

Chương V

Hồi quy và tương quan

Nội dung chính

1. Đặt vấn đề
2. Hồi quy và tương quan tuyến tính giữa 2 tiêu thức số lượng
3. Tương quan phi tuyến
4. Hồi quy đa biến

Một số khái niệm

- **Hồi quy:** là quy về trung bình (regression to mediocrity). Phân tích hồi quy nghiên cứu **quan hệ giữa một biến phụ thuộc với một hoặc nhiều biến độc lập khác.**
- **Tương quan:** phân tích tương quan nhằm **đo mức độ tuyến tính** giữa hai biến và các biến này có tính chất đối xứng.

Các mối quan hệ khác và giả thiết “*các yếu tố khác không đổi*”

- Mục tiêu của các nhà kinh tế học là xem xét liệu rằng một biến số này có *tác động đến* một biến khác.
- Do vậy, giả thiết “*các yếu tố khác không đổi*” thường được sử dụng trong nghiên cứu kinh tế.
- Hàm hồi quy và quan hệ nhân quả: Hồi quy không đòi hỏi mối quan hệ nhân quả giữa biến độc lập và biến phụ thuộc
- Hồi quy và tương quan: Tương quan đo mức độ kết hợp tuyến tính giữa hai biến, không có sự phân biệt các biến, có tính đối xứng; còn hồi quy ước lượng hoặc dự báo 1 biến trên cơ sở giá trị đã cho của các biến khác, hồi quy không có tính chất đối xứng

Các mối liên hệ

1. Liên hệ hàm số

- Là mối liên hệ hoàn toàn chặt chẽ. Sự thay đổi của hiện tượng này có tác dụng quyết định đến sự thay đổi của hiện tượng liên quan theo một tỷ lệ xác định.
- Liên hệ hàm số được viết dưới dạng: $y = f(x)$, có nghĩa là cứ mỗi giá trị của x thì theo một hàm nào đó có một giá trị của y tương ứng.
- Mối liên hệ này thường có trong tự nhiên. Ví dụ trong vật lý: $S = v.t...$
- Đặc điểm: liên hệ hàm số không những được biểu hiện ở tổng thể mà còn được biểu hiện trên từng đơn vị cá biệt.

Liên hệ tương quan

- Liên hệ tương quan: là mối liên hệ không hoàn toàn chặt chẽ. Sự thay đổi của hiện tượng này có thể làm hiện tượng liên quan thay đổi theo nhưng không có ảnh hưởng hoàn toàn quyết định.
- Mối liên hệ này rất phổ biến và thường gặp trong các hiện tượng KTXH.
- Đặc điểm: liên hệ tương quan không được biểu hiện trên từng đơn vị cá biệt mà phải thông qua hiện tượng số lớn (là tổng thể).

Ví dụ

- Mỗi liên hệ giữa tuổi nghề và NSLĐ. Tuổi nghề có tác động đến NSLĐ nhưng NSLĐ không chỉ chịu ảnh hưởng của tuổi nghề mà còn chịu ảnh hưởng của các nhân tố khác. Mặt khác, nếu nghiên cứu riêng lẻ từng đơn vị cá biệt, có những đơn vị tuổi nghề hoàn toàn không ảnh hưởng tới NSLĐ.
- Mỗi liên hệ giữa năng lực làm việc và bằng cấp

Nhiệm vụ của Hồi quy và tương quan

1. Xác định mô hình hồi qui phản ánh mối liên hệ
2. Đánh giá mức độ chặt chẽ của mối liên hệ tương quan

Các bước xác định mô hình hồi qui phản ánh mối liên hệ

- Bước 1: Giải thích sự tồn tại thực tế và bản chất của mối liên hệ bằng phân tích lý luận. Tùy theo mục đích nghiên cứu cụ thể mà xác định đâu là nguyên nhân (biến độc lập- nguyên nhân), đâu là kết quả (phụ thuộc- kết quả).
 - Ví dụ 1: Tuổi nghề có ảnh hưởng tới NSLĐ. Như vậy, tuổi nghề là nguyên nhân có ảnh hưởng đến NSLĐ.
 - NSLĐ và khối lượng sản phẩm? Khối lượng sản phẩm và giá thành?
- Phải chú ý đến mục đích nghiên cứu là gì để xác định đâu là tiêu thức nguyên nhân, đâu là tiêu thức kết quả
 - Ví dụ: chi phí quảng cáo và doanh thu?
- Trong mối liên hệ này, có thể có nhiều nguyên nhân nhưng chỉ có một kết quả mà thôi.

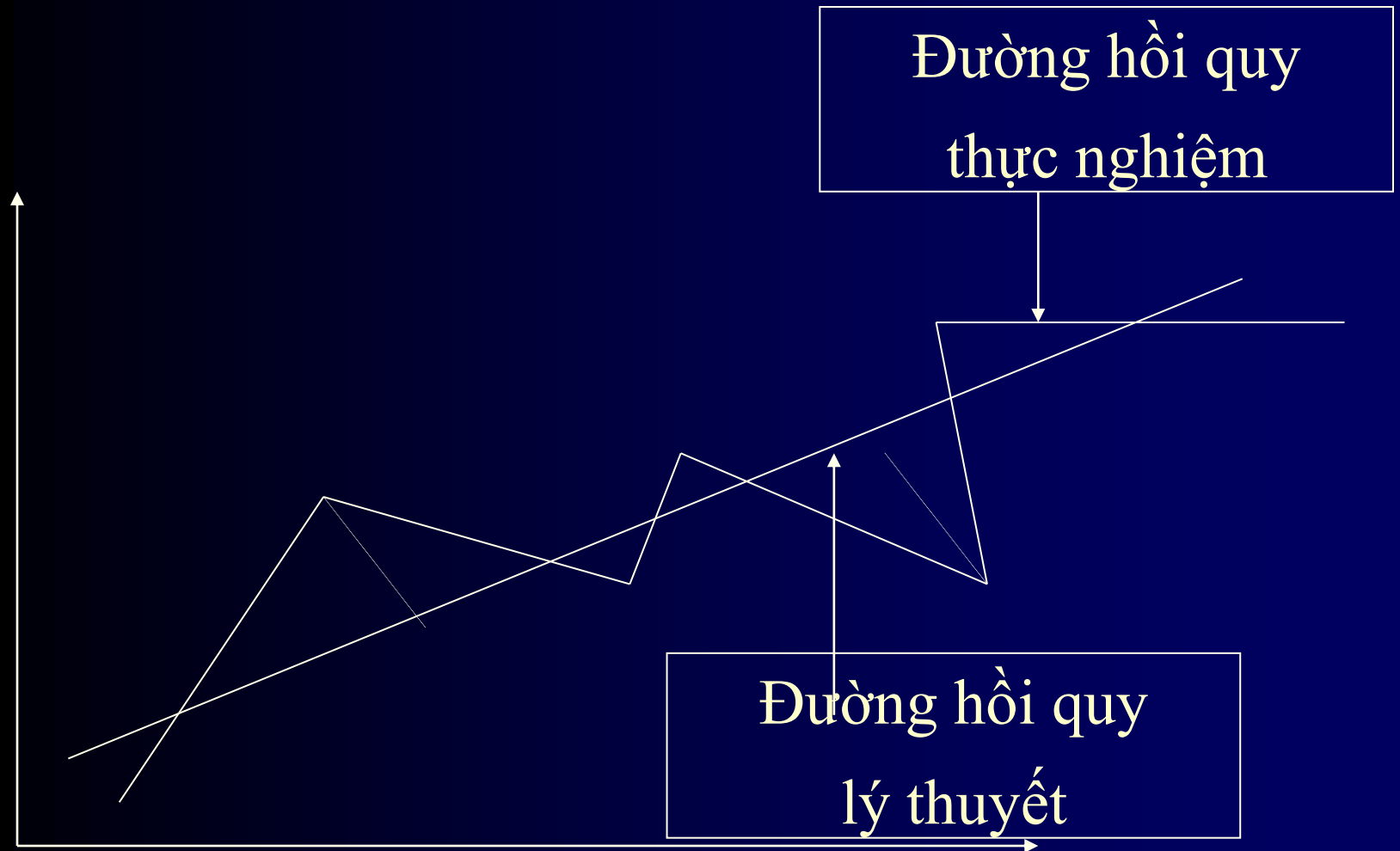
Xác định mô hình hồi quy phản ánh mối liên hệ

- Bước 2: Thăm dò mối liên hệ bằng các phương pháp thống kê: phương pháp đồ thị, phân tổ, số bình quân, phương pháp quan sát 2 dãy số song song...
- Bước 3: Lập phương trình/hàm hồi quy biểu hiện mối liên hệ.
- Ví dụ: các phương trình $y = a + bx$; $y = a + bx + cx^2$...
- Bước 4: Tính toán các tham số và giải thích ý nghĩa. Đánh giá mức độ (cường độ) chặt chẽ của liên hệ

Mô hình hồi quy

- **Đường hồi qui thực nghiệm:** là đường được hình thành bởi các tài liệu thực tế.
- **Đường hồi qui lý thuyết:** là đường điều chỉnh bù trừ các chênh lệch ngẫu nhiên vạch ra xu hướng cơ bản của hiện tượng.

Sơ đồ hóa



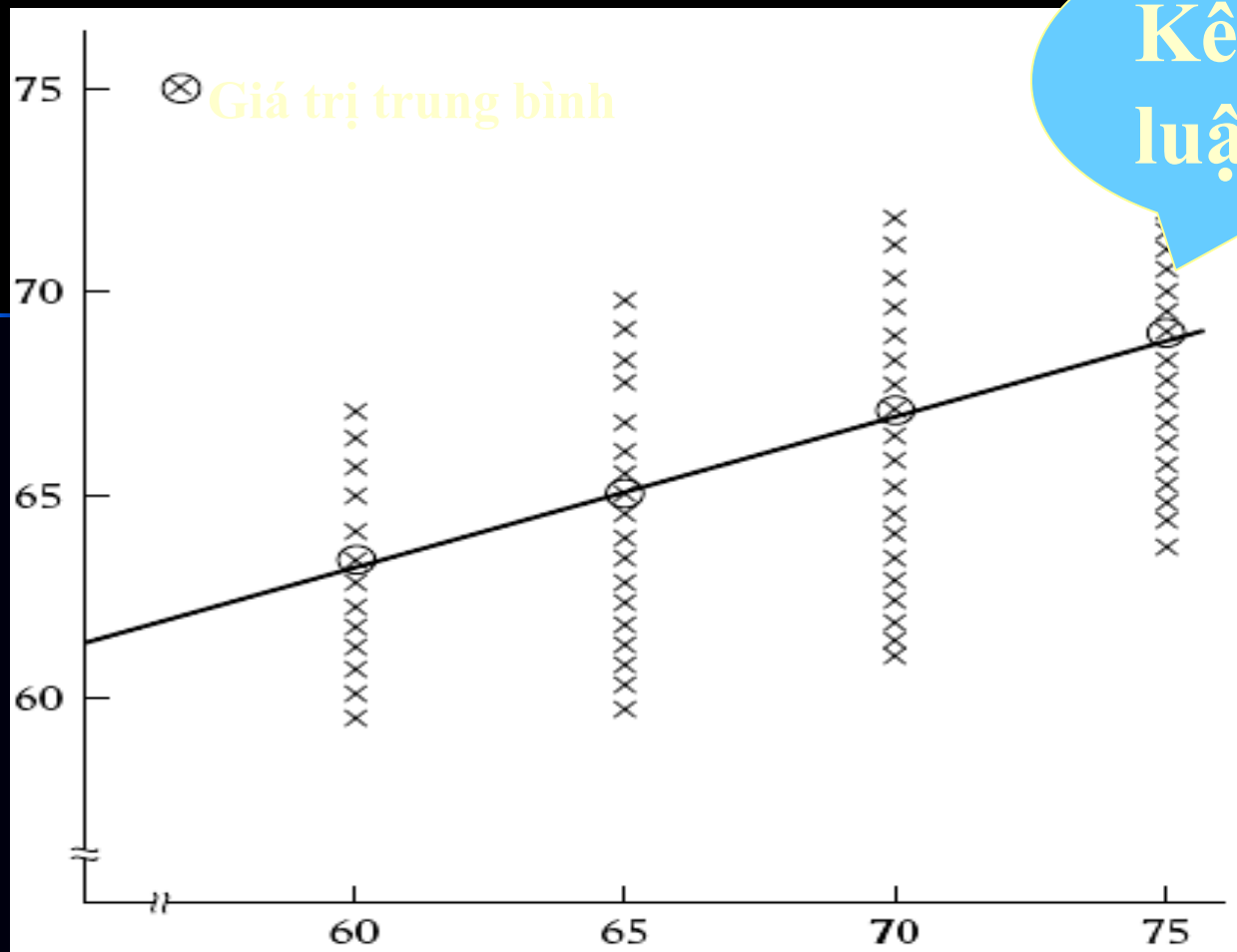
Phân tích hồi quy

- Phân tích hồi quy nghiên cứu sự phụ thuộc của một biến số, *biến phụ thuộc*, vào một hay nhiều biến số khác, *biến độc lập*, với ý định ước lượng và/hoặc dự đoán giá trị trung bình (tổng thể) của biến phụ thuộc dựa trên những giá trị đã biết hay cố định của biến độc lập.

Ví dụ 1

- Chúng ta quan tâm đến việc dự báo chiều cao trung bình của những người con khi biết chiều cao của người cha.
- Vẽ ***biểu đồ phân tán*** để biểu diễn phân phối chiều cao của những người con trong một tổng thể tương ứng với chiều cao của những người cha được cho trước hay cố ***định***
- ***Mục đích vẽ biểu đồ để làm gì?***

Chiều
cao
của
người
con
(tính
bằng
inch)

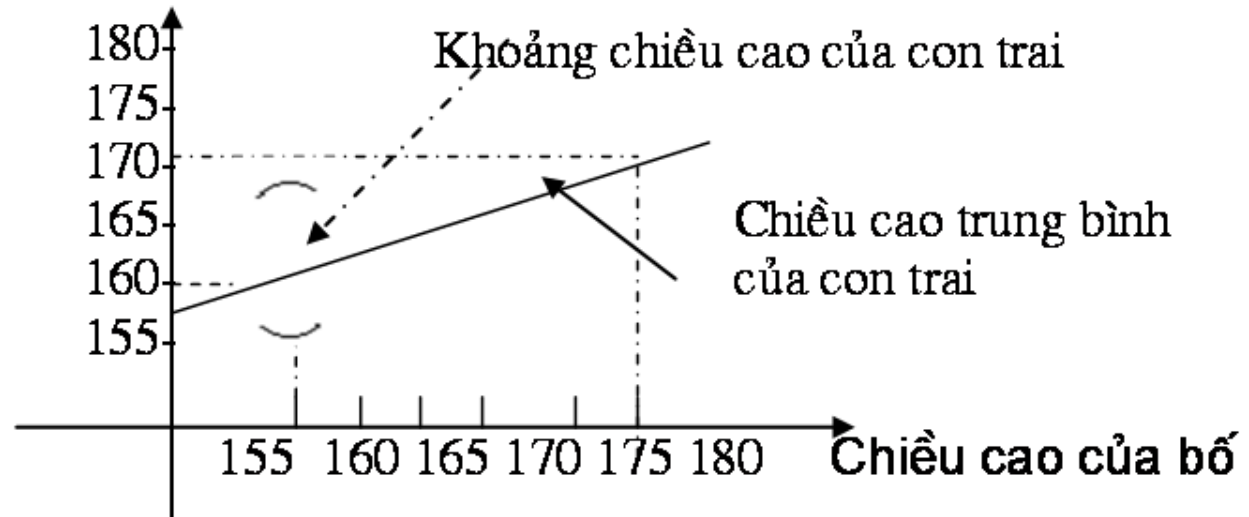


Kết
luận gì?

Chiều cao của người cha
(tính bằng inch)

Phân phối giả thiết của chiều cao của những người con trai
tương ứng với chiều cao của người cha được cho trước

Chiều cao của con trai



Từ đồ thị trên ta thấy:

- * Với chiều cao đã biết của bố thì chiều cao của con trai là một khoảng
- * Bố cao thì con trai cao
- * Chiều cao trung bình của các cậu con trai của những ông bố thấp lớn hơn chiều cao của bố
- * Chiều cao trung bình của các cậu con trai của những ông bố cao thì nhỏ hơn chiều cao của bố

Ví dụ 2:

Một nhà kinh tế nghiên cứu mối quan hệ giữa **lượng cầu của một loại hàng hóa** phụ thuộc vào: **giá của bản thân hàng hóa**, **thu nhập của người tiêu dùng**, **giá của một số loại hàng hóa cạnh tranh**.

Trong ví dụ này thì:

Lượng cầu loại hàng hóa này là biến phụ thuộc

Giá của bản thân hàng hóa	}	là các biến độc lập
Thu nhập của người tiêu dùng		
Giá của các hàng hóa cạnh tranh		

Và chúng ta có thể xây dựng mô hình hồi quy?

Mô hình hồi qui hai biến

- **Hàm hồi qui tổng thể** (population regression function – PRF) có dạng:

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

X_i là biến độc lập, nếu PRF có 1 biến độc lập thì được gọi là hàm hồi qui đơn (hồi qui hai biến), nếu có từ 2 biến độc lập trở lên được gọi là hàm hồi qui bội

- Hàm hồi qui tổng thể cho biết giá trị trung bình của biến Y sẽ thay đổi như thế nào khi biến X nhận các giá trị khác nhau.

Còn có những tên gọi khác cho biến độc lập, biến phụ thuộc

- Biến ngoại sinh, biến giải thích, biến hồi quy, biến dự báo
- Biến nội sinh, biến được giải thích, biến được hồi quy, biến được dự báo

Ví dụ

- Giả sử có một tổng thể gồm 60 hộ gia đình, có thu nhập (X) và chi tiêu (Y) hàng tuần như sau

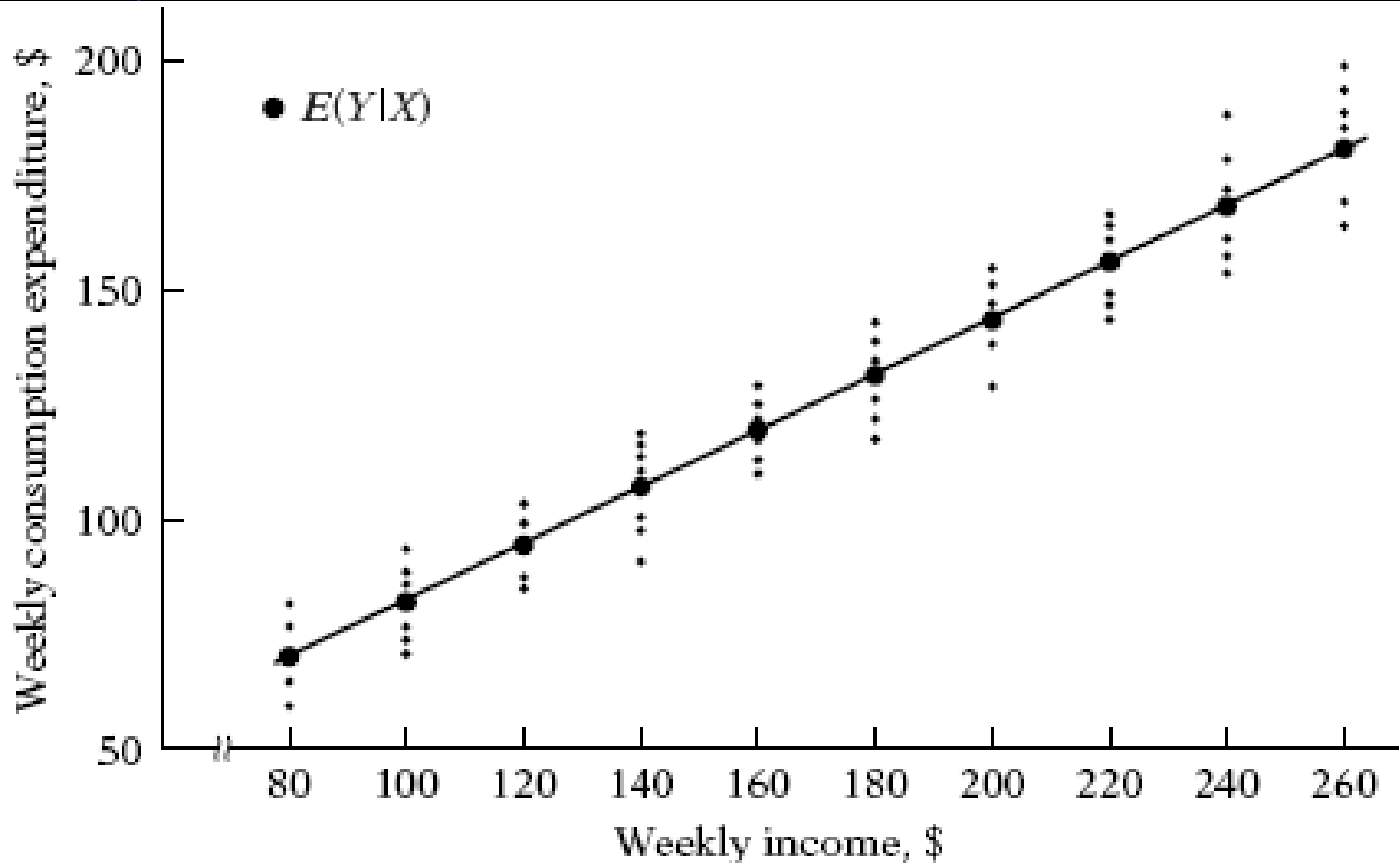
WEEKLY FAMILY INCOME X , \$

$Y \downarrow \quad X \rightarrow$	80	100	120	140	160	180	200	220	240	260
Weekly family consumption expenditure Y , \$	55	65	79	80	102	110	120	135	137	150
	60	70	84	93	107	115	136	137	145	152
	65	74	90	95	110	120	140	140	155	175
	70	80	94	103	116	130	144	152	165	178
	75	85	98	108	118	135	145	157	175	180
	–	88	–	113	125	140	–	160	189	185
	–	–	–	115	–	–	–	162	–	191
Total	325	462	445	707	678	750	685	1043	966	1211
Conditional means of Y , $E(Y X)$	65	77	89	101	113	125	137	149	161	173

Ví dụ (tiếp)

- Mặc dù có sự biến động lớn của Y ứng với mỗi giá trị của X , nhưng, ***một cách tổng quát***,
 $X \uparrow$ thì $Y \uparrow$
- Giá trị kỳ vọng của Y ứng với một giá trị nào đó của X được gọi là ***Giá trị kỳ vọng có điều kiện***, ký hiệu:
 $E(Y|X)$
- Ví dụ: $E(Y|X=80) = 65$; $E(Y|X=260) = 173$
- **Giá trị kỳ vọng không có điều kiện:**
 $E(Y) = 7273/60 = 121,20$

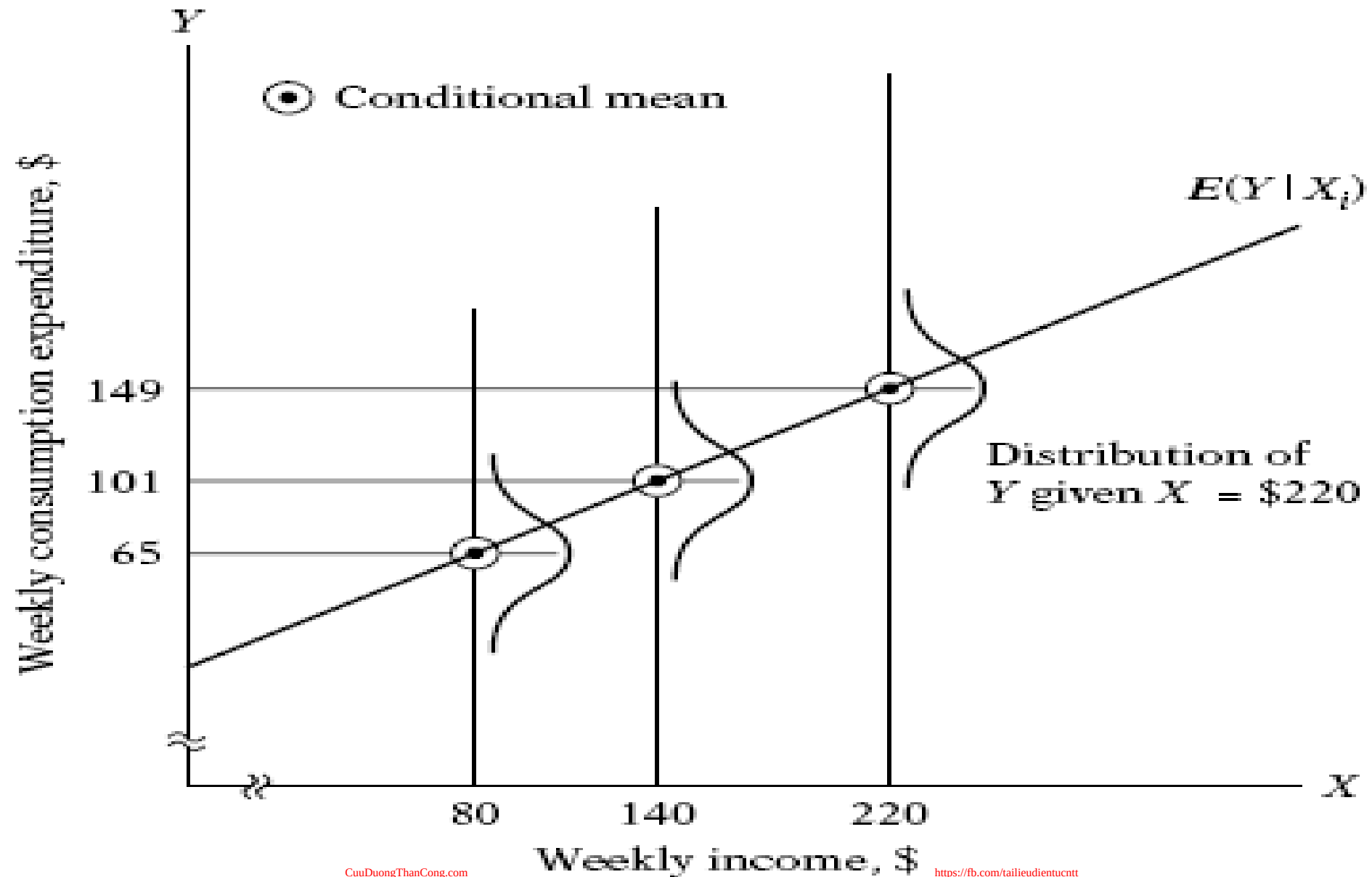
Phân phối có điều kiện của chi tiêu ứng với các mức thu nhập khác nhau



Hàm hồi quy tổng thể

- Đường nối các điểm tròn đen trong hình là **đường hồi quy tổng thể**, biểu diễn sự **hồi quy của Y vào X** .
- Về mặt hình học, một đường hồi quy tổng thể là tập hợp các giá trị trung bình có điều kiện của biến phụ thuộc ứng với mỗi giá trị cố định của biến giải thích.
- Ứng với mỗi giá trị của X , có một tổng thể các giá trị của Y , dao động xung quanh giá trị trung bình có điều kiện của Y .

Đường hồi quy tổng thể



Hàm hồi quy tuyến tính

- Vậy kỳ vọng có điều kiện $E(Y|X_i)$ là một hàm số của X_i :

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

- Dạng hàm $f(X_i)$ phụ thuộc vào các mối quan hệ kinh tế (thường được xác định dựa vào các lý thuyết kinh tế).
- Ở đây, ta thường sử dụng hàm số tuyến tính:

Mô hình hồi qui hai biến tuyến tính

$$E(Y/X_i) = b_0 + b_1 X_i$$

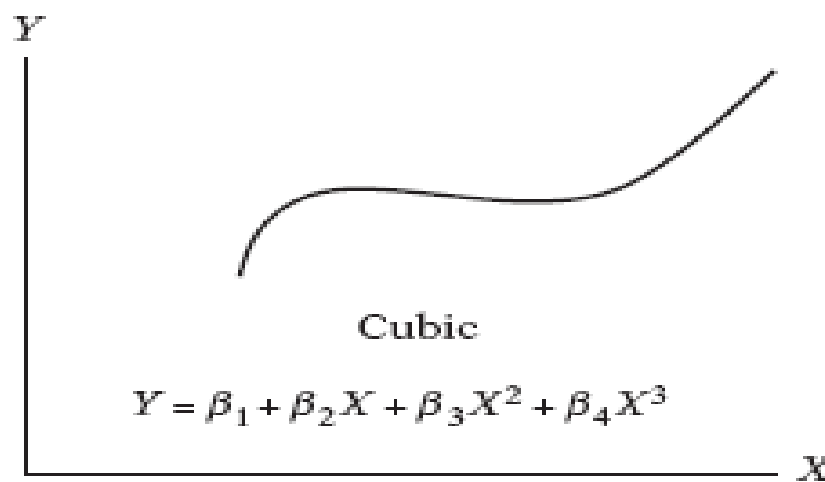
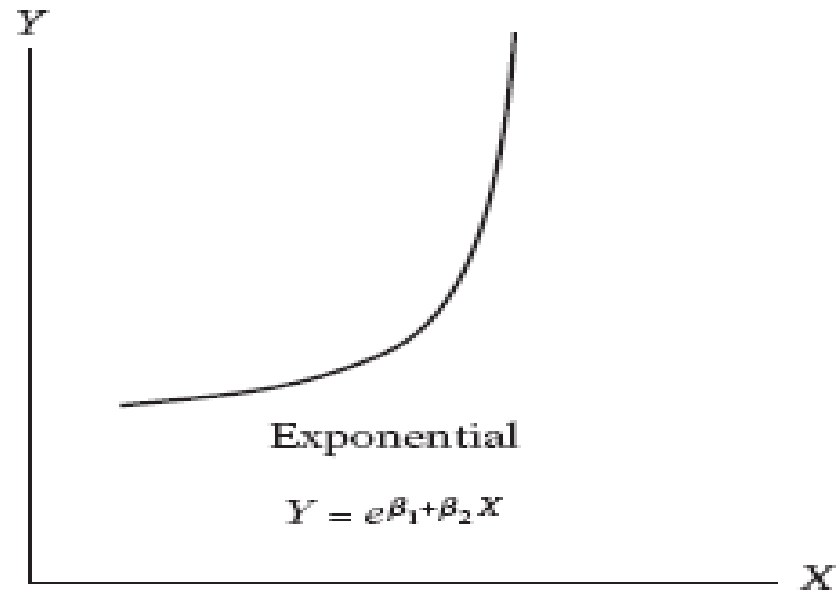
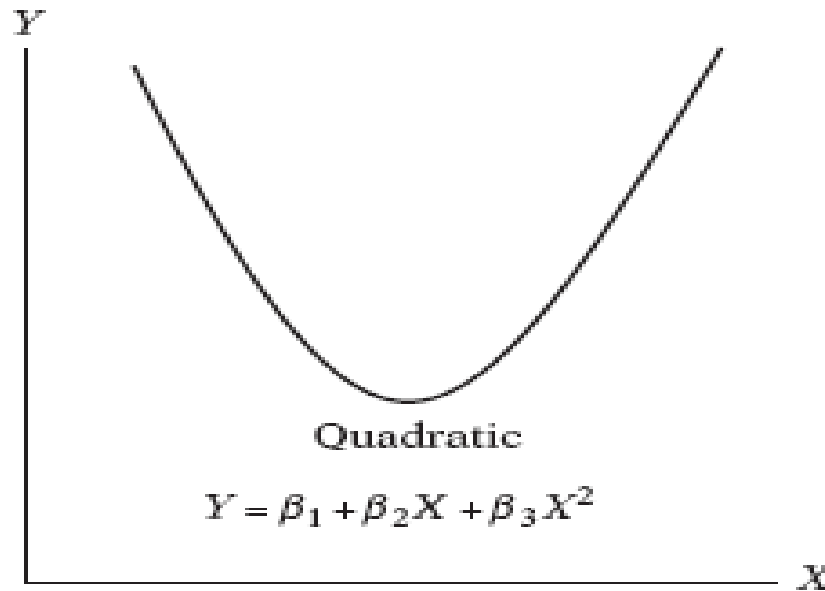
Trong đó b_0 , b_1 là các tham số chưa biết nhưng cố định – các tham số hồi qui.

- b_0 là hệ số tự do, cho biết giá trị trung bình của biến phụ thuộc Y sẽ thay đổi như thế nào khi biến X nhận giá trị 0.
- b_1 là hệ số góc, cho biết giá trị trung bình của biến phụ thuộc Y sẽ thay đổi (tăng / giảm) bao nhiêu đơn vị khi giá trị của biến độc lập X tăng 1 đơn vị với điều kiện các yếu tố khác không thay đổi.

Mô hình hồi qui hai biến tuyến tính

- Thuật ngữ “tuyến tính” ở đây được hiểu theo hai nghĩa: tuyến tính đối với tham số và tuyến tính đối với biến độc lập.
- $E(Y/X_i) = b_0 + b_1 X_i$ là *tuyến tính tham số*
- $E(Y/X_i) = b_0 + b_1^2 X_i$ là *tuyến tính biến số*.
- Hàm hồi qui tuyến tính luôn được hiểu là tuyến tính đối với tham số, nó có thể không tuyến tính đối với biến.

Các hàm số tuyến tính đối với tham số



Linear-in-parameter functions.

Mô hình hồi qui hai biến tuyến tính

- Ứng với mỗi giá trị của X , giá trị Y của một số quan sát có độ lệch so với giá trị kỳ vọng.
- Giá trị quan sát thứ i của biến phụ thuộc Y được ký hiệu là Y_i .
- Ký hiệu U_i là chênh lệch giữa Y_i và $E(Y/X_i)$
 - $U_i = Y_i - E(Y/X_i)$
 - hay $Y_i = E(Y/X_i) + U_i$ (dạng ngẫu nhiên PRF)
- U_i được gọi là đại lượng ngẫu nhiên hay sai số ngẫu nhiên

Mô hình hồi qui hai biến tuyến tính

- Trong thực tế, ta thường phải ước lượng các hệ số hồi quy của tổng thể từ hệ số hồi quy của mẫu.
- **Hàm hồi qui mẫu** (sample regression function – SRF): sử dụng khi chúng ta không thể lấy tất cả thông tin từ tổng thể mà chỉ thu thập được từ các mẫu riêng lẻ từ tổng thể.
- Nếu hàm PRF có dạng tuyến tính ($E(Y/X_i) = b_0 + b_1 X_i$), ta có SRF:

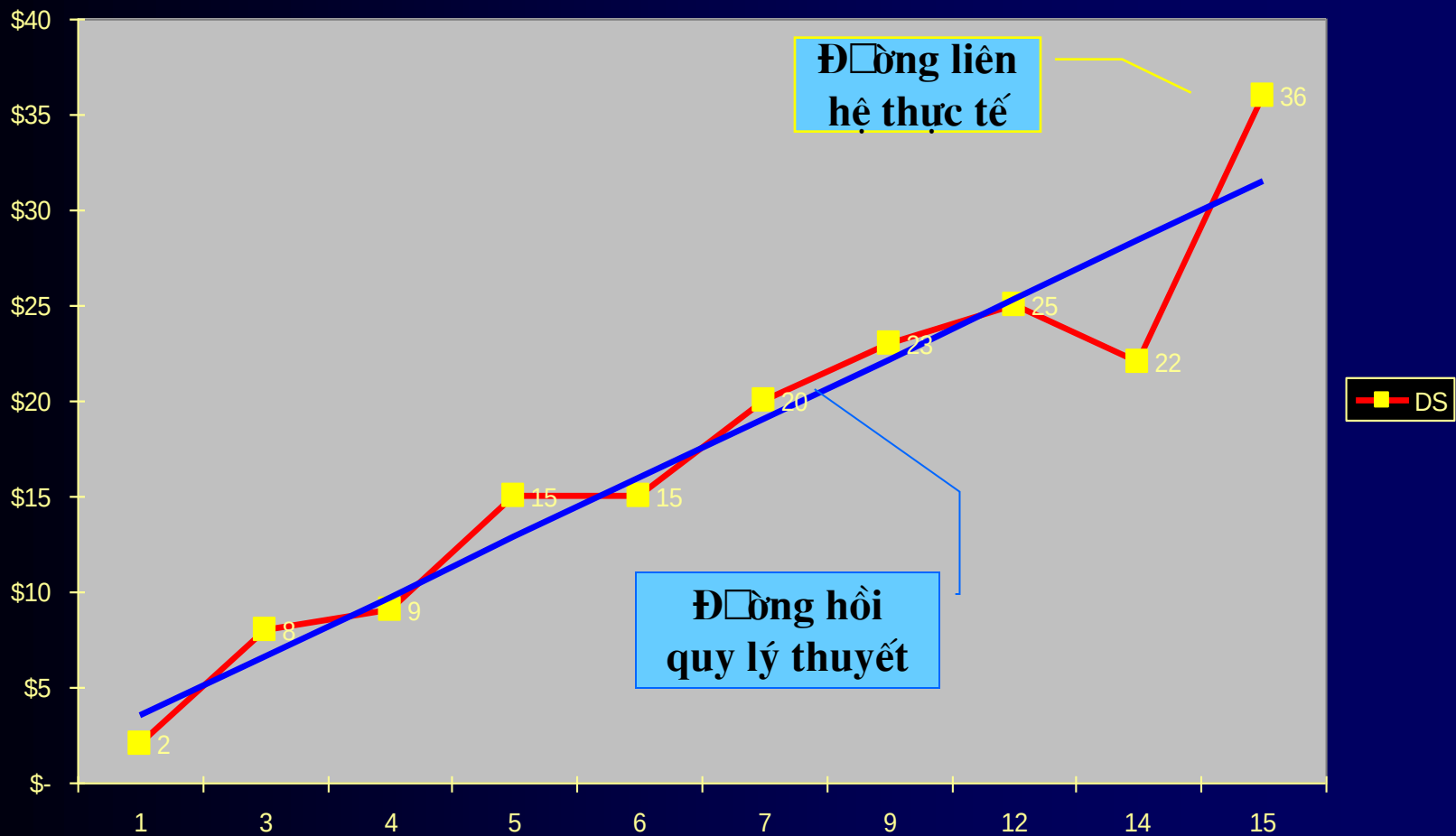
$$\hat{Y}_i = \hat{b}_0 + \hat{b}_1 X_i$$

trong đó $\hat{Y}_i$ là ước lượng điểm của $E(Y/X_i)$
 $\hat{b}_0$ là ước lượng điểm của b_0 ;
 $\hat{b}_1$ là ước lượng điểm của b_1 ;

Xét ví dụ: theo dõi liên hệ giữa chi phí quảng cáo (CPQC) (nghìn USD) và doanh số (DS) (nghìn sp) của một mặt hàng mới

CP QC (\$)	1	3	4	5	6	7	9	12	14	15
DS (ngh sp)	2	8	9	15	15	20	23	25	22	36

➤ Biểu diễn mối liên hệ giữa 2 tiêu thức



Biến độc lập: CP quảng cáo (x)

Biến phụ thuộc: doanh số (y)

- Có thể xây dựng hàm hồi quy mẫu hai biến như sau: $\hat{y}_x = b_0 + b_1x$
- Sử dụng các tham thống nhất với giáo trình
- Trong đó:
 - x: trị số của tiêu thức gây ảnh hưởng (nguyên nhân) (biến độc lập).
 - $\hat{y}_x$: trị số điều chỉnh/ước lượng của tiêu thức chịu ảnh hưởng (kết quả) (biến phụ thuộc) theo quan hệ với x.
 - b_0 : hệ số tự do (hệ số chặn), là điểm xuất phát của đường hồi quy lý thuyết, nêu lên ảnh hưởng của các nhân tố khác (tiêu thức nguyên nhân khác) ngoài x tới sự biến động của y.

- b_1 : hệ số hồi quy (hệ số góc, độ dốc), phản ánh ảnh hưởng trực tiếp của tiêu thức nguyên nhân x đến tiêu thức kết quả y . Mỗi khi x tăng lên 1 đơn vị thì y sẽ thay đổi trung bình b_1 đơn vị.
 - b_1 nói lên chiều hướng của mối liên hệ:
 - $b_1 > 0$: mối liên hệ thuận
 - $b_1 < 0$: mối liên hệ nghịch

- b_0 , b_1 phải được xác định sao cho đường hồi quy lý thuyết mô tả gần đúng nhất mối liên hệ thực tế. Trên hình vẽ, khoảng cách từ điểm thực tế đến điểm thuộc đường hồi quy lý thuyết nhỏ nhất sẽ là tốt nhất.
- Sử dụng phương pháp bình phương nhỏ nhất (OLS-Ordinary Least Square) với nội dung: tổng bình phương các độ lệch giữa giá trị thực tế và giá trị lý thuyết của biến phụ thuộc (tiêu thức kết quả) là nhỏ nhất.

$$S = \sum (y_i - \hat{y}_x)^2 = \min$$

$$\text{Hay } S = \sum (y_i - b_0 - b_1 x)^2 = \min$$

Tại sao lại là tổng bình phương độ lệch?

Để thỏa mãn điều kiện trên, cần tính đạo hàm riêng theo 2 tham số cần tìm. Từ đó, b_0 và b_1 phải thỏa mãn hệ phương trình:

$$\begin{cases} \sum y = nb_0 + b_1 \sum x \\ \sum xy = b_0 \sum x + b_1 \sum x^2 \end{cases}$$

- Giải phương trình trên theo 2 hệ số ta được:

$$\begin{cases} b_1 = \frac{\overline{xy} - \bar{x} \times \bar{y}}{\sigma_x^2} \\ b_0 = \bar{y} - b_1 \times \bar{x} \end{cases}$$

Áp dụng cho VD trên

x	y	xy	x^2	y^2
1	2	2	1	4
3	8	24	9	64
4	9	36	14	81
5	15	75	25	225
6	15	90	36	225
7	20	140	49	400
9	23	207	81	529
12	25	300	144	625
14	22	308	196	484
15	36	540	225	1296
76	175	1722	782	3933

Giải hệ phương trình

- Thay số:

$$\begin{cases} 175 = 10b_0 + 76b_1 \\ 1722 = 76b_0 + 782b_1 \end{cases}$$

- Giải hệ:

$$\begin{cases} b_0 = 2,92 \\ b_1 = 1,92 \end{cases}$$

- *Kết luận gì về các hệ số ?*

— b_0

— b_1

■ Kiểm tra sự phù hợp của các hệ số?

- So sánh sự thay đổi của biến phụ thuộc với mỗi giá trị của biến độc lập để xác định dấu của hệ số góc, so sánh với kết quả tính toán
- So sánh giữa giá trị ước lượng và giá trị thực tế (Y) xem có gần nhau không?

Đánh giá mức độ chặt chẽ của mối liên hệ

- Sử dụng hệ số tương quan r:
- Hệ số tương quan là chỉ tiêu đánh giá mức độ chặt chẽ của mối liên hệ tương quan tuyến tính đơn.

$$r = b_1 \frac{\sigma_x}{\sigma_y}$$

$$r = \frac{\overline{xy} - \bar{x} \times \bar{y}}{\sigma_x \times \sigma_y}$$

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Tác dụng của hệ số tương quan - r

- Xác định cường độ của mối liên hệ từ đó chọn ra nguyên nhân chủ yếu hoặc thứ yếu đối với hiện tượng nghiên cứu.
- Xác định chiều hướng cụ thể của mối liên hệ (thuận - nghịch).
- Hệ số tương quan còn dùng trong nhiều trường hợp dự đoán thống kê và tính sai số của dự đoán.

Ý nghĩa của hệ số tương quan

■ Biểu thị cường độ của liên hệ

- $r = \pm 1 \Rightarrow$ liên hệ hoàn toàn chặt chẽ (hàm số)
- $|r| \rightarrow 1 \Rightarrow$ liên hệ càng chặt chẽ
- $r = 0 \Rightarrow$ không có liên hệ
- ≥ 0.9 : mối liên hệ rất chặt chẽ
- $0.7 \leq |r| \leq 0.9$: mối liên hệ tương đối chặt chẽ
- $0.5 \leq |r| \leq 0.7$: mối liên hệ bình thường
- $|r| < 0.5$: mối liên hệ hết sức lỏng lẻo

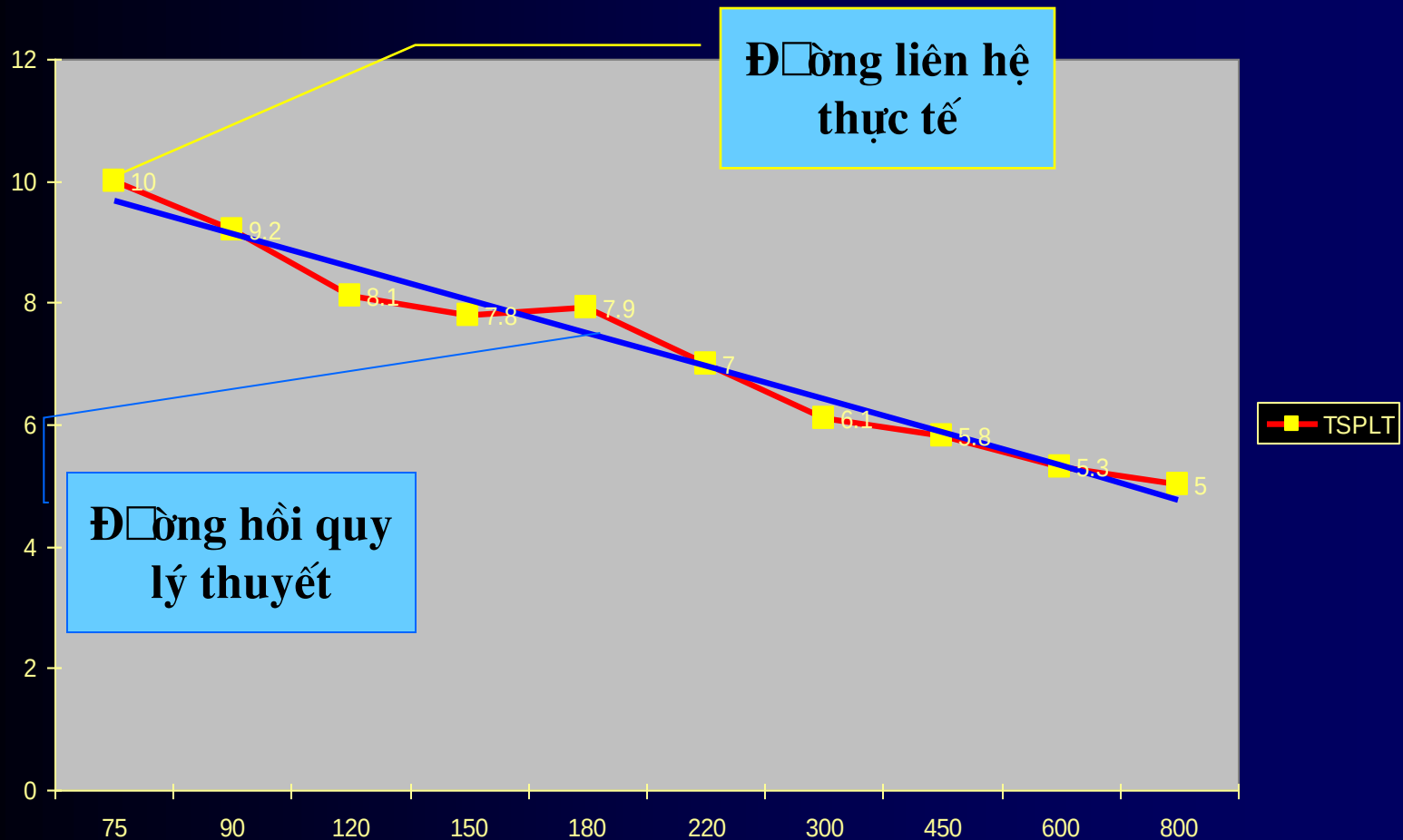
■ Biểu hiện tính chất của liên hệ

- $r > 0 \Rightarrow$ tương quan thuận
- $r < 0 \Rightarrow$ tương quan nghịch

Bài tập thực hành 1

Mức tiêu thụ hàng hoá (trVND)	75	90	120	150	180	220	300	450	600	800
Tỷ suất phí lưu thông (%)	10.0	9.2	8.1	7.8	7.9	7.0	6.1	5.8	5.3	5.0

➤ Biểu diễn mối liên hệ giữa 2 tiêu thức



Xác định mô hình hồi quy

- Hồi quy tuyến tính, mô hình có dạng:
- $y = b_0 + b_1x$
- trong đó: x – mức tiêu thụ hàng hóa là tiêu thức nguyên nhân (biến độc lập)
- y : tỷ suất phí lưu thông là biến phụ thuộc hay kết quả
- b_0 : tham số tự do
- b_1 : hệ số hồi quy tuyến tính

	x	y	xy	x^2	y^2
	75	10.0	750	5625	100.00
	90	9.2	828	8100	84.64
	120	8.1	972	14400	65.61
	150	7.8	1170	22500	60.84
	180	7.9	1422	32400	62.41
	220	7.0	1540	48400	49.00
	300	6.1	1830	90000	37.21
	450	5.8	2610	202500	33.64
	600	5.3	3180	360000	28.09
	800	5.0	4000	640000	25.00
Σ	=2985	72.2	18302	1423925	546.44

Xác định giá trị của b_0 , b_1

$$\begin{cases} \sum y = nb_0 + b_1 \sum x \\ \sum xy = b_0 \sum x + b_1 \sum x^2 \end{cases}$$

$$\begin{cases} 72.2 = 10b_0 + 2985b_1 \\ 18302 = 2985b_0 + 1423925b_1 \end{cases} \Rightarrow \begin{cases} b_0 = 9.04 \\ b_1 = -0.0061 \end{cases}$$

Mô hình hồi quy có dạng:

$$\hat{y}_x = 9.04 - 0,0061x$$

ý nghĩa của b_0 và b_1 ?

Đánh giá trình độ chặt chẽ của liên hệ

$$r = \frac{\overline{xy} - \bar{x} \times \bar{y}}{\sigma_x \times \sigma_y} = \frac{\overline{xy} - \bar{x} \times \bar{y}}{\sqrt{\frac{\sum x^2}{n} - \left(\bar{x}\right)^2} \times \sqrt{\frac{\sum y^2}{n} - \left(\bar{y}\right)^2}}$$

$$r = \frac{1830,2 - 298,5 \times 7,22}{\sqrt{\frac{1423925}{10} - (298,5)^2} \times \sqrt{\frac{54,644}{10} - (7,22)^2}} = -0,888$$

Kết luận gì về r

Hệ số xác định R^2 : một thước đo Độ tin cậy của mô hình

- Gọi TSS (Tổng bình phương sai số tổng cộng):

$$TSS = \sum (Y_i - \bar{Y})^2$$

- ESS: bình phương sai số được giải thích bởi mô hình

$$ESS = \sum (\hat{Y}_i - \bar{Y})^2$$

- RSS: tổng bình phương sai số/phần dư:

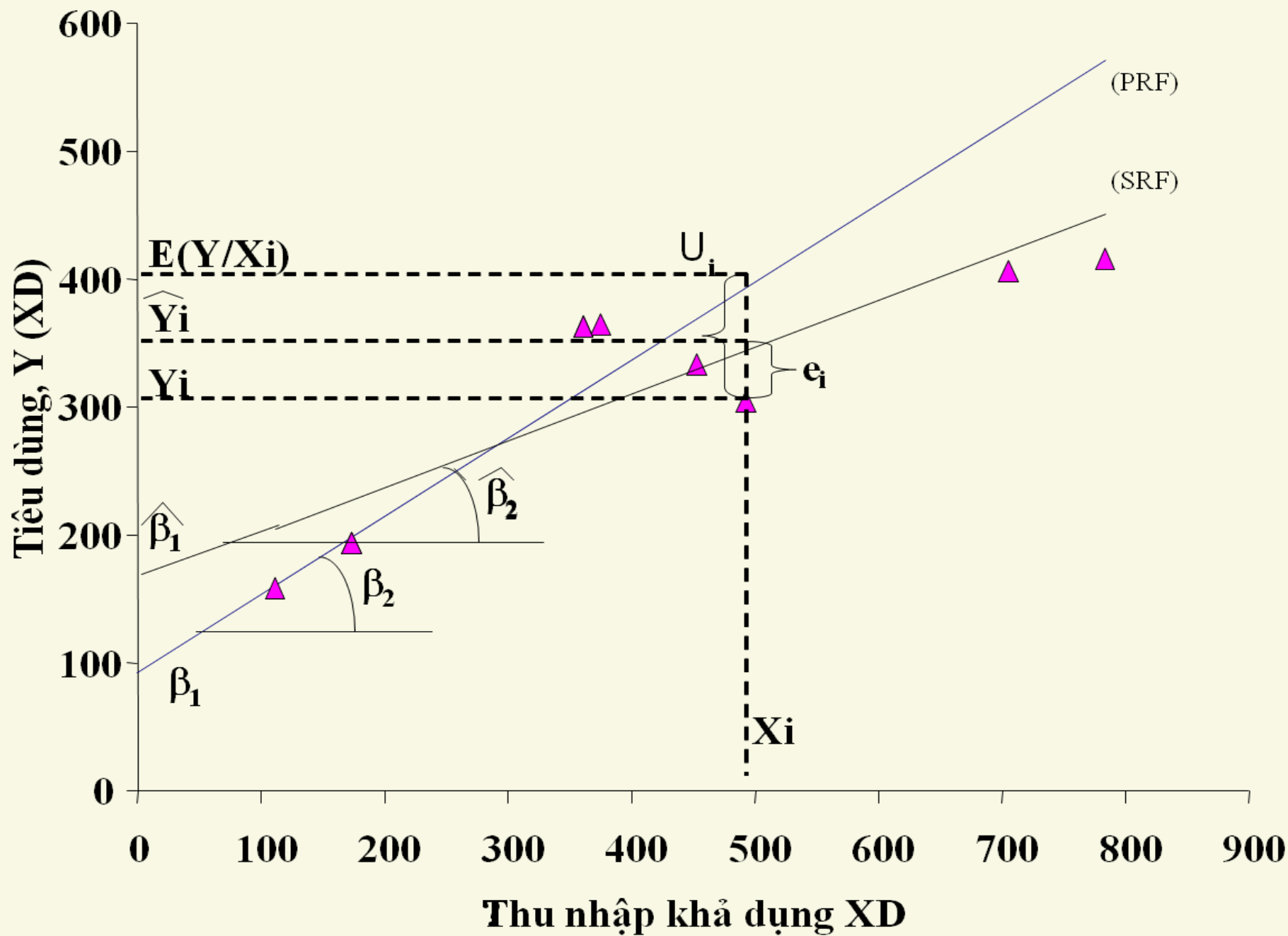
$$RSS = \sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2$$

- Ta chứng minh được: $TSS = ESS + RSS$

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

Hệ số xác định R^2

- R^2 cho biết % sự biến động của Y được giải thích bởi các biến số X trong mô hình.
- $0 < R^2 < 1$
- $R^2 \rightarrow 1$: mô hình giải thích được càng nhiều sự biến động của $Y \rightarrow$ mô hình càng đáng tin cậy.
- Một nhược điểm của R^2 là giá trị của nó tăng khi số biến X đưa vào mô hình tăng, bất chấp biến đưa vào không có ý nghĩa.
- Cần sử dụng R^2 điều chỉnh (adjusted $R^2 - \bar{R}^2$) để quyết định việc đưa thêm biến vào mô hình.



2. Liên hệ tương quan tuyến tính đa biến

- Nghiên cứu mối liên hệ giữa nhiều tiêu thức nguyên nhân với một tiêu thức kết quả

- Hàm số:

$$\hat{y}_x = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p$$

- $b_1, \dots, b_p$ là các hệ số hồi quy riêng, nói lên chiều hướng (thuận hay nghịch) tác động của tiêu thức nguyên nhân x_i đến tiêu thức kết quả y

Hệ phương trình xác định giá trị tham số

$$\sum y = b_0 n + b_1 \sum x_1 + b_2 \sum x_2 + \dots + b_n \sum x_n$$

$$\sum x_1 y = b_0 \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2 + \dots + b_n \sum x_1 x_n$$

...

$$\sum x_n y = b_0 \sum x_n + b_1 \sum x_1 x_n + b_2 \sum x_2 x_n + \dots + b_n \sum x_n^2$$

- Vấn đề đặt ra là mức độ ảnh hưởng của từng tiêu thức nguyên nhân đến tiêu thức kết quả như thế nào?

Hệ số tương quan bội

Hệ số tương quan tuyến tính bội là chỉ tiêu dùng để đánh giá mức độ chặt chẽ của mối liên hệ tương quan giữa tiêu thức kết quả y với tất cả các tiêu thức nguyên nhân x_i được nghiên cứu

$$R = \sqrt{1 - \frac{\sum (y - \hat{y}_x)^2}{\sum (y - \bar{y})^2}}$$

$$R = \sqrt{\frac{\sigma_y^2 - \sigma_{y(x_1, x_2, x_3, \dots, x_n)}^2}{\sigma_y^2}}$$

$$R = \sqrt{\frac{\sigma_{y x_1, x_2, x_3, \dots, x_n}^2}{\sigma_y^2}}$$

- R nằm trong khoảng $[0,1]$.
- $R = 0$: không có mối liên hệ tương quan tuyến tính bội.
- $R = 1$: mối liên hệ tương quan hoàn toàn chặt chẽ.
- R càng gần 1, mối liên hệ càng chặt chẽ.

Giải thích ý nghĩa các ký hiệu trong CT

- σ_y^2 = phương sai của tiêu thức kết quả/phụ thuộc y
- $\sigma_{y(x)}^2$ = phương sai của tt kết quả y do ảnh hưởng của các tt nguyên nhân khác ngoài x
- σ_{yx}^2 = phương sai của tt kết quả y do ảnh hưởng của x

$$\sigma_{y(x_1x_1....x_n)} = \sqrt{\frac{\sum (y - \hat{y}_x)^2}{n}}$$

$$\sigma_y = \sqrt{\frac{\sum (y - \bar{y})^2}{n}}$$

$$\sigma_{yx_1x_2...x_n}^2 = \frac{\sum (\hat{y} - \bar{y})^2}{n}$$

Hệ số hồi quy chuẩn hoá

để đánh giá mức độ ảnh hưởng của từng tiêu thức nguyên nhân tới tiêu thức kết quả

$$\beta_j = \frac{b_j \times \sigma_{x_j}}{\sigma_y}$$

- Với: β_j : tham số tương quan chuẩn hoá của tiêu thức nguyên nhân j
- Dấu của beta j phụ thuộc vào dấu của b_j , b_j phản ánh chiều hướng tác động của x_j đến y .
- Trị tuyệt đối của beta j nói lên mức độ ảnh hưởng của tiêu thức nguyên nhân x_j đến tiêu thức kết quả y

Bài tập thực hành 2

Tiền lương (\$/CN)	1,00	1,20	1,25	1,27	1,30	1.32	1.35
Chi phí NVL (\$/sp)	5,0	5,3	5,5	6,0	6,7	6,9	7,0
Giá thành (\$/sp)	6,5	6,9	7,2	7,8	8,3	8,9	9,2

Phân tích

- Tt nguyên nhân 1: tiền lương CN: x_1
- Tt nguyên nhân 2: chi phí NVL: x_2
- Tt kết quả: giá thành sản phẩm: y
- Hàm hồi quy tuyến tính đa biến có dạng:

$$y = b_0 + b_1x_1 + b_2x_2$$

x_1	x_2	y	x_1x_2	x_1^2	x_2^2	x_1y	x_2y
1.00	5.0	6.5	5.000	1.0000	25.00	6.500	32.50
1.20	5.3	6.9	6.360	1.4400	28.09	8.280	36.57
1.25	5.5	7.2	6.875	1.5625	30.25	9.000	39.60
1.27	6.0	7.8	7.620	1.6129	36.00	9.906	46.80
1.30	6.7	8.3	8.710	1.6900	44.89	10.790	55.61
1.32	6.9	8.9	9.108	1.7424	47.61	11.748	61.41
1.35	7.0	9.2	9.450	1.8225	49.00	12.420	64.40
8.69	42.4	54.8	53.123	10.8703	260.84	68.644	336.89

Xác định giá trị các tham số b_0, b_1, b_2

$$\begin{cases} \sum y = nb_0 + b_1 \sum x_1 + b_2 \sum x_2 \\ \sum x_1 y = b_0 \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2 \\ \sum x_2 y = b_0 \sum x_2 + b_1 \sum x_1 x_2 + b_2 \sum x_2^2 \end{cases}$$

$$\begin{cases} b_0 = 0.06323 \\ b_1 = 0.56831 \\ b_2 = 1.16554 \end{cases}$$

- Phương trình hồi quy tuyến tính đa biến có dạng:

$$y = 0.06323 + 0.56831x_1 + 1.16554x_2$$

Đánh giá trình độ chặt chẽ của liên hệ

- Dùng hệ số tương quan bội R
- Dùng hệ số tương quan chuẩn hoá β

x_1	x_2	y	$\hat{Y}_i$	$\sum (y - \hat{y}_x)^2$	$\sum (y - \bar{y})^2$
1.00	5.0	6.5	6.459	0.002	1.765
1.20	5.3	6.9	6.923	0.001	0.862
1.25	5.5	7.2	7.184	0.000	0.395
1.27	6.0	7.8	7.778	0.000	0.001
1.30	6.7	8.3	8.611	0.097	0.222
1.32	6.9	8.9	8.856	0.002	1.148
1.35	7.0	9.2	8.989	0.044	1.881
8.69	42.4	54.8	54.800	0.146	6.274

Hệ số tương quan bội

$$R = \sqrt{1 - \frac{\sum (y - \hat{y}_x)^2}{\sum (y - \bar{y})^2}}$$

$$R = \sqrt{1 - \frac{0.146}{6.274}} = 0,988$$

Hệ số hồi quy chuẩn hoá

$$\beta_j = \frac{b_j \times \sigma_{x_j}}{\sigma_y}$$

$$\beta_1 = \frac{b_1 \times \sigma_{x_1}}{\sigma_y} = 0.56831 \times \sqrt{\frac{\sum x_1^2 - 7\bar{x}_1^2}{\sum y^2 - 7\bar{y}^2}} = +0.0651$$

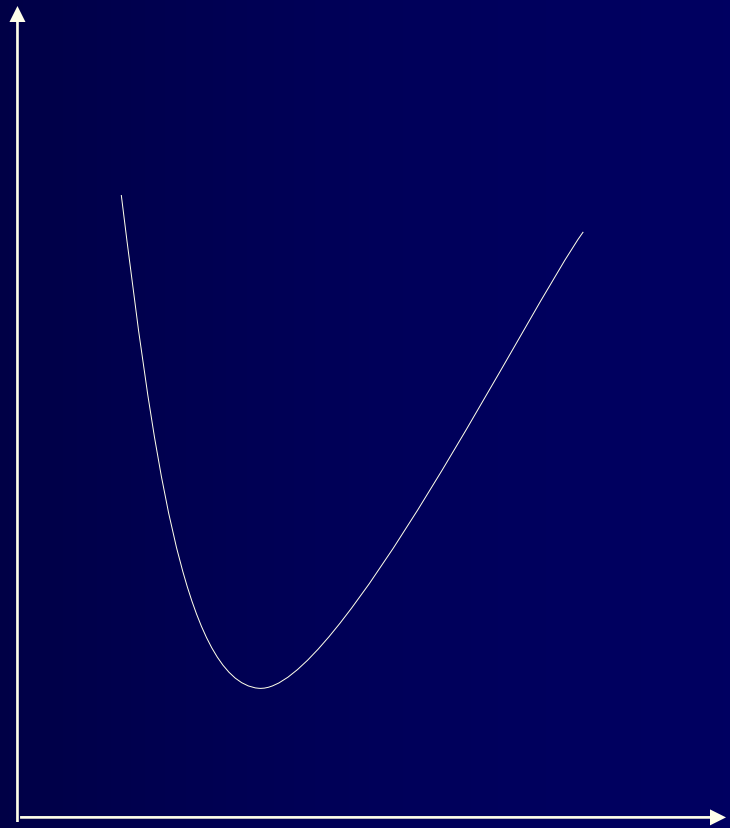
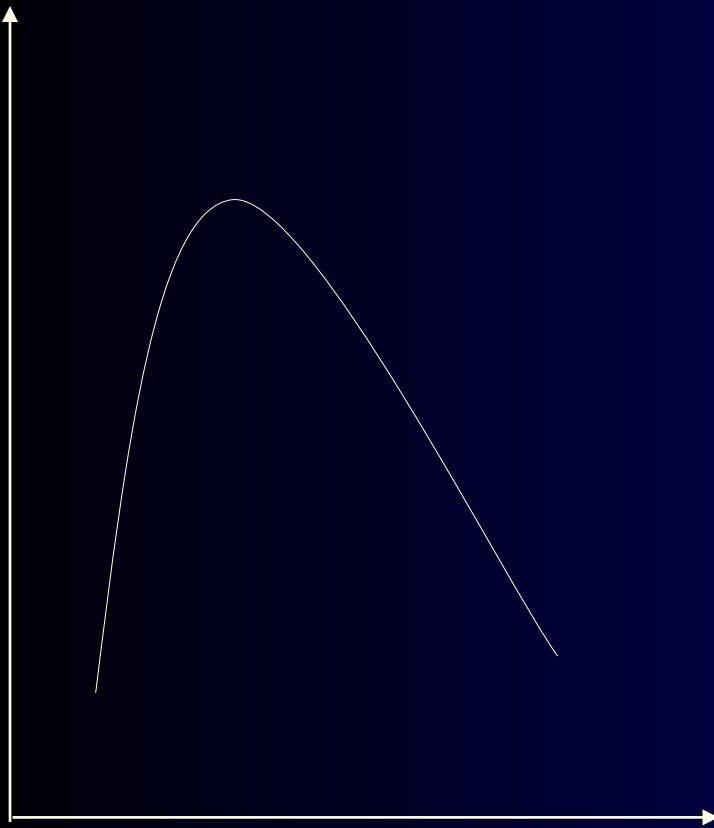
$$\beta_2 = \frac{b_2 \times \sigma_{x_2}}{\sigma_y} = 1.16554 \times \sqrt{\frac{\sum x_2^2 - 7\bar{x}_2^2}{\sum y^2 - 7\bar{y}^2}} = +0.9330$$

III. Liên hệ tương quan phi tuyến

- Nghiên cứu một số hàm hồi quy phi tuyến (đối biến độc lập):
- **Hàm parabol/bậc 2:** Vận dụng khi tiêu thức nguyên nhân tăng hay giảm với 1 lượng đều nhau thì tiêu thức kết quả biến động với 1 lượng không đều nhau (nhanch hơn hoặc chậm hơn)

$$\hat{y}_x = b_0 + b_1x + b_2x^2$$

Dạng số liệu khi vẽ



- Tiến hành các bước phân tích tương tự.
- b_0, b_1, b_2 là các tham số của mô hình hồi quy, được xác định bằng phương pháp bình phương nhỏ nhất và phải thỏa mãn hệ phương trình

$$\begin{cases} \sum y = nb_0 + b_1 \sum x + b_2 \sum x^2 \\ \sum xy = b_0 \sum x + b_1 \sum x^2 + b_2 \sum x^3 \\ \sum x^2 y = b_0 \sum x^2 + b_1 \sum x^3 + b_2 \sum x^4 \end{cases}$$

Dạng Hypebol:

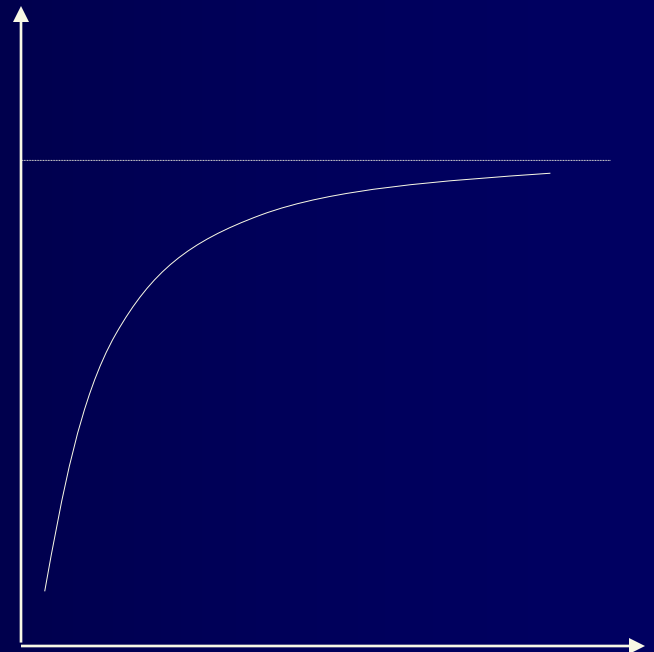
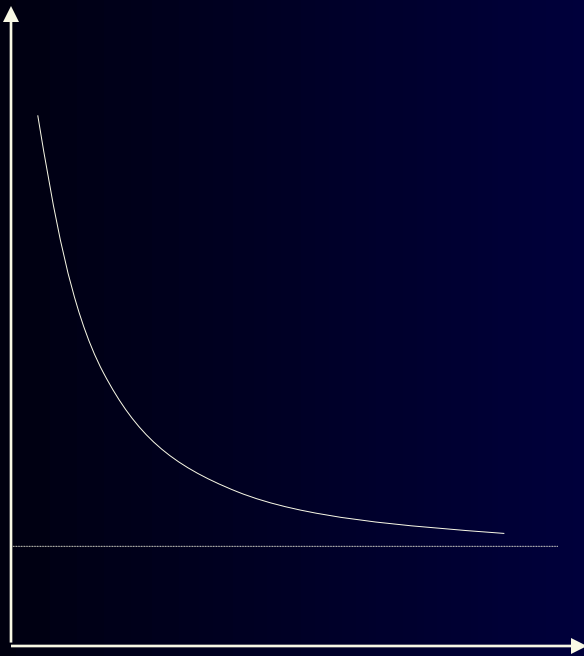
$$\hat{y}_x = b_0 + \frac{b_1}{x}$$

- Vận dụng khi tiêu thức nguyên nhân tăng thì tiêu thức kết quả giảm với tốc độ không đều.
- Tiến hành các bước phân tích tương tự
- Hệ phương trình xác định tham số

$$\sum y = b_0 n + b_1 \sum \frac{1}{x}$$

$$\sum \frac{1}{x} y = b_0 \sum \frac{1}{x} + b_1 \sum \frac{1}{x^2}$$

Dạng số liệu khi vẽ để xác định mô hình Hypebol

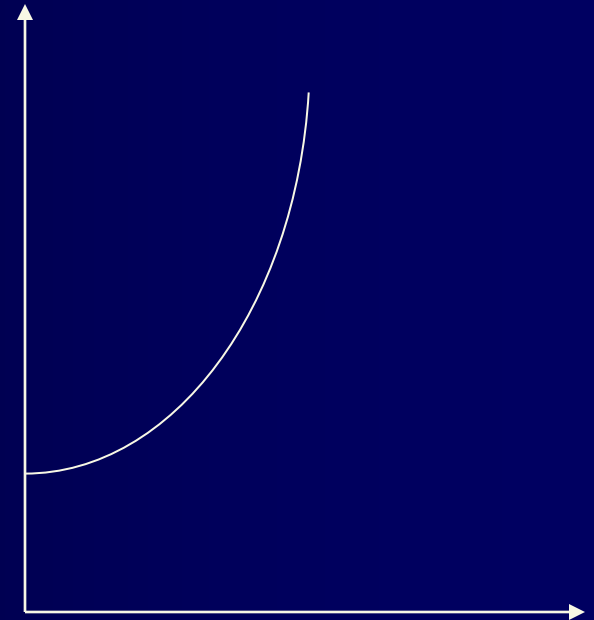


Phương trình hàm mũ

- Vận dụng khi trị số của tiêu thức kết quả thay đổi theo cấp số nhân.
- Mô hình hồi quy

$$\hat{y}_x = b_0 \cdot b_1^x$$

- *Giải bài toán cực trị này thế nào?*



- Biến đổi sang dạng tuyến tính:
 - $\ln y = \ln b_0 + x \cdot \ln b_1$
- $\ln b_0, \ln b_1$ phải thỏa mãn hệ phương trình:
 - $\sum \ln y = \ln b_0 \cdot n + \ln b_1 \cdot \sum x$
 - $\sum x \ln y = \ln b_0 \cdot \sum x + \ln b_1 \cdot \sum x^2$

Tỷ số tương quan

Tỷ số tương quan là chỉ tiêu đánh giá trình độ chặt chẽ của mối liên hệ tương quan phi tuyến

Các công thức

$$\eta = \sqrt{\frac{\sigma_y^2 - \sigma_{y(x)}^2}{\sigma_y^2}}$$

$$\eta = \sqrt{\frac{\sigma_{yx}^2}{\sigma_y^2}}$$

$$\eta = \sqrt{1 - \frac{\sigma_{y(x)}^2}{\sigma_y^2}}$$

$$\eta = \sqrt{1 - \frac{\sum (y - \hat{y}_x)^2}{\sum (y - \bar{y})^2}}$$

Giải thích ý nghĩa các ký hiệu trong CT

- σ_y^2 = phương sai của tiêu thức kết quả/phụ thuộc y
- $\sigma_{y(x)}^2$ = phương sai của tt kết quả y do ảnh hưởng của các tt nguyên nhân khác ngoài x
- σ_{yx}^2 = phương sai của tt kết quả y do ảnh hưởng của x

$$\sigma_{y(x)} = \sqrt{\frac{\sum (y - \hat{y}_x)^2}{n}}$$

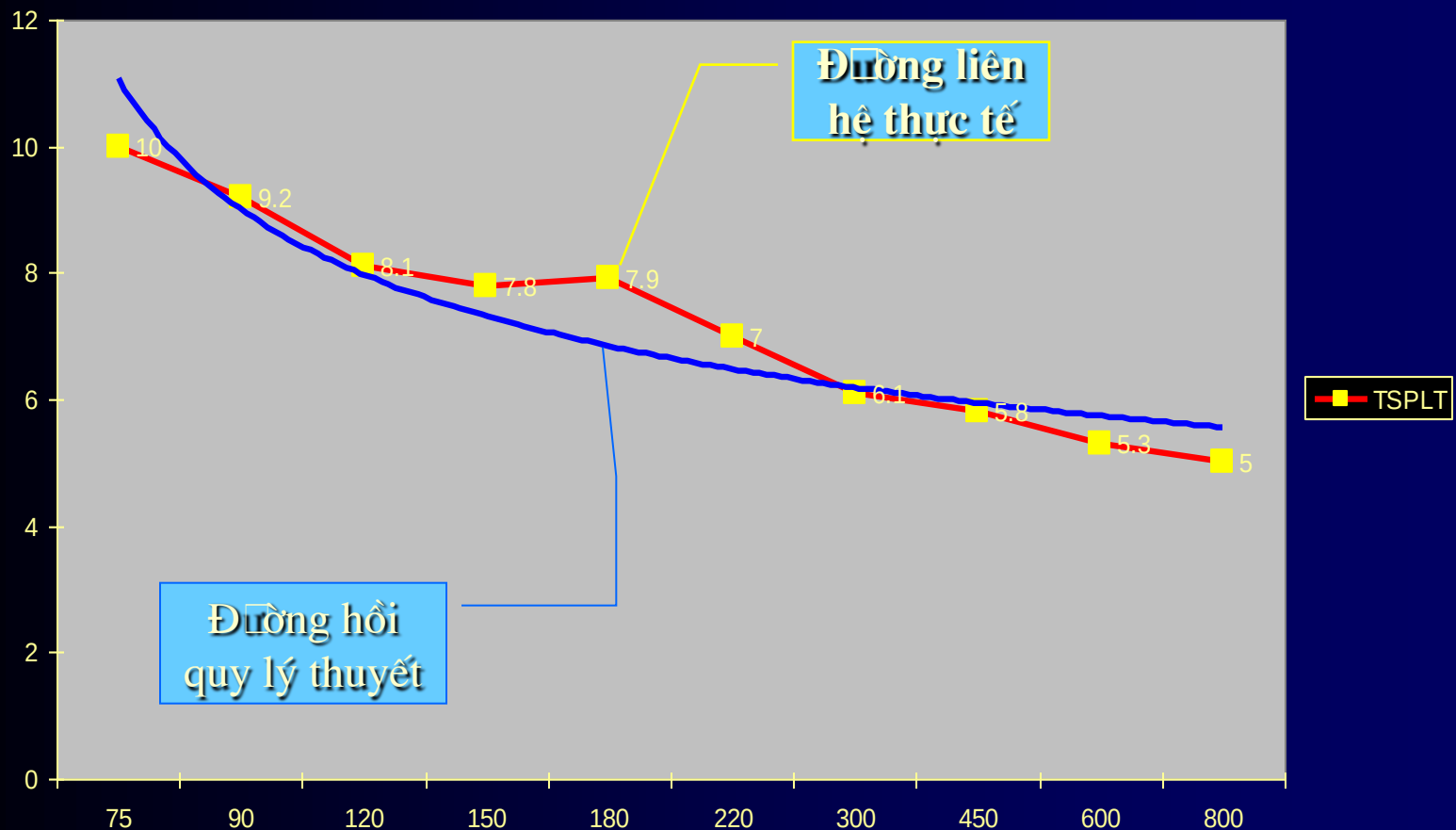
$$\sigma_y = \sqrt{\frac{\sum (y - \bar{y})^2}{n}}$$

$$\sigma_{yx}^2 = \frac{\sum (\hat{y} - \bar{y})^2}{n}$$

Bài tập thực hành 3

Mức tiêu thụ hàng hoá (trVND)	75	90	120	150	180	220	300	450	600	800
Tỷ suất phí lưu thông (%)	10.0	9.2	8.1	7.8	7.9	7.0	6.1	5.8	5.3	5.0

Biểu diễn mối liên hệ giữa 2 tiêu thức



Tiêu thức nguyên nhân: Mức tiêu thụ: x
Tiêu thức kết quả: Tỷ suất phí LT: y

- Đường hồi quy lý thuyết là đường hyperbol được biểu diễn bằng hàm số:

- $$\hat{y}_x = b_0 + \frac{b_1}{x}$$

- trong đó: x: tt nguyên nhân
- y: tt kết quả
- b_0 : tham số tự do
- b_1 : hệ số hồi quy phi tuyến

Dùng phương pháp bình phương nhỏ nhất để xác định giá trị của b_0, b_1

Giải hệ phương trình để xác định giá trị của b_0, b_1

$$\sum y = b_0 n + b_1 \sum \frac{1}{x}$$

$$\sum \frac{1}{x} y = b_0 \sum \frac{1}{x} + b_1 \sum \frac{1}{x^2}$$

x	y	1/x	1/x ²	y/x
0.75	10.0	1.33	1.7778	13.3333
0.90	9.2	1.11	1.2346	10.2222
1.20	8.1	0.83	0.6944	6.7500
1.50	7.8	0.67	0.4444	5.2000
1.80	7.9	0.56	0.3086	4.3889
2.20	7.0	0.45	0.2066	3.1818
3.00	6.1	0.33	0.1111	2.0333
4.50	5.8	0.22	0.0493	1.2889
6.00	5.3	0.17	0.0278	0.8833
8.00	5.0	0.13	0.0156	0.6250
	72.2	5.80	4.8702	47.9067

$$\begin{cases} 72.2 = 10b_0 + 5.80b_1 \\ 47.9067 = 5.80b_0 + 4.8702b_1 \end{cases} \Rightarrow \begin{cases} b_0 = 4.898 \\ b_1 = 4.004 \end{cases}$$

Phương trình hồi quy lý thuyết có dạng:

$$y = 4,898 + 4,004 \cdot 1/x$$

Đánh giá trình độ chặt chẽ của liên hệ

x	y	$\hat{y}_Q$	$(\hat{y} - \bar{y})^2$	y^2
0.75	10.0	10,1853	8,7932	100,00
0.90	9.2	9,3071	4,3560	84,64
1.20	8.1	8,2093	0,9788	65,61
1.50	7.8	7,5507	0,1093	60,84
1.80	7.9	7,1116	0,0118	62,41
2.20	7.0	6,7124	0,2577	49,00
3.00	6.1	6,2333	0,9735	37,21
4.50	5.8	5,7942	2,0328	33,64
6.00	5.3	5,5747	2,7071	28,09
8.00	5.0	5,4100	3,2761	25,00
			23,4964	546,44

$$\eta = \sqrt{\frac{\sigma_{y_x}^2}{\sigma_y^2}} = \sqrt{\frac{\Sigma(\hat{y} - \bar{y})^2}{\Sigma y^2 - n * (\bar{y})^2}}$$

$$\eta = \sqrt{\frac{23,4964}{546,44 - 10 * 7,22^2}} = \sqrt{\frac{23,4964}{25,1560}} = 0,966$$