

Ch- ơng 6

HỒI QUY – T_ƯƠNG QUAN

- Mỗi liên hệ giữa các hiện t- ợng và nhiệm vụ của ph- ơng pháp hồi quy - t- ơng quan.
- Liên hệ t- ơng quan tuyến tính

I. Mối liên hệ giữa các hiện tượng và nhiệm vụ của phương pháp hồi quy - tương quan

1. Mối liên hệ giữa các hiện tượng

- Các hiện tượng tồn tại trong mối liên hệ phổ biến và ràng buộc lẫn nhau.
- Xét theo mức độ chặt chẽ, có thể phân mối liên hệ thành 2 loại:
 - + Liên hệ hàm số
 - + Liên hệ tương quan

1. Mối liên hệ giữa các hiện tượng

- Liên hệ hàm số:

- Dạng TQ của liên hệ hàm số: $y = f(x)$

- Ví dụ:

- Hoàn toàn chặt chẽ?

1. Mối liên hệ giữa các hiện tượng

Liên hệ tượng quan:

- *T/c không hoàn toàn chặt chẽ:*

cuu duong than cong. com

■ Liên hệ t- ơng quan



Ví dụ:

■ *Nguyên nhân:*

cuu duong than cong. com

2. Nhiệm vụ của ph- ơng pháp hồi quy ☐ t- ơng quan


- Xác định ph- ơng trình HQ phản ánh mối liên hệ

cuu duong than cong. com

- Đánh giá mức độ chặt chẽ của mối liên hệ

cuu duong than cong. com

3. □ nghĩa của việc nghiên cứu HQ -TQ

- 
- Nghiên cứu mối liên hệ giữa các hiện t- ợng.
 - Xây dựng các mô hình KT
 - Tiến hành một số dự đoán.
 - Dùng trong một số ph- ơng pháp TK khác

4. Các bước tiến hành



cuu duong than cong. com

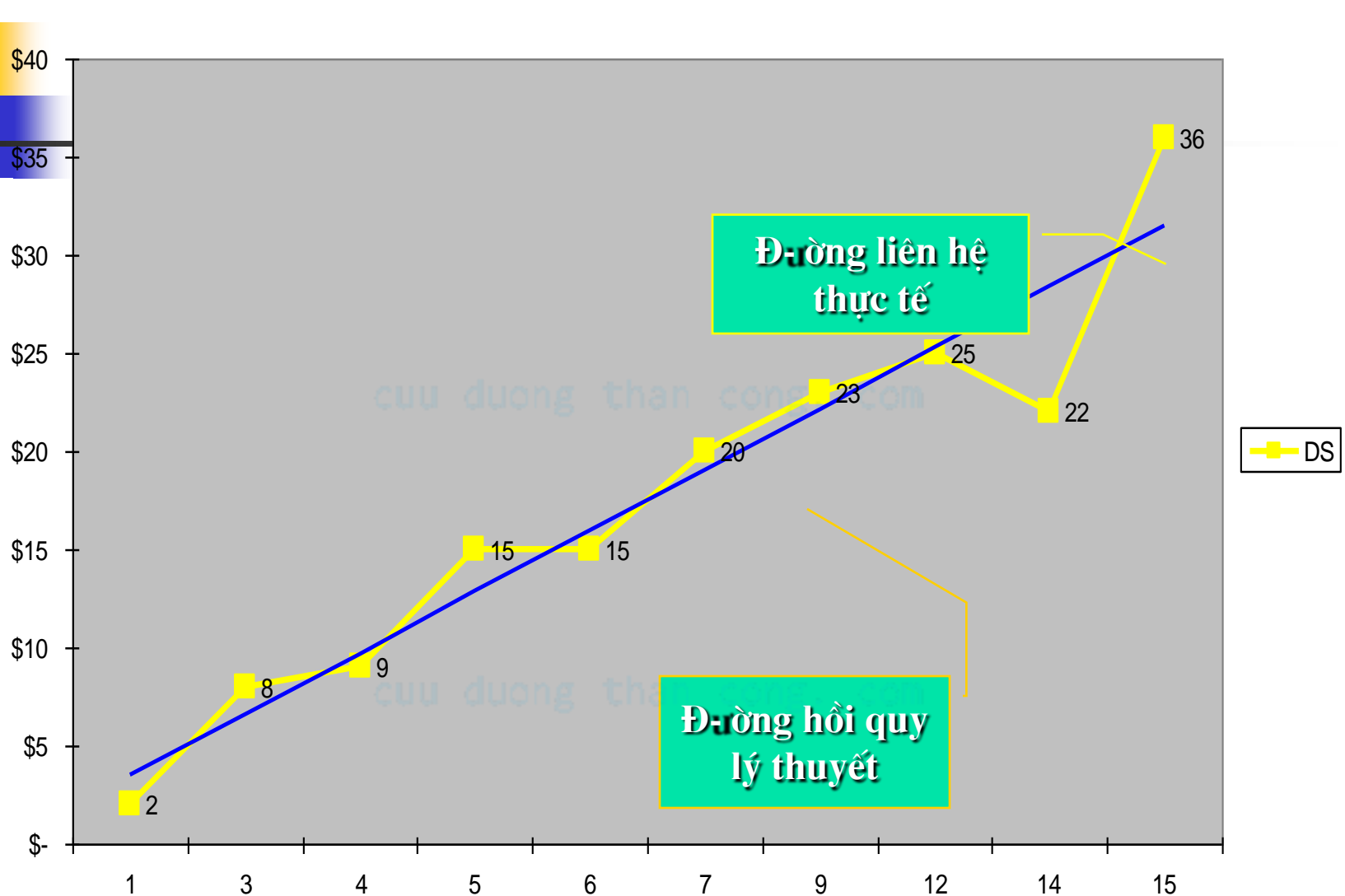
cuu duong than cong. com

II. Liên hệ t- ơng quan tuyến tính

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



Tiêu thức nguyên nhân:

Tiêu thức kết quả:

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

Trong đó:

x :

\hat{y} :

y :

a :

b :

Xác định giá trị của a và b

Ph- ơng pháp th- ờng dùng: *ph ơng pháp bình phương nhỏ nhất.*

■ *Nội dung:*

$$Q = \sum (y - \hat{y})^2 = \sum (y - a - bx)^2 = \sum e_i^2 \quad \min$$

(e_i : phần d)

$$\Leftrightarrow \begin{cases} \frac{\partial Q}{\partial a} = 0 \\ \frac{\partial Q}{\partial b} = 0 \end{cases} \Rightarrow \begin{cases} \sum y = na + b \sum x \\ \sum xy = a \sum x + b \sum x^2 \end{cases}$$

Xác định giá trị của a và b

- Giải hệ phương trình để xác định giá trị của a, b

cuu duong than cong. com

cuu duong than cong. com

áp dụng cho VD trên

x	y	xy	x^2	y^2
1	2			
3	8			
4	9			
5	15			
6	15			
7	20			
9	23			
12	25			
14	22			
15	36			
76	175			



Giải hệ ph- ơng trình

- Thay số:

[cuu duong than cong. com](http://cuuduongthancong.com)

- Giải hệ:

[cuu duong than cong. com](http://cuuduongthancong.com)

- Ph- ơng trình HQ:

⇒ Có thể xác định đ- ợc a, b bằng cách sử dụng công thức

cuu duong than cong. com

cuu duong than cong. com



Ý nghĩa của tham số: a ? b ?

- a (*hệ số tự do*):



- b (*hệ số HQ*):

Tính chất của mô hình HQ mẫu

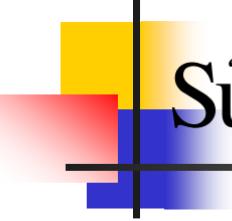
+ Đ- ờng HQ mẫu đi qua điểm (\bar{x}, \bar{y})

nghĩa là: $\bar{y} = a + b \bar{x}$

$$+ \hat{\bar{y}} = \bar{y}$$

$$+ \sum e_i = 0$$

$$+ \sum e_i^2 \text{ min}$$



➤ **Đánh giá mức độ chặt chẽ của liên hệ**
Sử dụng hệ số t- ơng quan r:

cuu duong than cong. com

Nhân xét

cuu duong than cong. com



Ý nghĩa của hệ số t- ơng quan

■ Biểu thị c- ờng độ của liên hệ

cuu duong than cong. com

■ Biểu hiện tính chất của liên hệ

cuu duong than cong. com

Đánh giá mức độ chặt chẽ của mối liên hệ giữa CP quảng cáo và doanh số

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

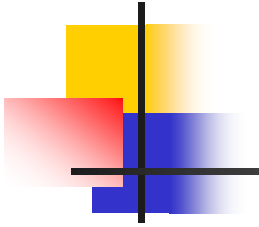
cuu duong than cong. com



Bài tập

Mức tiêu thụ hàng hoá (trVND)	75	90	120	150	180	220	300	450	600	800
Tỷ suất phí l- u 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



cuu duong than cong. com

cuu duong than cong. com

Tiêu thức nguyên nhân:

Tiêu thức kết quả:

➤ Đ- ồng hồi quy lý thuyết

Trong đó:

$x:$

$\hat{y} :$

$y:$

$a :$

$b:$



Dùng phương pháp bình phương nhỏ nhất để xác định giá trị của các tham số a và b

Giải hệ phương trình để xác định giá trị của a, b

cuu duong than cong. com

cuu duong than cong. com



x	y			
75	10.0			
90	9.2			
120	8.1			
150	7.8			
180	7.9			
220	7.0			
300	6.1			
450	5.8			
600	5.3			
800	5.0			

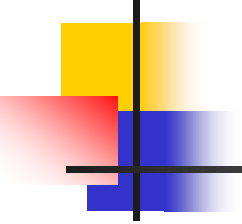


Xác định giá trị của a , b

- cuu duong than cong. com
- Ph- ơng trình hồi quy lý thuyết có dạng:

- cuu duong than cong. com
- ý nghĩa của a và b

Đánh giá trình độ chặt chẽ của liên hệ


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

cuu duong than cong. com

III. Liên hệ t- ơng quan phi tuyến

- Một số hàm hồi quy phi tuyến:
 - Hàm parabol: $y = a + bx + cx^2$
 - Hàm hyperpol: $y = a + b.1/x$
- Tỷ số t- ơng quan: đánh giá trình độ chặt chẽ của liên hệ


$$y = a + bx + cx^2$$

- Tiến hành các bước phân tích tổng tự
- Hệ phương trình xác định tham số

$$\begin{cases} \sum y = na + b \sum x + c \sum x^2 \\ \sum xy = a \sum x + b \sum x^2 + c \sum x^3 \\ \sum x^2 y = a \sum x^2 + b \sum x^3 + c \sum x^4 \end{cases}$$


$$y = a + b \cdot 1/x$$

- Tiến hành các bước phân tích tương tự
- Hệ phương trình xác định tham số

$$\begin{cases} \sum y = na + b \sum \frac{1}{x} \\ \sum y \frac{1}{x} = a \sum \frac{1}{x} + b \sum \frac{1}{x^2} \end{cases}$$



Tỷ số t- ơng quan

Các công thức:

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

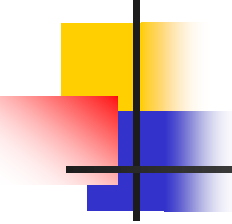
$$\eta = \sqrt{\frac{\sigma_{yx}^2}{\sigma_y^2}}$$

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

Giải thích ý nghĩa các ký hiệu trong CT

- σ^2_y = ph- ơng sai của tiêu thức kết quả y
- σ^2_{yx} = ph- ơng sai của tt kết quả y do ảnh h- ớng của x
- $\sigma^2_{y(x)}$ = ph- ơng sai của tt kết quả y do ảnh h- ớng của các tt nguyên nhân khác ngoài x

Công thức tính σ_y^2 ; σ_{yx}^2 ; $\sigma_{y(x)}^2$


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

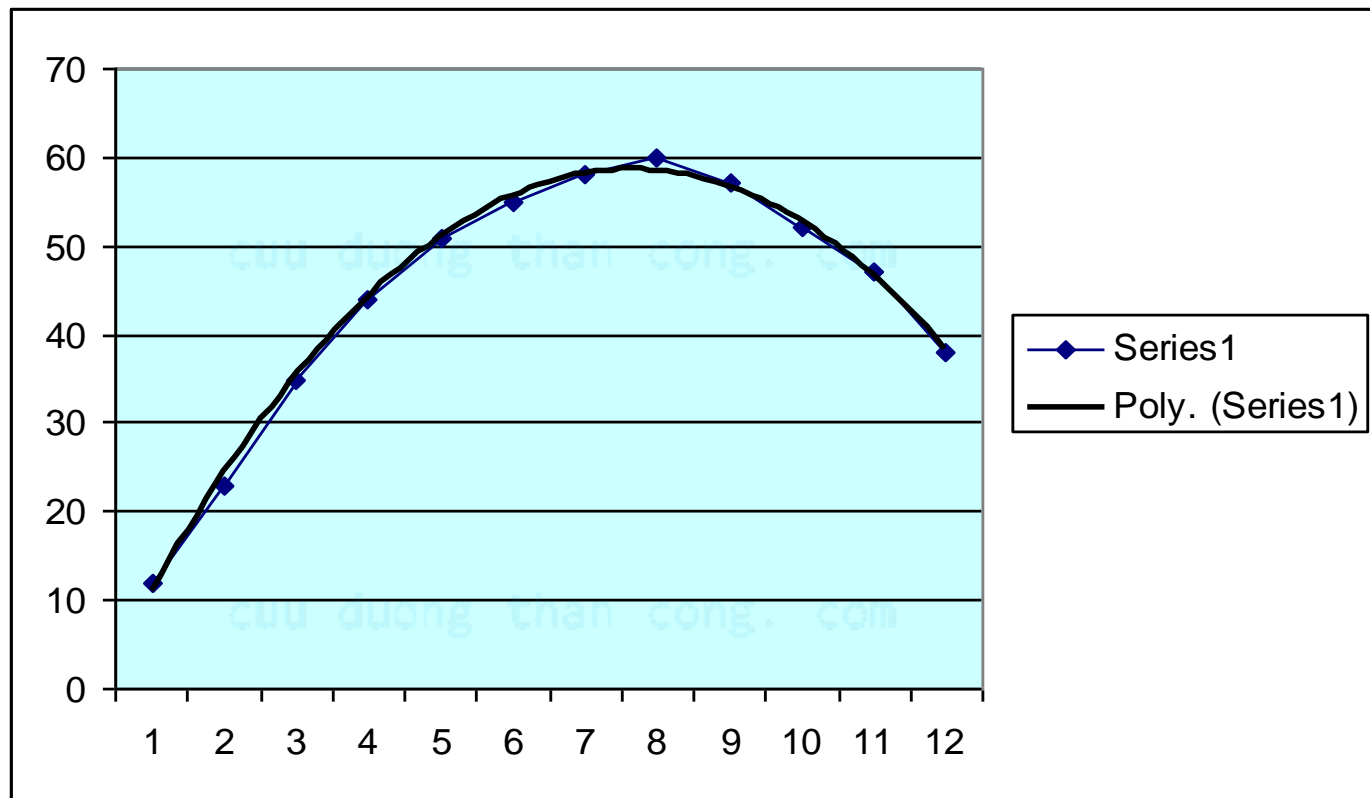
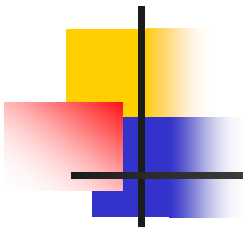
$$\sigma_{yx}^2 = \frac{\sum (\hat{y} - \bar{y})^2}{n}$$

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



Bài tập

Tuổi nghề (năm)	3	6	9	12	15	18	21	24	27	30	33	36
Năng suất LĐ (sp/tg)	12	23	35	44	51	55	58	60	57	52	47	38



- 
-
- Hàm hồi quy lý thuyết có dạng

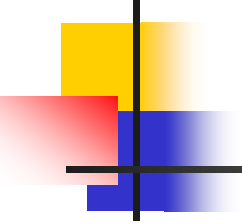
$$y = a + bx + cx^2$$

trong đó:

- tiêu thức nguyên nhân: tuổi nghề: x
- tiêu thức kết quả: NSLĐ: y

x	y	x^2	x^3	x^4	xy	x^2y
3	12	9	27	81	36	108
6	23	36	216	1296	138	828
9	35	81	729	6561	315	2835
12	44	144	1728	20736	528	6336
15	51	225	3375	50625	765	11475
18	55	324	5832	104976	990	17820
21	58	441	9261	194481	1218	25578
24	60	576	13824	331776	1440	34560
27	57	729	19683	531441	1539	41553
30	52	900	27000	810000	1560	46800
33	47	1089	35937	1185921	1551	51183
36	38	1296	46656	1679616	1368	49248

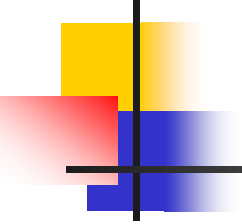
Các giá trị tính đ- ợc

- 
-
- $\Sigma x = 234$
 - $\Sigma y = 532$
 - $\Sigma x^2 = 5850$
 - $\Sigma x^3 = 164268$
 - $\Sigma x^4 = 4917510$
 - $\Sigma xy = 11448$
 - $\Sigma x^2y = 288324$

Hệ phương trình xác định giá trị a,b,c

$$\begin{cases} \sum y = na + b \sum x + c \sum x^2 \\ \sum xy = a \sum x + b \sum x^2 + c \sum x^3 \\ \sum x^2 y = a \sum x^2 + b \sum x^3 + c \sum x^4 \end{cases}$$

$$\begin{cases} 532 = 12a + 234b + 5850c \\ 11448 = 234a + 5850b + 164268c \\ 288324 = 5850a + 164268b + 4917510c \end{cases}$$


$$\begin{cases} a = -4.545 \\ b = 5.492 \\ c = -0.119 \end{cases}$$

- Phương trình hồi quy có dạng:

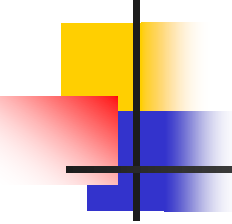
$$\hat{y} = -4,545 + 5,492 x + 0,119 x^2$$



Xác định tỷ số t-ơng quan

$$\eta = \sqrt{1 - \frac{\sigma_{y(x)}^2}{\sigma_y^2}} = \sqrt{1 - \frac{\sum y^2 - \hat{y}^2}{\sum y^2 - n * (\bar{y})^2}}$$

x	y	\hat{y}	$(y - \hat{y})^2$	y^2
3	12	10,680	1,2996	144
6	23	24,123	1,2611	529
9	35	35,244	0,0595	1225
12	44	44,223	0,0497	1936
15	51	51,060	0,0036	2601
18	55	55,755	0,5700	3025
21	58	58,308	0,0949	3364
24	60	58,719	1,6410	3600
27	57	56,988	0,0001	3249
30	52	53,115	1,2432	2704
33	47	47,100	0,0100	2209
36	38	38,943	0,8893	1444



Đánh giá trình độ chặt chẽ của mlh

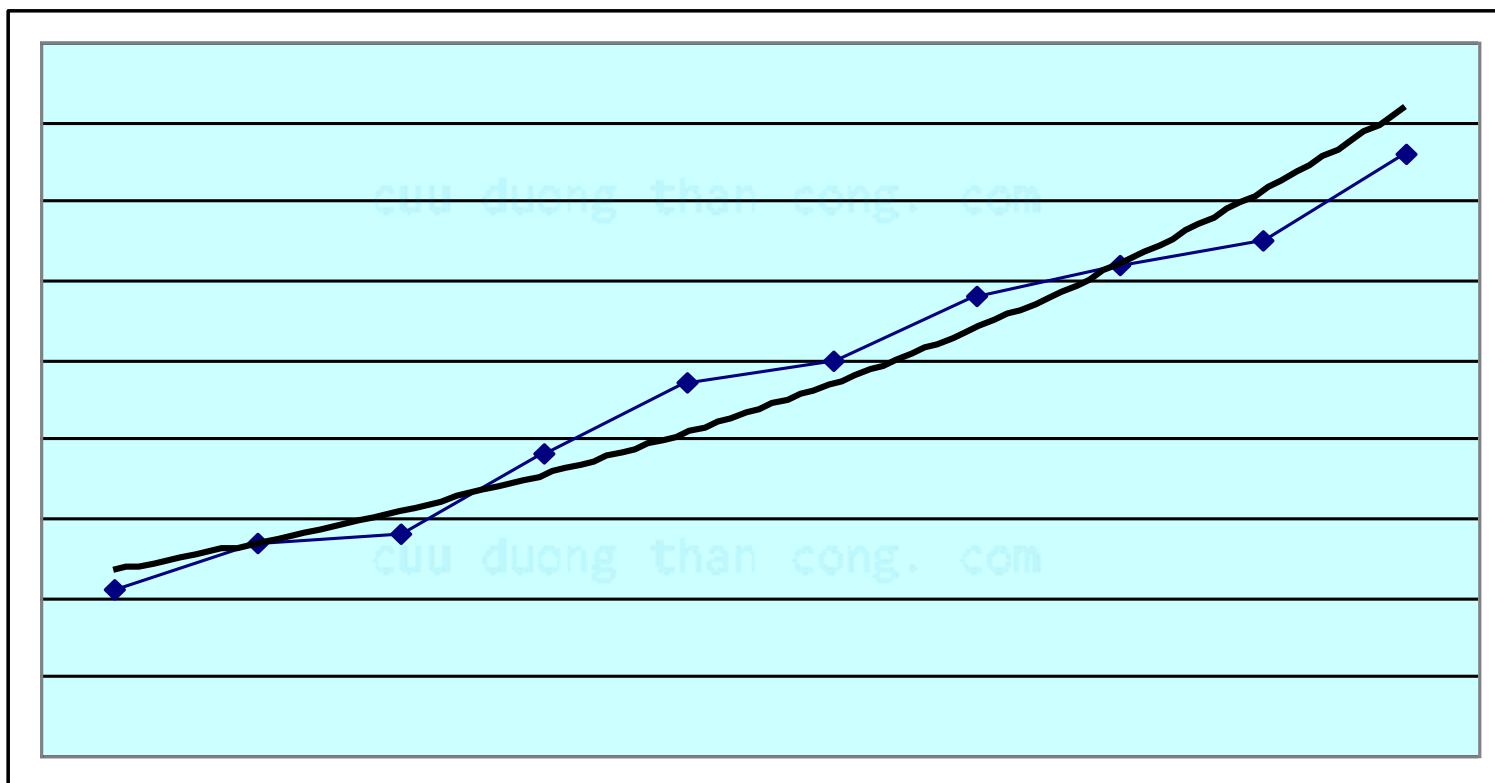
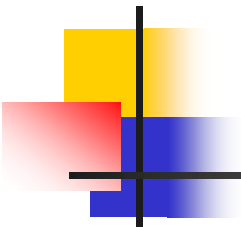
$$\eta = \sqrt{1 - \frac{\sum y - \hat{y}}{\sum y^2 - n * (\bar{y})^2}}$$


$$\eta = \sqrt{1 - \frac{7,122}{26030 - 12 * 1965,44}} = 0,998$$




Bài tập

Giá trị xuất khẩu (\$)	32	42	43	52	70	70	75	92	100	115
Chi phí l- u thông (\$)	2.1	2.7	2.8	3.8	4.7	5.0	5.8	6.2	6.5	7.6






x	y	1/x	1/x ²	y/x
32	2.1	0.03125	0.000976563	0.06563
42	2.7	0.02381	0.000566893	0.06429
43	2.8	0.02326	0.000540833	0.06512
52	3.8	0.01923	0.000369822	0.07308
70	4.7	0.01429	0.000204082	0.06714
70	5.0	0.01429	0.000204082	0.07143
75	5.8	0.01333	0.000177778	0.07733
92	6.2	0.01087	0.000118147	0.06739
100	6.5	0.0100	0.000100000	0.06500
115	7.6	0.0087	0.00007569	0.06609
	47.2	0.16902	0.003333814	0.68249



$$\begin{cases} \sum y = na + b \sum \frac{1}{x} \\ \sum y \frac{1}{x} = a \sum \frac{1}{x} + b \sum \frac{1}{x^2} \end{cases}$$

$$\begin{cases} 47.2 = 10a + 0.16902b \\ 0.68249 = 0.16902a + 0.003333814b \end{cases}$$

$$\begin{cases} a = 8.805 \\ b = -241.667 \end{cases}$$



x	y	\hat{y}	$(y - \hat{y})^2$	y^2
32	2.1	1.25291	0.71757	4.41
42	2.7	3.05102	0.12322	7.29
43	2.8	3.18484	0.1481	7.84
52	3.8	4.15756	0.12785	14.44
70	4.7	5.35261	0.42591	22.09
70	5.0	5.35261	0.12434	25.00
75	5.8	5.58277	0.04719	33.64
92	6.2	6.17818	0.00048	38.44
100	6.5	6.38833	0.01247	42.25
115	7.6	6.70355	0.80363	57.76
	47.2		2.53073	253.16

Đánh giá trình độ chặt chẽ của liên hệ

$$\eta = \sqrt{1 - \frac{\sigma^2_{y(x)}}{\sigma^2_y}} = \sqrt{1 - \frac{\sum y - y^{LT}^2}{\sum y^2 - n * (\bar{y})^2}}$$

$$\eta = \sqrt{1 - \frac{2.53073}{253.16 - 10 * 4.72^2}} = 0.957$$

IV. Liên hệ t- ơng quan tuyến tính đa biến

- Nghiên cứu mối liên hệ giữa nhiều tiêu thức nguyên nhân với một tiêu thức kết quả
- Hàm số HQ biểu diễn mối liên hệ

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

Đánh giá trình độ chặt chẽ của liên hệ

(Dùng hệ số t-ơng quan bội - R)

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

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

Đánh giá mức độ ảnh hưởng của từng tiêu thức nguyên nhân tới tiêu thức kết quả (tham số t-ơng quan chuẩn hóa — beta)

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

cuu duong than cong. com

- Với:
- β_j : tham số t-ơng quan chuẩn hoá của tiêu thức nguyên nhân j
 - b_j : tham số hồi quy của tiêu thức nguyên nhân j
 - σ_{x_j} : độ lệch tiêu chuẩn của tiêu thức nguyên nhân j
 - σ_y : độ lệch tiêu chuẩn của tiêu thức kết quả y



Bài tập

Tiền l- ơng (\$/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



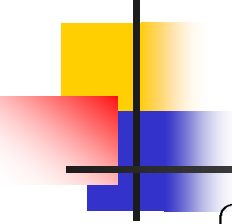
Phân tích

- Tt nguyên nhân 1: tiền l- ơng 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:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2$$

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


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

$$\begin{cases} 54,8 = 7b_0 + 8,69b_1 + 42,4b_2 \\ 68,644 = 8,69b_0 + 10,8703b_1 + 53,123b_2 \\ 336,89 = 42,2b_0 + 53,123b_1 + 260,84b_2 \end{cases} \quad \begin{cases} b_0 = 0.06323 \\ b_1 = 0.56831 \\ b_2 = 1.16554 \end{cases}$$

- 
- Ph- ơng trình hồi quy tuyến tính đa biến có dạng:

$$\hat{y} = 0.06323 + 0.56831x_1 + 1.16554x_2$$

- Đánh giá trình độ chặt chẽ của liên hệ (dùng hệ số t- ơng quan bội R)
- Đánh giá mức độ ảnh h- ưởng (dùng hệ số t- ơng quan chuẩn hóa β_j).

x_1	x_2	y	y^2	x_1^2	x_2^2	\hat{y}	$(\hat{y} - \bar{y})^2$
1.00	5.0	6.5	42.25	1.0000	25.00	6.45923	1.879010393
1.20	5.3	6.9	47.61	1.4400	28.09	6.92255	0.823461873
1.25	5.5	7.2	51.84	1.5625	30.25	7.18408	0.417219106
1.27	6.0	7.8	60.84	1.6129	36.00	7.77821	0.002682101
1.30	6.7	8.3	68.89	1.6900	44.89	8.61114	0.610176575
1.32	6.9	8.9	79.21	1.7424	47.61	8.85561	1.051879975
1.35	7.0	9.2	84.64	1.8225	49.00	8.98922	1.343779416
8.69	42.4	54.8	435.28	10.8703	260.84		6.128209438



Hệ số t-ơng quan bội

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

$$R = \sqrt{\frac{6.1282}{435.28 - 7 * 7.828^2}} = 0,983$$



Tham số t- ơng quan chuẩn hoá

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

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

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