1. **Hồi quy tuyến tính đa biến**

**1.1 Định nghĩa**

Phân tích hồi quy là một trong những công cụ cơ bản của kinh tế lượng. Phân tích hồi quy là mô tả mối quan hệ phụ thuộc của một biến (thường được gọi là biến phụ thuộc hay biến được giải thích) vào một hay nhiều biến khác (thường được gọi là biến độc lập hay biến giải thích).



Ta ký hiệu biến được giải thích (biến phụ thuộc) là Y và các biến giải thích (biến độc lập) là X1 , X2 ,..., Xk.

**Ví dụ 1:** Mô hình hồi quy bội với biến phụ thuộc Y và hai biến độc lập X1 và X2 , trong đó:

+ Y = doanh thu của công ty

+ X1 = chi phí cho quảng cáo

+ X2 = lương trả cho nhân viên tiếp thị.

* Trong ví dụ 2 ta xác định mối quan hệ giữa doanh thu Y và chi phí cho quảng cáo X1 và tiền lương trả cho nhân viên tiếp thị X2 .

**1.2 ANOVA Test:**

The ANOVA test for significance of the entire model is:

A screenshot of a computer

Description automatically generated with low confidence

* One can also test for significance of individual regression coefficients.
  1. **Phân tích hồi quy giải quyết những vấn đề sau:**
* Ước lượng giá trị trung bình của biến phụ thuộc với giá trị đã cho của biến độc lập.
* Dự báo giá trị của Y khi biết được giá trị của biến giải thích X.
* Kiểm định giả thuyết và bản chất của sự phụ thuộc và xác định hiệu quả tác động của biến độc lập lên biến phụ thuộc. Khi mô tả mối quan hệ giữa biến phụ thuộc và các biến độc lập cũng cần phân biệt rõ các kiểu quan hệ sau:

o Quan hệ tất định hay còn gọi là quan hệ toán học giữa X và Y, quan hệ này được cho dưới dạng một hàm số Y f(X) = , cứ cho một giá trị của X ta sẽ xác định được một giá trị của Y.

o Quan hệ thống kê là quan hệ mà không xác định được giá trị duy nhất của Y khi cho biết giá trị của X mà ta chỉ biết được giá trị của Y theo một xác suất nào đó, hay nói cách khác ta chỉ xác định được phân bố xác suất của Y khi biết giá trị của X.

**Ví dụ về hồi quy tuyến tính:** Dữ liệu bên dưới đây sử dụng các mẫu từ 205 loại xe ô tô khác nhau. Và được chia thành các cột thuộc tính như sau: Giá xe, Loại nguyên liệu (1: xăng; 2: dầu diesel), Cylinder number, Enginesize, Horsepower. Sau đó, chúng ta sẽ sử dụng phương pháp hồi quy tuyến tính đa biến để phân tích mối quan hệ giữa thuộc tính “Giá xe” với các thuộc tính còn lại. Để từ đó, ta có thể đưa ra kết luận rằng thuộc tính “Giá xe” có bị ảnh hưởng bởi những thuộc tính khác hay không.

Table

Description automatically generated

**Ta có phương trình như sau:**

Price = β0 + fueltype\*β1 + cylindernumber\*β2 + enginesize\*β3 + horsepower\*β4

**Trong đó:**

* **Price:** Là biến phụ thuộc
* **Fueltype, cylindernumber, enginesize, horsepower**: Là biến độc lập.
* **β0**: Là hằng số hồi quy
* **β1, β2, β3, β4:** Là hệ số hồi quy

Trước khi bắt đầu, chúng ta phải thực hiện kiểm thử ANOVA để đánh giá độ phù hợp mô hình một cách chính xác qua kiểm định giả thuyết.

*H0: β1 = β2 = β3 = β4 = 0*

*H1: at least one βj is not 0*

**Bài 4d.**

1. EXERCISE 4D

**Problem statement:** This is a survey on satisfaction with insurance services collected from 25 students. These students are divided into three main groups of education level: college graduate, graduate degree and some colleges. The analysis aims to find out whether the level of satisfaction affected by education level.

Hypothesis for ANOVA test:

*H0: The level of satisfacion is the same at all education level (μ1 = μ2 = μ3)*

*H1: at least one education level is different from the others*

* 1. Excel
  2. R
  3. Pythons