

# VỀ PHẦN DƯ TRONG PHƯƠNG TRÌNH HỒI QUY TUYẾN TÍNH

## ON THE RESIDUAL PARTS OF THE LINEAR REGRESSION EQUATIONS

CAO VĂN NUÔI

*Trường Đại học Sư phạm, Đại học Đà Nẵng*

CAO NGỌC CHÂU

*Học viên cao học khoá 2005-2008*

### TÓM TẮT

Bài báo trình bày phương pháp tìm phân phối của phần dư trong phương trình hồi quy tuyến tính. Các kết luận trong bài báo được chứng minh một cách chi tiết. Khi biết được phân phối của phần dư trong phương trình hồi quy tuyến tính, người ta đánh giá được sai số của dữ liệu đầu ra và hiểu rõ thêm quy luật phân phối của phần dư. Vì vậy, phân phối của phần dư trong phương trình hồi quy tuyến tính là rất quan trọng.

### ABSTRACT

This paper presents on the distributions of the residual parts of the linear regression equations. The results in this paper proved in detail. If we known the distribution of this residual parts then we can estimate errors of output data and to study the distributions of this residual parts of the linear regression equations. So, the distributions of residual parts of the linear regression equation is the most importance.

## 1. Khái niệm

**Định nghĩa 1.1.** Nếu  $Z_1, Z_2, \dots, Z_n$  là các biến ngẫu nhiên chuẩn tiêu chuẩn độc lập thì  $X$  được xác định bởi

$$X = Z_1^2 + Z_2^2 + \dots + Z_n^2,$$

được gọi là có phân phối khi-bình phương với  $n$  bậc tự do và ký hiệu:

$$X \sim \chi_n^2.$$

Thông thường ta ký hiệu  $E(\xi), V(\xi)$  lần lượt là kỳ vọng và phương sai của đại lượng ngẫu nhiên  $\xi$ .

**Định nghĩa 1.2.** Một đại lượng ngẫu nhiên  $\xi$  được gọi là chuẩn hoá, nếu  $E(\xi)=0$  và  $V(\xi)=1$ . Mọi đại lượng ngẫu nhiên (khác hằng số, tức là  $P(\xi \neq C)=1$ ) đều có thể đưa về dạng chuẩn hoá bằng cách đặt

$$\xi' = \frac{\xi - E(\xi)}{\sqrt{V(\xi)}}.$$

Trong khuôn khổ bài báo này ta xét phương trình hồi quy tuyến tính có dạng:

$$Y = \alpha + \beta x + \varepsilon,$$

trong đó :  $x$  là biến độc lập;

$Y$  là biến ngẫu nhiên phụ thuộc vào biến độc lập  $x$  ;

$\alpha, \beta$  gọi là các tham số hồi quy;

$\varepsilon$  gọi là sai số ngẫu nhiên và giả thiết  $E(\varepsilon) = 0$ .

Với mẫu hai chiều  $(x_i, Y_i), i = 1, \dots, n$  cỡ mẫu  $n$  ta có:

$$Y_i = \alpha + \beta x_i + \varepsilon_i, \quad i = \overline{1, n}$$

với  $\varepsilon_i$  là sai số ngẫu nhiên.

Ta xét trong trường hợp các  $\varepsilon_i$  thỏa mãn các điều kiện sau:

a)  $E(\varepsilon_i) = 0, \forall i = \overline{1, n}$ .

b)  $E(\varepsilon_i \varepsilon_j) = \sigma^2 = \begin{cases} \sigma^2 & \text{nếu } i = j \\ 0 & \text{nếu } i \neq j \end{cases}$

c)  $\varepsilon_i \sim \square(0, \sigma^2), \forall i = \overline{1, n}$ .

## 2. Phân phối của phần dư

**Bổ đề 2.1.** [2] Cho  $X = Y + K$  với giả thiết  $X$  là đại lượng ngẫu nhiên có phân phối  $\chi_n^2$ ,  $K$  là đại lượng phân phối  $\chi_1^2$ ; và  $Y$  và  $K$  là độc lập. Khi đó,  $Y$  có phân phối  $\chi_{n-1}^2$ .

**Bổ đề 2.2.** [2] Cho  $Z_1, \dots, Z_n$  là các biến ngẫu nhiên chuẩn tiêu chuẩn độc lập và  $X$  là đại lượng ngẫu nhiên được xác định bởi  $X = Z_1^2 + Z_2^2 + \dots + Z_n^2, i = 1, \dots, n$  thì:

a)  $E(X) = n$ .

b)  $V(X) = 2n$ .

**Bổ đề 2.3.** [2] Nếu  $(X_1, \dots, X_n)$  là mẫu ngẫu nhiên sinh ra bởi phân phối chuẩn có  $E(X_i) = \mu, V(X_i) = \sigma^2, i = \overline{1, n}$ ; thì:

a)  $E(\bar{X}) = \mu, V(\bar{X}) = \frac{\sigma^2}{n}$ .

b)  $\frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$

trong đó  $\bar{X} = \sum_{i=1}^n \frac{X_i}{n}$  và  $s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$ .

**Định lý 2.1.** Giả sử rằng đầu ra  $Y_i, i = 1, \dots, n$  là các biến ngẫu nhiên chuẩn độc lập có  $E(Y_i) = \alpha + \beta x$  và  $V(Y_i) = \sigma^2, \forall i = 1, n$  với  $A$  và  $B$  lần lượt là ước lượng bình phương bé nhất của  $\alpha$  và  $\beta$  thì:

$$\frac{\sum_{i=1}^n (Y_i - A - Bx_i)^2}{\sigma^2} \sim \chi_{n-2}^2.$$

**Chứng minh.** Từ bổ đề 2.3. ta có:

$$\frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} + \left[ \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \right]^2$$

Từ đó suy ra,

$$i) \frac{\sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2}{\sigma^2} = \frac{\sum_{i=1}^n (Y_i - \alpha - Bx_i)^2}{\sigma^2} + \frac{n(B - \beta)^2 x_i^2}{\sigma^2}.$$

$$ii) \frac{\sum_{i=1}^n (Y_i - Bx_i)^2}{\sigma^2} = \frac{\sum_{i=1}^n (Y_i - Bx_i - A)^2}{\sigma^2} + \frac{n(A - \alpha)^2}{\sigma^2}.$$

\* Bây giờ ta sẽ chứng minh i)

Từ phương trình  $Y = \alpha + \beta x + \varepsilon$  (2.1)

Xét  $Y = A + \beta x + \varepsilon^{(1)}$  (2.2)

trong đó  $A$  là ước lượng của  $\alpha$ .

Lấy (2.2) – (2.1) về theo về ta được:

$$(A - \alpha) + (\varepsilon^{(1)} - \varepsilon) = 0$$

$$\Rightarrow E(A - \alpha) + E(\varepsilon^{(1)} - \varepsilon) = 0,$$

Mà  $E(A - \alpha) = 0 \Rightarrow E(\varepsilon^{(1)} - \varepsilon) = 0 \Rightarrow E(\varepsilon^{(1)}) + E(\varepsilon) = 0.$

Bây giờ, ta đặt:  $Z = Y_i - \alpha \Rightarrow \bar{Z} = \overline{Y_i - \alpha} = \bar{Y}_i - \alpha = \alpha + Bx_i - \alpha = Bx_i$

Suy ra i) được chứng minh.

\* Để chứng minh ii) ta đặt:

$$T = Y_i - Bx_i$$

$$\Rightarrow \bar{T} = \overline{Y_i - Bx_i} = \bar{Y}_i - Bx_i = A.$$

Vậy ii) được chứng minh.

Từ  $Y_i$  là các biến ngẫu nhiên chuẩn độc lập nên ta có:  $\frac{(Y_i - E[Y_i])}{\sqrt{V(Y_i)}}$ ,  $i = 1, \dots, n.$

Suy ra,  $\frac{\sum_{i=1}^n (Y_i - E[Y_i])^2}{V(Y_i)} = \frac{\sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2}{\sigma^2} \sim \chi_n^2.$

Từ i) có: 
$$\frac{\sum_{i=1}^n (Y_i - \alpha - \beta x_i)}{\sigma^2} \sim \chi_n^2,$$

và 
$$\frac{n(B - \beta)^2 x_i^2}{\sigma^2} \sim \chi_1^2,$$

kết hợp với bổ đề 2.1. suy ra: 
$$\frac{\sum_{i=1}^n (Y_i - Bx_i - \alpha)^2}{\sigma^2} \sim \chi_{n-1}^2.$$

Hơn nữa từ ii) có: 
$$\frac{n(A - \alpha)^2}{\sigma^2} \sim \chi_1^2,$$

Suy ra, 
$$\frac{\sum_{i=1}^n (Y_i - A - Bx_i)^2}{\sigma^2} \sim \chi_{n-2}^2.$$

Vậy định lý được chứng minh hoàn toàn.

**Định lý 2.2.** Giả sử đầu ra  $Y_i, i=1, \dots, n$  là các biến ngẫu nhiên chuẩn độc lập có  $E(Y_i) = \alpha + \beta x$  và  $V(Y_i) = \sigma^2, \forall i = \overline{1, n}$  với  $A$  và  $B$  lần lượt là ước lượng bình phương bé nhất của  $\alpha$  và  $\beta$  thì:

$$a) E \left[ \frac{\sum_{i=1}^n (Y_i - A - Bx_i)^2}{\sigma^2} \right] = n - 2.$$

$$b) V \left[ \frac{\sum_{i=1}^n (Y_i - A - Bx_i)^2}{\sigma^2} \right] = 2(n - 2).$$

**Chứng minh.**

a) Từ định lý 2.1. có: 
$$\frac{\sum_{i=1}^n (Y_i - A - Bx_i)^2}{\sigma^2} \sim \chi_{n-2}^2.$$

Nghĩa là,  $\frac{\sum_{i=1}^n (Y_i - A - Bx_i)^2}{\sigma^2}$  có phân phối khi bình phương với  $n-1$  bậc tự do,

nên theo bổ đề 2.2. ta có: 
$$E \left[ \frac{\sum_{i=1}^n (Y_i - A - Bx_i)^2}{\sigma^2} \right] = n - 2.$$

b) Từ bổ đề 2.1 có: với  $X \sim \chi_n^2$  thì  $V(X) = 2n$ ,

Từ định lý 2.1 ta có: 
$$\frac{\sum_{i=1}^n (Y_i - A - Bx_i)^2}{\sigma^2} \sim \chi_{n-2}^2.$$

Suy ra,

$$V \left[ \frac{\sum_{i=1}^n (Y_i - A - Bx_i)^2}{\sigma^2} \right] = 2(n - 2).$$

Vậy định lý được chứng minh.

### **TÀI LIỆU THAM KHẢO**

- [1] William Feller, *An Introduction to Probability Theory and Its Application*, New York-Chichester-Brisbane-Toronto, Singapore, 1971.
- [2] Shedon M. Ross, *Introduction to prabbility and statistics for engineers and scientists*, New York-Chichester-Brisbane-Toronto, Singapore, 1987.