



# Khái niệm về hồi quy

Bởi:

Phạm Trí Cao

## Khái niệm

Phân tích hồi quy là tìm quan hệ phụ thuộc của một biến, được gọi là biến phụ thuộc vào một hoặc nhiều biến khác, được gọi là biến độc lập nhằm mục đích ước lượng hoặc tiên đoán giá trị kỳ vọng của biến phụ thuộc khi biết trước giá trị của biến độc lập

Một số tên gọi khác của biến phụ thuộc và biến độc lập như sau:

Biến phụ thuộc: biến được giải thích, biến được dự báo, biến được hồi quy, biến phản ứng, biến nội sinh.

Biến độc lập: biến giải thích, biến dự báo, biến hồi quy, biến tác nhân hay biến kiểm soát, biến ngoại sinh.

Sau đây là một ví dụ về phân tích hồi quy

Ngân hàng XYZ muốn tăng lượng tiền huy động. Ngân hàng này muốn biết mối quan hệ giữa lượng tiền gửi và lãi suất tiền gửi, cụ thể hơn họ muốn biết khi tăng lãi suất thêm 0,1% thì lượng tiền gửi sẽ tăng trung bình là bao nhiêu.

Một nhà nghiên cứu nông nghiệp muốn biết năng suất tôm sú nuôi trong hệ thống thâm canh phụ thuộc thế nào vào diện tích ao nuôi, mật độ thả tôm giống, chi phí hoá chất xử lý môi trường, trình độ nhân công. Từ phân tích hồi quy này ông ta đề ra các chỉ tiêu kỹ thuật phù hợp cho loại hình này.

## Sự khác nhau giữa các dạng quan hệ

Quan hệ tất định và quan hệ thống kê

Quan hệ tất định là loại quan hệ có thể biểu diễn bằng một hàm số toán học. Một số quan hệ trong vật lý, hoá học và một số ngành khoa học tự nhiên khác là quan hệ tất định.

## Khái niệm về hồi quy

Ví dụ định luật Ohm trong vật lý : gọi  $U$  là điện áp,  $R$  là điện trở của mạch điện thì dòng điện  $I$  sẽ là  $I = \frac{U}{R}$ , nói cách khác khi điện áp và điện trở được cố định trước thì chúng ta chỉ nhận được một và chỉ một giá trị dòng điện.

Đa số các biến số kinh tế không có quan hệ tất định. Thí dụ ta không thể nói với diện tích nuôi tôm cho trước và kỹ thuật nuôi được chọn thì năng suất sẽ là bao nhiêu. Lý do là có rất nhiều biến số được kể đến trong mô hình cũng tác động lên năng suất, ngoài ra trong số các biến số vắng mặt này có những biến không thể kiểm soát được như thời tiết, dịch bệnh... Nhà nghiên cứu nông nghiệp kể trên chỉ có thể tiên đoán một giá trị trung bình của năng suất ứng với kỹ thuật nuôi đã chọn. Quan hệ giữa các biến số kinh tế có tính chất quan hệ thống kê.

### Hồi quy và quan hệ nhân quả

Mặc dù phân tích hồi quy dựa trên ý tưởng sự phụ thuộc của một biến số kinh tế vào biến số kinh tế khác nhưng bản thân kỹ thuật phân tích hồi quy không bao hàm quan hệ nhân quả. Một ví dụ điển hình của sự nhầm lẫn hai khái niệm này tiến hành hồi quy số vụ trộm ở một thành phố với số nhân viên cảnh sát của thành phố. Gọi  $Y$  là số vụ trộm trong một năm và  $X$  là số nhân viên cảnh sát. Khi chúng ta hồi quy  $Y$  theo  $X$ , nếu chúng ta tìm được mối quan hệ đồng biến của  $Y$  và  $X$  có ý nghĩa thống kê thì phân tích hồi quy này cho kết luận: “Tăng số lượng nhân viên cảnh sát sẽ làm tăng số vụ trộm”. Rõ ràng phân tích này sai lầm trong việc nhận định mối quan hệ nhân quả. Số cảnh sát tăng lên là do sự tăng cường của lực lượng cảnh sát trong bối cảnh số vụ trộm tăng lên. Vậy đúng ra chúng ta phải hồi quy số cảnh sát theo số vụ trộm hay  $X$  theo  $Y$ . Vậy trước khi phân tích hồi quy chúng ta phải nhận định chính xác mối quan hệ nhân quả.

Một sai lầm phổ biến nữa trong phân tích kinh tế lượng là quy kết mối quan hệ nhân quả giữa hai biến số trong khi trong thực tế chúng đều là hệ quả của một nguyên nhân khác. Ví dụ chúng ta phân tích hồi quy giữa số giáo viên và số phòng học trong toàn ngành giáo dục. Sự thực là cả số giáo viên và số phòng học đều phụ thuộc vào số học sinh. Như vậy phân tích mối quan hệ nhân quả dựa vào kiến thức và phương pháp luận của môn khác chứ không từ phân tích hồi quy.

### Hồi quy và tương quan

Phân tích tương quan chỉ cho thấy độ mạnh yếu của mối quan hệ tuyến tính giữa hai biến số. Phân tích tương quan cũng không thể hiện mối quan hệ nhân quả. Ví dụ chúng ta xét quan hệ giữa hai biến số  $X$  là số bệnh nhân bị xơ gan và  $Y$  là số lít rượu được tiêu thụ của một nước. Chúng ta có thể nhận được hệ số tương quan cao giữa  $X$  và  $Y$ . Hệ số tương quan được xác định như sau:

$$r_{YX} = \frac{\text{cov}(X, Y)}{S_X S_Y} = \frac{\text{cov}(Y, X)}{S_Y S_X} = r_{XY}$$

## Khái niệm về hồi quy

Qua đẳng thức này chúng ta cũng thấy trong phân tích tương quan vai trò của hai biến là như nhau và hai biến đều là ngẫu nhiên.

Phân tích hồi quy của X theo Y cho ta biết trung bình số bệnh nhân bị xơ gan là bao nhiêu ứng với lượng tiêu dùng rượu cho trước. Chúng ta không thể đảo ngược hồi quy thành Y theo X. Phân tích hồi quy dựa trên giả định biến độc lập là xác định trong khi biến phụ thuộc là ngẫu nhiên. Chúng ta tìm giá trị kỳ vọng của biến phụ thuộc dựa vào giá trị cho trước của của biến độc lập.