



CHƯƠNG 1 – NHỮNG VẤN ĐỀ CƠ BẢN TRONG MÔ HÌNH HỒI QUY ĐƠN

Nội dung

1.1. Phân tích hồi quy

1.2. Số liệu trong phân tích hồi quy

1.3. Hồi quy tổng thể

1.4. Sai số ngẫu nhiên

1.5. Hồi quy mẫu

1.1. Phân tích hồi quy

1.1.1. Bản chất của phân tích hồi quy

Thuật ngữ “hồi quy - regression” được Francis Galton sử dụng vào năm 1886. Đến nay hồi quy vẫn là một công cụ cơ bản, hữu hiệu của Kinh tế lượng.

- **Khái niệm:** PTHQ nghiên cứu mối liên hệ phụ thuộc của một biến, gọi là biến phụ thuộc (dependent variable) vào một hay nhiều biến khác, gọi là biến độc lập (independent variable), trong đó ước lượng giá trị trung bình của biến phụ thuộc theo các giá trị xác định của biến độc lập.

1.1. Phân tích hồi quy

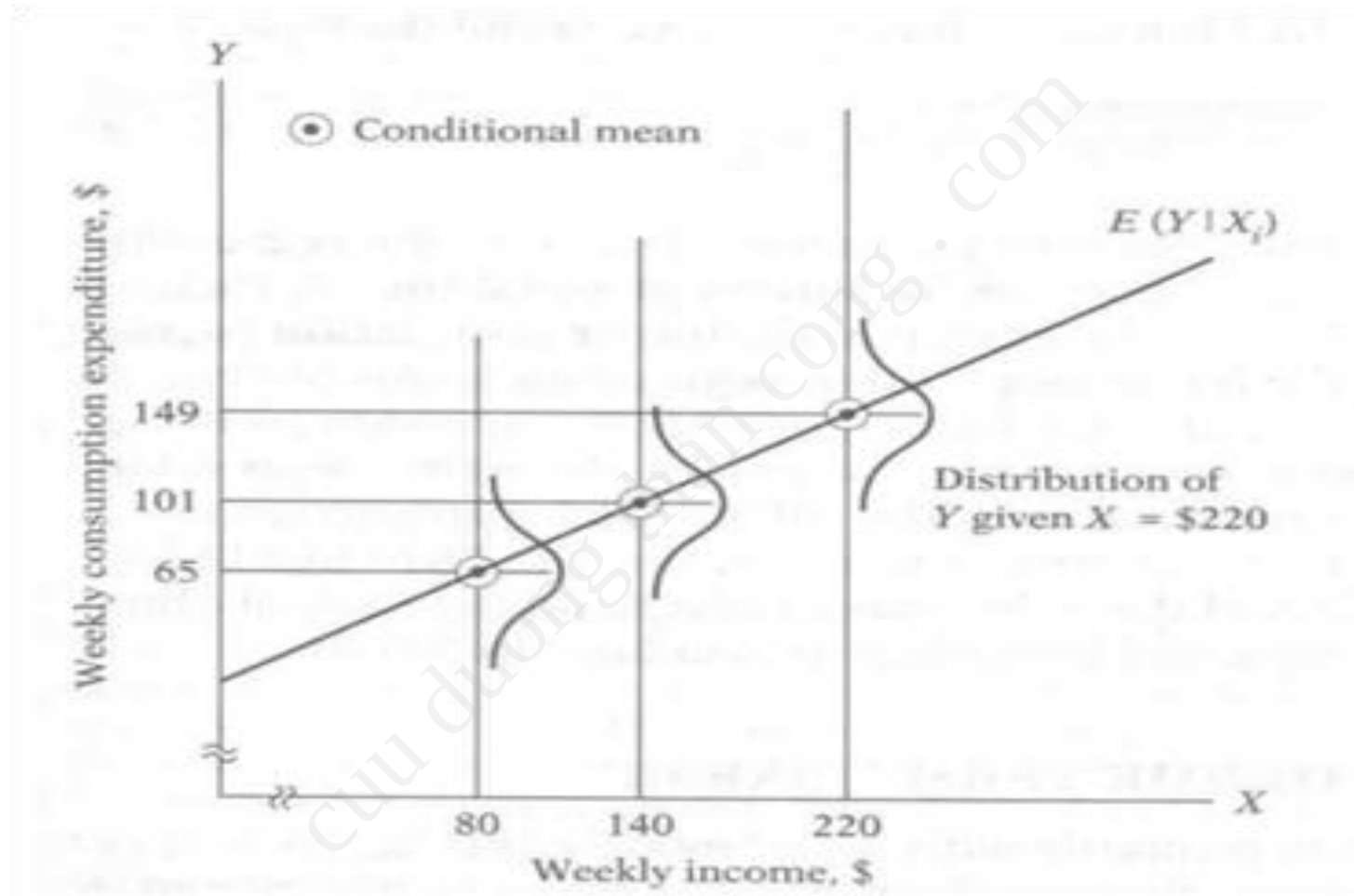
- **Ví dụ:** Giả sử dự đoán mức tiêu dùng trung bình hàng tuần của hộ gia đình trong một khu dân cư khi biết thu nhập khả dụng.
- **Các biến nghiên cứu:**
 - Biến phụ thuộc: Y là chi tiêu hàng tuần của hộ gia đình (đơn vị: trăm nghìn đồng).
 - Biến độc lập: X là thu nhập khả dụng hàng tuần của hộ gia đình (đơn vị: trăm nghìn đồng).
 - Mẫu nghiên cứu: Có số liệu về chi tiêu và thu nhập của 60 hộ gia đình. Kết quả sắp xếp theo thu nhập tăng dần.

1.1. Phân tích hồi quy

Bảng 1. Xác suất có điều kiện của tiêu dùng theo các mức thu nhập

$P(Y/X_i)$ \ X	80	100	120	140	160	180	200	220	240	260
	1/5	1/6	1/5	1/7	1/6	1/6	1/5	1/7	1/6	1/7
	1/5	1/6	1/5	1/7	1/6	1/6	1/5	1/7	1/6	1/7
	1/5	1/6	1/5	1/7	1/6	1/6	1/5	1/7	1/6	1/7
	1/5	1/6	1/5	1/7	1/6	1/6	1/5	1/7	1/6	1/7
	1/5	1/6	1/5	1/7	1/6	1/6	1/5	1/7	1/6	1/7
	1/5	1/6	1/5	1/7	1/6	1/6	1/5	1/7	1/6	1/7
	1/5	1/6	1/5	1/7	1/6	1/6	1/5	1/7	1/6	1/7
		1/6		1/7	1/6	1/6		1/7	1/6	1/7
				1/7				1/7		1/7
$E(Y/X_i)$	65	77	89	101	113	125	137	149	161	173

1.1. Phân tích hồi quy



Hình 1. Hàm mật độ xác suất của Y với từng giá trị thu nhập X

1.1. Phân tích hồi quy

■ Xét mỗi các mối quan hệ:

- Lượng cầu của người tiêu dùng về một hàng hóa (ký hiệu Y) với giá của nó (X)
- Tỷ lệ thay đổi của tiền lương (Y) theo tỷ lệ thất nghiệp (X)
- Tỷ lệ tiền mặt nắm giữ trong tổng thu nhập (Y) với tỷ lệ lạm phát (X)
- Mức cầu (Y) phụ thuộc mức chi cho quảng cáo (X)
- Sản lượng của một loại nông sản (Y) phụ thuộc lượng phân bón, lượng mưa, nhiệt độ, v.v...

1.1. Phân tích hồi quy

1.1.2. Mục đích của phân tích hồi quy

- Ước lượng giá trị trung bình của biến phụ thuộc khi biết giá trị của biến độc lập.
- Kiểm định các giả thuyết về bản chất của mối quan hệ giữa biến phụ thuộc và biến độc lập mà lý thuyết kinh tế đưa ra.
- Dự báo giá trị trung bình và giá trị cá biệt của biến phụ thuộc khi biết giá trị của các biến độc lập.

1.1. Phân tích hồi quy

1.1.3. Phân tích hồi quy và các mối quan hệ khác

- Phân tích hồi quy và quan hệ hàm số (functional relationship)
- Phân tích hồi quy và phân tích tương quan (correlation analysis)
- Phân tích hồi quy và quan hệ nhân quả (causation relationship)

1.2. Số liệu trong phân tích hồi quy

■ Các loại số liệu:

- Số liệu chuỗi thời gian (Time series data): thu thập tại nhiều thời điểm trên cùng một đối tượng quan sát.
- Số liệu chéo (Undate-Cross section data): thu thập tại một thời điểm trên nhiều đối tượng quan sát.
- Số liệu kết hợp: Kết hợp hai loại số liệu trên là số liệu thu thập tại nhiều thời điểm, trên nhiều đối tượng quan sát. Số liệu bảng (Panel data) là dạng đặc biệt của số liệu kết hợp.

1.2. Số liệu trong phân tích hồi quy

Bảng 2. Tốc độ tăng trưởng GDP của VN qua một số năm

Năm	DGDP (%)
1995	9.5
1996	9.3
1997	8.2
1998	5.8
1999	4.8
2000	6.8
2001	6.9
2002	7.1
2003	7.3
2004	7.8
2005	8.4
2006	8.2
2007	8.48
2008	6.18
2009	5.32
2010	6.78
2011	5.89
2012	5.25
2013	5.42
2014	5.98
2015	6.68
2016	6.21
2017	6.81

Đơn vị tính: Triệu % – Nguồn Tổng cục Thống kê

1.2. Số liệu trong phân tích hồi quy

Bảng 3. Tổng sản phẩm trong nước theo giá hiện hành của một số nước Đông Nam Á năm 2016

Đông Nam Á	..
Bru-nây	11.401,0
Cam-pu-chia	20.017,0
In-đô-nê-xi-a	1.783,0
Lào	932.259,0
Ma-lai-xi-a	15.806,0
Mi-an-ma	296.536,0
Phi-li-pin	63.225,0
Xin-ga-po	304.905,0
Thái Lan	296.976,0
Đông Ti-mo	407.026,0
Việt Nam	205.305,0

Đơn vị tính: Triệu USD – Nguồn Tổng cục Thống kê

1.2. Số liệu trong phân tích hồi quy

Bảng 4. Tốc độ tăng tổng sản phẩm trong nước của một số nước Đông Nam Á trong 4 năm

	2013	2014	2015	2016
Đông Nam Á	..	..	..	..
Bru-nây	-2,10	-2,30	-0,60	-2,50
Cam-pu-chia	7,40	7,10	7,00	7,00
In-đô-nê-xi-a	5,60	5,00	4,90	5,00
Lào	8,00	7,60	7,30	7,00
Ma-lai-xi-a	4,70	6,00	5,00	4,20
Mi-an-ma	8,40	8,00	7,00	5,90
Phi-li-pin	7,10	6,10	6,10	6,90
Xin-ga-po	5,00	3,60	1,90	2,00
Thái Lan	2,70	0,90	2,90	3,20
Đông Ti-mo	2,60	4,20	4,10	5,70
Việt Nam	5,40	6,00	6,70	6,20

Đơn vị tính: % – Nguồn Tổng cục Thống kê

1.2. Số liệu trong phân tích hồi quy

■ Nguồn gốc của số liệu

- Số liệu từ các nguồn được phát hành như: Niên giám thống kê, tạp chí,...Hoặc có thể do các cơ quan nhà nước, các tổ chức quốc tế, các công ty tư nhân thu thập.
- Số liệu từ các cuộc điều tra thực tế hoặc đi mua.

1.2. Số liệu trong phân tích hồi quy

- **Bản chất chung của số liệu kinh tế - xã hội**
 - Phần lớn là các số liệu phi thực nghiệm có độ tin cậy chưa cao.
 - Đối với kinh tế học nói riêng và khoa học xã hội nói chung, rất khó bố trí thí nghiệm có kiểm soát.
- ✓ *Kết quả của nghiên cứu sẽ không chỉ phụ thuộc vào mô hình được lựa chọn mà còn phụ thuộc rất nhiều vào chất lượng của số liệu.*

1.2. Số liệu trong phân tích hồi quy

■ Hạn chế của số liệu

- Hầu hết các số liệu trong kinh tế là các loại số liệu phi thực nghiệm, do vậy có độ tin cậy không cao.
- Các cuộc điều tra bằng câu hỏi có thể không nhận được câu trả lời chính xác hoặc không trả lời hết câu hỏi.
- Các mẫu thu thập trong các cuộc điều tra rất khác nhau về kích thước, cho nên rất khó khăn trong việc so sánh các kết quả giữa các đợt điều tra.
- Các số liệu kinh tế thường rất tổng hợp, không cho phép đi sâu vào phân tích các đơn vị nhỏ.
- Ngoài ra còn có những số liệu thuộc bí mật quốc gia mà không phải ai cũng tiếp cận và sử dụng được.

1.3. Hồi quy tổng thể

1.1.3. Hàm hồi quy tổng thể

- Tổng thể (Population) là toàn bộ tập hợp các phần tử đồng nhất theo một dấu hiệu nghiên cứu định tính hoặc định lượng nào đó.
- Giả sử một tổng thể nghiên cứu gồm N phần tử với hai dấu hiệu nghiên cứu X, Y tạo thành một biến ngẫu nhiên hai chiều (X, Y) , có bảng phân phối xác suất đồng thời như sau:

1.3. Hồi quy tổng thể

Bảng 5. Phân phối xác suất đồng thời của X và Y.

	X_1	X_2	...	X_k
Y_1	$P(Y_1, X_1)$	$P(Y_1, X_2)$	...	$P(Y_1, X_k)$
Y_2	$P(Y_2, X_1)$	$P(Y_2, X_2)$		$P(Y_2, X_k)$
...	...	...	...	...
Y_h	$P(Y_h, X_1)$	$P(Y_h, X_2)$	...	$P(Y_h, X_k)$

$$\sum_{j=1}^h \sum_{i=1}^k P(Y_j, X_i) = 1$$

1.3. Hồi quy tổng thể

- Các bảng phân phối xác suất có điều kiện của Y theo X_i ($i = \overline{1, k}$)

(Y/X_i)	Y_1	Y_2	$\dots$	Y_h
$P(Y/X_i)$	$P(Y_1, X_i)$	$P(Y_2, X_i)$	$\dots$	$P(Y_h, X_i)$

- Kỳ vọng toán của Y với điều kiện của X_i :

$$E(Y / X_i) = \sum_{j=1}^h Y_j * P(Y_j / X_i)$$

$$X = X_i \Rightarrow \exists! E(Y/X_i) \Rightarrow E(Y / X_i) = f(X_i) \quad (i = 1 \div k)$$

$E(Y/X)$ là một hàm số theo X và gọi là hàm hồi quy tổng thể của Y đối với X (Population Regression Function - PRF). Biểu diễn mối quan hệ giữa giá trị trung bình của biến phụ thuộc Y theo mỗi giá trị của biến độc lập X .

1.3. Hồi quy tổng thể

■ Ví dụ 1:

Y	X		
		2	3
	1	0,2	0,4
	2	0,3	0,1
	P(X)	0,5	0,5
			P(Y)
			0,6
			0,4
			1,0

Y X ₁	Y ₁	Y ₂	Y X ₂	Y ₁	Y ₂
P	$P(Y_1 X_1) = \frac{0,2}{0,5} = \frac{2}{5}$	$P(Y_2 X_1) = \frac{0,3}{0,5} = \frac{3}{5}$	P	$P(Y_1 X_2) = \frac{0,4}{0,5} = \frac{4}{5}$	$P(Y_2 X_2) = \frac{0,1}{0,5} = \frac{1}{5}$

$$E(Y | X_1) = Y_1.P(Y_1 | X_1) + Y_2.P(Y_2 | X_1) = 1.\frac{2}{5} + 2.\frac{3}{5} = \frac{8}{5}$$

$$E(Y | X_2) = Y_1.P(Y_1 | X_2) + Y_2.P(Y_2 | X_2) = 1.\frac{4}{5} + 2.\frac{1}{5} = \frac{6}{5}$$

1.3. Hồi quy tổng thể

- Nếu hàm hồi quy tổng thể có một biến độc lập thì gọi là hàm hồi quy đơn - Simple regression:

$$E(Y/X_i) = f(X_i)$$

- Nếu hàm hồi quy tổng thể có hơn một biến độc lập thì gọi là hàm hồi quy bội - Multiple regression:

$$E(Y/X_{1i}, X_{2i}, \dots) = f(X_{1i}, X_{2i}, \dots)$$

1.3. Hồi quy tổng thể

- Giả sử hàm hồi quy có dạng tuyến tính:

$$\text{PRF: } E(Y / X_i) = \beta_1 + \beta_2 X_i \quad (i = 1 \div k)$$

Trong đó: $\beta_1 = E(Y / X_i = 0)$ là hệ số chặn (intercept coefficient)

$$\beta_2 = \frac{dE(Y / X)}{dX} \text{ là hệ số góc (slope coefficient)}$$

1.3. Hồi quy tổng thể

1.3.2. Mô hình hồi quy tổng thể

- Nếu hàm hồi quy tổng thể là tuyến tính, khi đó mô hình hồi quy tổng thể tương ứng có dạng tuyến tính (Population Regression Model – PRM):

PRM:
$$Y_i = \beta_1 + \beta_2 X_i + U_i \quad (i = 1 \div N)$$

- Mô hình hồi quy tuyến tính có nghĩa là mô hình hồi quy có dạng tuyến tính đối với các tham số, nó có thể là tuyến tính hoặc phi tuyến đối với các biến số. Chẳng hạn một số dạng mô hình hồi quy tuyến tính:

$$Y_i = \beta_1 + \beta_2 X_i^2 + U_i, \quad \ln(Y_i) = \beta_1 + \beta_2 \ln(X_i) + U_i$$

1.3. Hồi quy tổng thể

1.3.3. Sai số ngẫu nhiên

- **Bản chất của sai số ngẫu nhiên:** U_i được gọi là sai số ngẫu nhiên (SSNN)

$$U_i = Y_i - E(Y / X_i)$$

$$E(U / X_i) = 0$$

- SSNN đại diện cho tất cả những yếu tố không biết, không có số liệu và không phải biến độc lập nhưng có tác động đến biến phụ thuộc. Trung bình ảnh hưởng của nó đến biến phụ thuộc bằng 0.
- Sự tồn tại của SSNN là tất yếu khách quan và nó có vai trò đặc biệt quan trọng trong phân tích hồi quy, nó phải thoả mãn những điều kiện nhất định thì việc phân tích trên mô hình hồi quy mới có ý nghĩa.

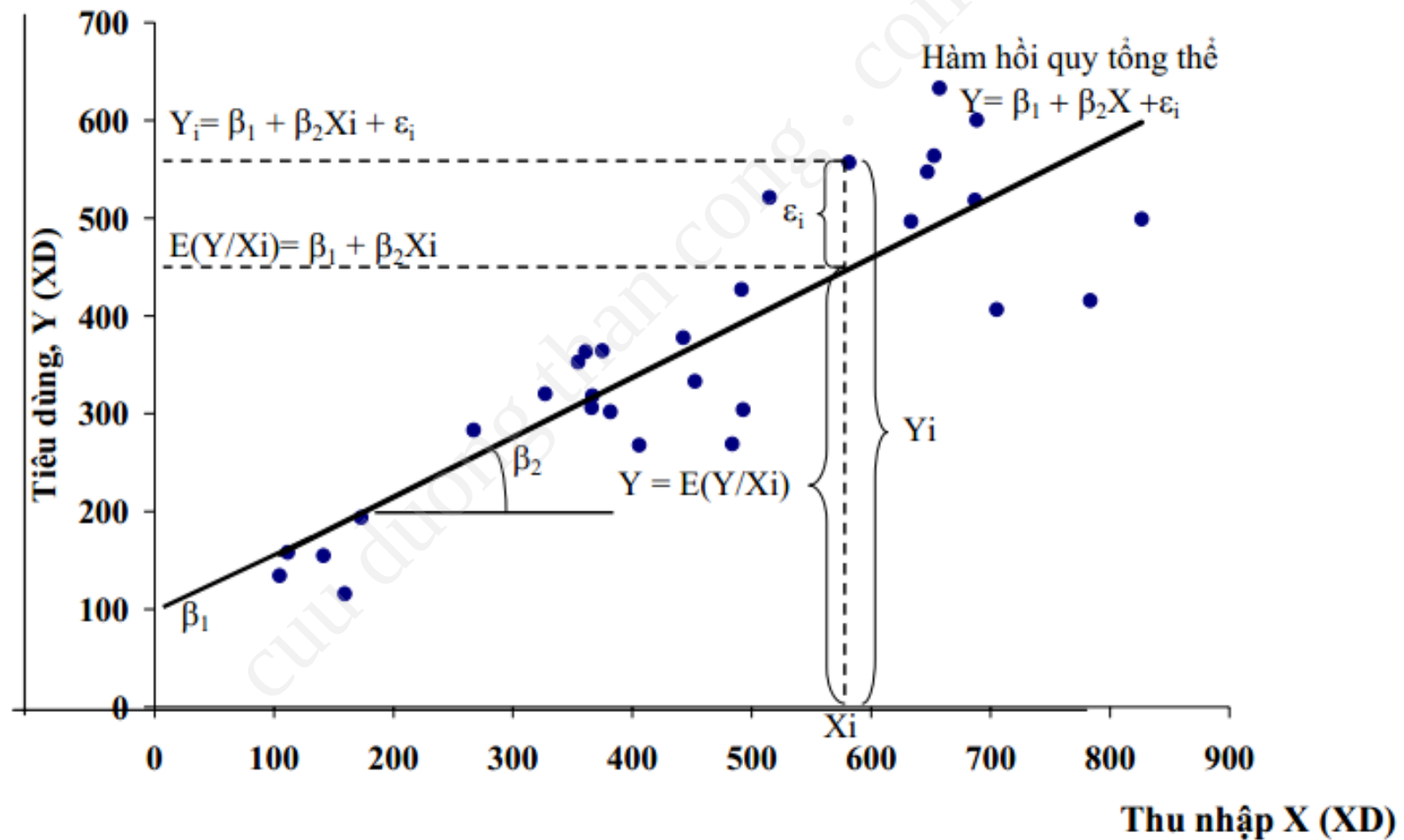
1.3. Hồi quy tổng thể

■ Nguyên nhân:

- Sự mập mờ về lý thuyết kinh tế
- Tâm quan trọng khác nhau giữa các biến giải thích
- Sự kém tin cậy của số liệu thống kê
- Khả năng chỉ định sai dạng hàm
- Sự tình cờ trong hành vi của con người đôi lúc mang tính ngẫu nhiên không tuân theo quy luật thường lệ.

1.3. Hồi quy tổng thể

Bảng 7. Mô tả hàm hồi qui tổng thể trên đồ thị



1.4. Hồi quy mẫu

1.4.1. Hàm hồi quy mẫu

- Giả sử từ tổng thể lập một mẫu ngẫu nhiên kích thước n :

$$W = \{(X_i, Y_i) ; i = 1 \div n\}$$

- Hàm hồi quy được xây dựng trên cơ sở của mẫu ngẫu nhiên được gọi là hàm hồi quy mẫu – SRF (Sample Regression Function). Về lý thuyết, hàm hồi quy mẫu được xây dựng tương tự như hàm hồi quy tổng thể. Hàm hồi quy mẫu là ước lượng của hàm hồi quy tổng thể.

1.4. Hồi quy mẫu

- Giả sử hàm hồi quy tổng thể là tuyến tính, thì hàm hồi quy mẫu cũng có dạng tuyến tính. SRF: $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$
Trong đó: $\hat{\beta}_1, \hat{\beta}_2$: là các hệ số hồi quy mẫu
 $\hat{Y}_i$: (Fitted value) là ước điểm của $E(Y/X_i)$.
- Khi mẫu chưa được chọn cụ thể thì các ước lượng $\hat{\beta}_1, \hat{\beta}_2$ là đại lượng ngẫu nhiên hay là biến ngẫu nhiên. Vì vậy, chúng có một quy luật phân phối xác suất và giữa chúng có hiệp phương sai.
- Khi mẫu được chọn cụ thể thì $\hat{\beta}_1, \hat{\beta}_2$ là con số cụ thể và là ước lượng điểm của tham số β_1, β_2 .

1.4. Hồi quy mẫu

1.4.2. Mô hình hồi quy mẫu (Sample Regression Model – SRM):

SRM:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i \quad (i = 1 \div n)$$

Trong đó: $e_i = Y_i - \hat{Y}_i$ gọi là phần dư (Residual).

- Phần dư e_i là sai số ngẫu nhiên của mẫu, thực chất chúng là các ước lượng điểm của các sai số ngẫu nhiên U_i trong tổng thể.
- Bản chất cũng như nguyên nhân tồn tại của e_i được giải thích giống như sai số ngẫu nhiên U_i .

Phân biệt tổng thể và mẫu

Tổng thể (Population)	Mẫu (Sample)
$PRF : E(Y / X_i) = \beta_1 + \beta_2 X_i$	$SRF : \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$
$PRM : Y_i = \beta_1 + \beta_2 X_i + U_i$ ($i = 1 \div N$)	$SRM : Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i$ ($i = 1 \div n$)
Sai số ngẫu nhiên U_i	Phần dư e_i

Các thuật ngữ cơ bản

Tiếng Anh	Tiếng Việt
Regression analysis	Phân tích hồi quy
Dependent variable	Biến phụ thuộc
Explanatory variable/ Independent variable	Biến giải thích/ biến độc lập
Time series data	Số liệu theo thời gian
Cross section data	Số liệu chéo
Pooled data	Số liệu kết hợp
Panel data	Số liệu bảng
Population	Tổng thể
PRF – Population Regression Function	Hàm hồi quy tổng thể
PRM - Population Regression Model	Mô hình hồi quy tổng thể

Các thuật ngữ cơ bản

Tiếng Anh	Tiếng Việt
Simple regression	Hồi quy đơn
Multiple regression	Hồi quy bội
Intercept coefficient	Hệ số chặn hoặc hệ số tự do
Slope coefficient	Hệ số góc
Random error	Sai số ngẫu nhiên
SRF – Sample Regression Function	Hàm hồi quy mẫu
SRM – Sample Regression Model	Mô hình hồi quy mẫu
Estimated regression coefficient	Các hệ số hồi quy ước lượng
Residual	Phần dư