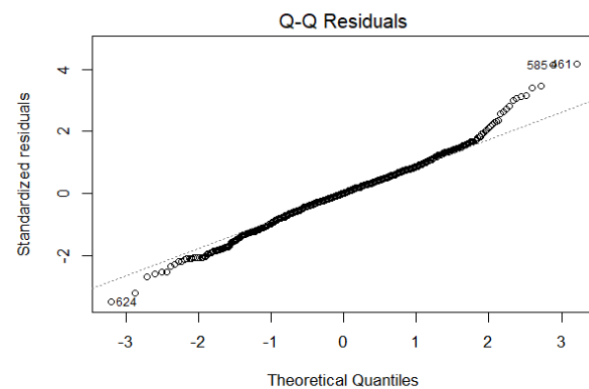
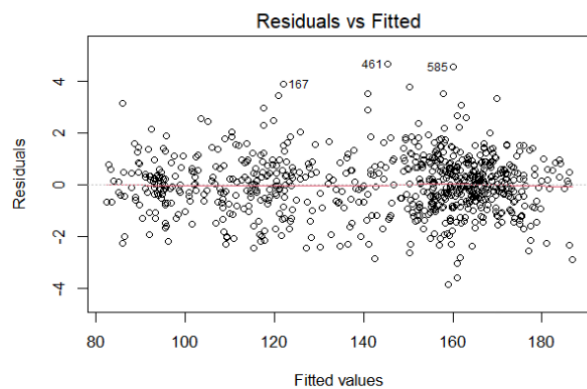
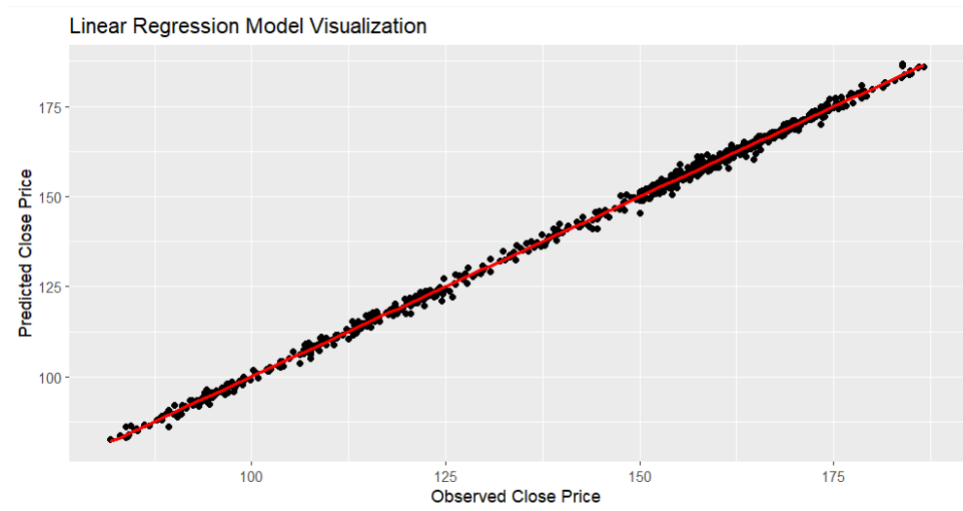
Sử dụng biến Open, High, Low, Volume để dự đoán giá Close.

```
dabien <- lm(data$Close ~ data$Open + data$High + data$Low + data$Volume)
```

predicted_data biểu thị cho giá trị dự đoán dựa trên mô hình.

```
predicted_data <- data.frame(Predicted = predict(dabien), Observed = data$Close)
```

Thực hiện vẽ đồ thị để xem xét sự tương quan:



Đánh giá mô hình:

Tính hệ số R_{squared} : 0.9984

Một số các thông số khác:

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.052e-02  2.595e-01   0.233   0.816
data$Open    -6.507e-01  2.930e-02 -22.208 <2e-16 ***
data$High     8.381e-01  3.169e-02  26.444 <2e-16 ***
data$Low      8.118e-01  2.936e-02  27.646 <2e-16 ***
data$Volume   5.242e-10  1.420e-09   0.369   0.712
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.123 on 751 degrees of freedom
Multiple R-squared:  0.9984,    Adjusted R-squared:  0.9984
F-statistic: 1.159e+05 on 4 and 751 DF,  p-value: < 2.2e-16
```

Nhận xét: Mô hình có độ chính xác cao (Multiple R-squared gần 1), giả sử rằng mô hình phản ánh tốt sự biến động của dữ liệu. F-statistic và p-value của nó cũng thể hiện sự ý nghĩa toàn cục của mô hình.