-------------Perdictive Analytics-------------

Tập Dữ Liệu: Breast Cancer

Phân tích phương sai ANOVA

1. Nhóm được tạo cho cột Mitoses có tác động thực sự đến khả năng khối u là ác tính và Lành tính hay không.

- Ta sử dụng 2 biến: Hạng(Class) (có 2 loại: 0: Lành tính(benign), 1: Ác tính(Malignant)) và biến ước

lượng Mitoses (Định lượng)

-Kết quả:Phân tích ANOVA cho thấy F=152.0 và P=9.68e-32(với mức 0,05)

+ Chúng ta có thể thấy rằng bắt đầu từ Mitoses là 2, tỷ lệ khối u lành tính và ác tính dường như vẫn gần nhau.

Do đó,có thể nhóm các giá trị này để làm cho thông tin đơn giản hơn, giúp giảm nguy cơ trang bị quá mức hoặc

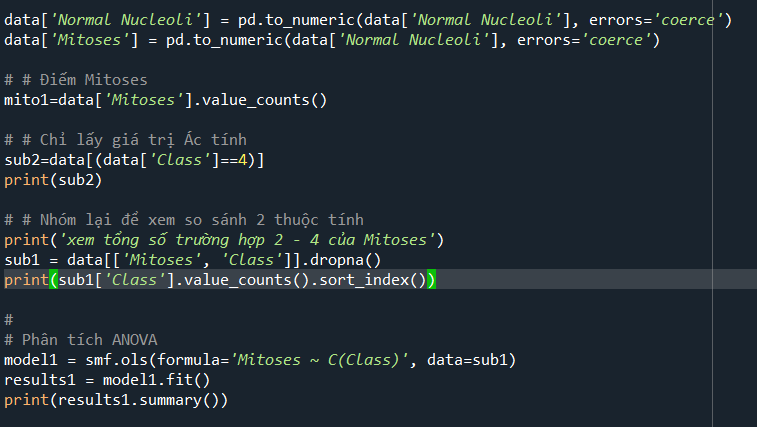
ảnh hưởng của các giá trị ngoại lai ở các cực, do đó làm cho các mô hình tương lai tốt hơn cho việc tổng quát hóa.

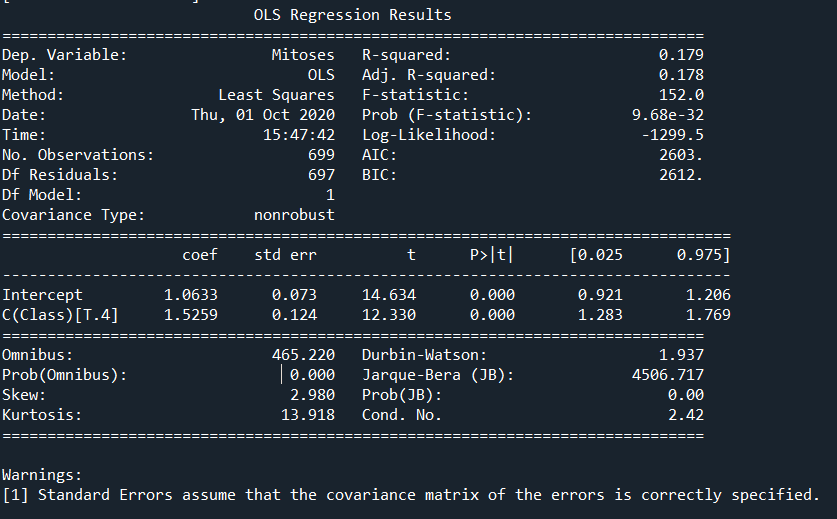
+ Với giá trị p<=9.68e-32 rất nhỏ, chúng ta có thể bác bỏ giả thuyết vô hiệu, xác nhận phát hiện của trên rằng nhóm được

tạo cho cột Mitoses có tác động thực sự đến khả năng khối u là ác tính. Cụ thể, chúng ta thấy rằng cơ hội xuất hiện

khối u ác tính cho các giá trị của Mitoses bằng 2 (hoặc nhiều hơn) lớn hơn cơ hội cho giá trị 1.

hình:





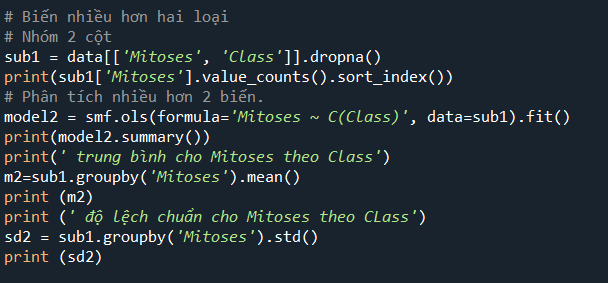
2. Liệu rằng các chỉ số đo được của Mitoses từ 1 – 10 có cần thiết lệ thuộc là Class hay không

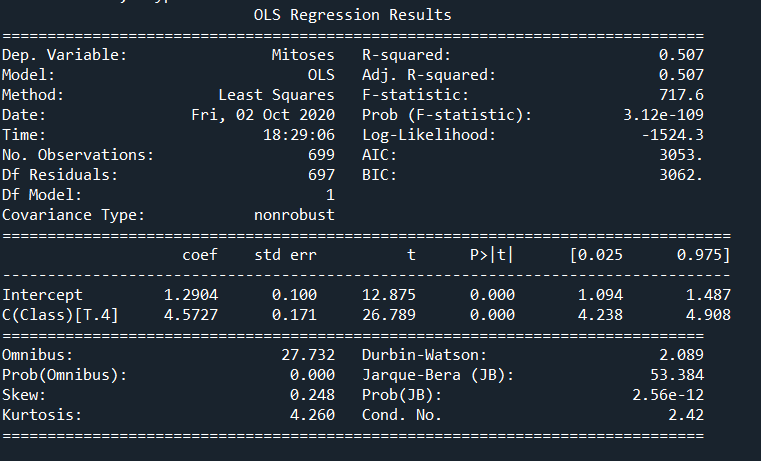
- Ta sử dụng 2 biến: Hạng(Class) (có 2 loại: 0: Lành tính(benign), 1: Ác tính(Malignant)) và biến ước

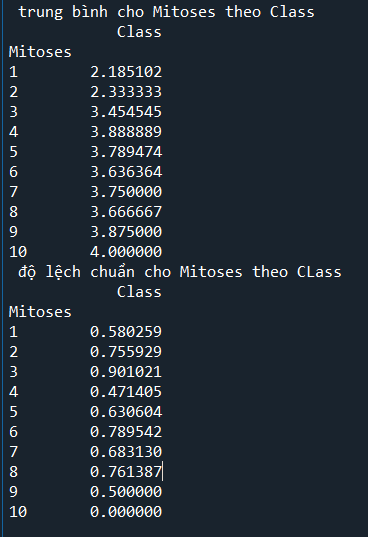
lượng Mitoses (Định lượng)

- Kết quả Phân tích ANOVA cho thấy con số ở đây rất rất nhỏ 3.12e-109 chúng ta có thể bác bỏ giả thuyết vô hiệu, xác nhận phát hiện của trên rằng nhóm được tạo cho cột Mitoses có lệ thuộc thực sự đến khả năng khối u là ác tính

- Độ lệch chuẩn cho thấy ở Mitoses rằng giá trị 10 là dường như không có và nó phụ thuộc nhiều nhất là ở 3







Phân tích sâu:

3. Mối liên quan giữ Độ dày cụm(Clump Thickness) và Nucleoli bình thường(Normal Nucleoli) chỉ lấy giá trị 1-5 của Nucleoli => các nhóm có sự khác biệt như thế nào

Kết quả:

Các nhóm 1-2,1-3,1-4,1-5,2-3,2-5 có khác biệt so với các nhóm còn lại cho ta thấy tỉ lệ phân bố giữa các nhóm tập trung nhiều nhất từ 1->2 của nhóm 1(Clump Thickness) và trải đều cho các nhóm 2(Normal Nucleoli)

Hình:

