

Bài tập

1. Vẽ đám mây điểm và tính hệ số tương quan của các tập số liệu sau :

a)

x	2	5	7	10	11
y	10	20	35	50	65

b)

x	5	8	10	12	15	22
y	50	42	40	35	33	28

2. Một công ty nhỏ quan tâm tới việc phân tích hiệu quả của việc quảng cáo. Trong thời gian 5 tháng công ty thu được kết quả sau

x	5	8	10	15	22
y	6	15	20	30	39

trong đó x là số tiền chi vào quảng cáo (đơn vị là trăm USD) còn y là tổng doanh thu (đơn vị là nghìn USD). Hãy xác định hệ số tương quan.

3. Một trường đại học thu thập các số liệu về số chứng chỉ mà một sinh viên theo học và số giờ học ở nhà của anh ta trong một tuần :

X	20	25	30	50	20	23
Y	12	13	12	15	16	16

ở đó X là số giờ học, Y là số chứng chỉ. Tìm hệ số tương quan giữa X và Y. Ở mức ý nghĩa $\alpha = 5\%$, có sự tương quan tuyến tính giữa hai biến nói trên không ?

4. Một nghiên cứu được tiến hành ở Mỹ để xác định mối quan hệ giữa chiều cao của một người và cỡ giày của họ.

Nhà nghiên cứu đã thu được số liệu sau

X	66	63	67	71	62
Y	9	7	$8\frac{1}{2}$	10	6

X	65	72	68	60	66
Y	$8\frac{1}{2}$	12	$10\frac{1}{2}$	$5\frac{1}{2}$	8

trong đó X là chiều cao (đơn vị là inches) còn Y là cỡ giày. Hãy tính hệ số tương quan giữa X và Y.

5. Tuổi và huyết áp của 10 bệnh nhân trẻ em (dưới 14 tuổi) chọn ngẫu nhiên được cho trong bảng sau đây :

X	14	1	9	7	9	12	1	3	9	1
Y	100	83	112	152	104	90	92	85	120	130

trong đó X là tuổi còn Y là huyết áp.

Tìm đường hồi quy mẫu của Y đối với X.

Tính sai số tiêu chuẩn của đường hồi quy.

6. Một nhà xã hội học tuyên bố rằng kết quả học tập của sinh viên (đo bằng điểm thi tốt nghiệp) không liên quan tới thu nhập của gia đình họ. Với một mẫu ngẫu nhiên gồm 20 sinh viên, người ta tính được hệ số tương quan là 0,4. Với mức ý nghĩa 0,01 hãy kiểm định giả thiết "Điểm thi tốt nghiệp và thu nhập của gia đình một sinh viên là không tương quan".

7. Khảo sát 20 trận đấu bóng đá cho thấy hệ số tương quan giữa số lần sút bóng vào khung thành đối phương và số bàn thắng là 0,21.

Với mức ý nghĩa 5%, kiểm định giả thiết

H_0 : "Số lần sút bóng và số bàn thắng không tương quan".

8. Từ 13 cặp quan sát (x_i, y_i) sau đây từ cặp hai biến (X, Y), hãy kiểm định giả thiết

H_0 : "X và Y không tương quan với nhau".

X	123,0	356,9	111,1	118,0	
Y	4,15	4,10	3,74	3,89	
X	265	390	67	24	156,9
Y	3,75	3,10	3,13	2,97	2,96
X	106	86	85,5	241,2	
Y	2,92	2,33	2,26	2,03	

9. Tính tỉ số tương quan của Y đối với X và hệ số xác định của tập số liệu sau đây :

$(X, Y) = (15, 13), (25, 22), (10, 6), (15, 17), (20, 21), (10, 10), (20, 25), (25, 18), (30, 14), (30, 10).$

Có nhận xét gì về mối quan hệ giữa X và Y ?

10. Tính tỉ số tương quan của Y đối với X và hệ số xác định của tập số liệu sau đây :

X	Y	X	Y
50	15	210	255
130	115	90	115
170	215	240	255
270	335	50	35
90	95	240	275
210	295	270	315
50	55	130	135
130	155	270	355
270	295	90	135
240	315	240	295
170	175	130	175
210	275	170	235
90	75	50	75
		170	195
		210	235

Có kết luận gì về mối tương quan giữa X và Y (phi tuyến hay tuyến tính) ?

11. Cũng hỏi như bài tập 10 đối với tập số liệu sau đây :

X	Y	X	Y
50	75	90	135
130	235	90	175
210	255	270	115
240	195	130	255
90	115	50	15
210	295	270	75
240	215	130	295
50	35	170	355
240	235	210	275
170	295	270	95
270	135	170	335
210	315	170	315
50	55	90	155
240	175	130	275

12. Một nhà xã hội học muốn tìm hiểu mối quan hệ giữa các dạng tội phạm (hình sự và không hình sự) với tuổi của phạm nhân. Chọn ngẫu nhiên 100 phạm nhân trong hồ sơ của tòa án, ông ta thu được số liệu sau :

Dạng tội phạm \ Tuổi	Tuổi		
	Dưới 25	Từ 25 đến 49	Trên 50
Hình sự	15	30	10
Không hình sự	5	30	10

Số liệu này có cho phép ta kết luận có mối tương quan giữa tuổi và bản chất của tội phạm hay không ? Sử dụng mức ý nghĩa 5%.

13. Một nhà xã hội học chọn một mẫu ngẫu nhiên gồm 500 người để trao một bản thăm dò với các câu hỏi sau :

– Ông (bà) có đi nhà thờ không ? (có ☐ không ☐)

– Mức độ thành kiến về chủng tộc của ông (bà) thế nào ?

(Rất cao ☐ Đôi khi ☐ không ☐).

Kết quả được ghi lại trong bảng sau đây :

Đi nhà thờ	Mức độ thành kiến chủng tộc		
	Cao	Đôi khi	Không
Có	70	160	170
Không	20	50	30

Với mức ý nghĩa 5%, có nhận định gì với mối tương quan giữa việc đi nhà thờ và vấn đề thành kiến chủng tộc ?

14. Một cuộc thăm dò được tiến hành ở Mỹ bởi viện nghiên cứu xã hội học nổi tiếng Gallup để nghiên cứu mối quan hệ giữa nghề nghiệp của một người với quan niệm của anh ta về tiêu chuẩn đạo đức và tính trung thực. Kết quả của việc khảo sát một mẫu ngẫu nhiên gồm 380 người cho ta số liệu sau đây :

Nghề nghiệp	Quan niệm		
	Cao	Trung bình	Thấp
Bác sĩ	53	35	10
Luật sư	24	43	27
Nhà kinh doanh	18	55	20
Nhà chính trị	14	43	38

Với mức ý nghĩa 10%, hãy xác định xem có mối quan hệ hay không ?

15. Bảng sau đây cho ta số liệu về màu tóc của 422 người :

Màu tóc	Nam	Nữ
Đen	56	32
Hung	37	66
Nâu	84	90
Vàng	19	38

Với mức ý nghĩa 1%, nhận định xem liệu có mối quan hệ giữa màu tóc và giới tính hay không ?

16. Một nhóm gồm 105 nhà doanh nghiệp Mỹ được phân loại căn cứ theo thu nhập hàng năm và tuổi của họ. Kết quả thu được như sau :

Tuổi	Thu nhập		
	Dưới 100000\$	Từ 100000\$ – 399599\$	Trên 400000\$
Dưới 40	6	9	5
Từ 40 đến 54	18	19	8
Trên 54	11	12	17

Với mức ý nghĩa 1%, kiểm định giả thiết cho rằng tuổi và mức thu nhập không có quan hệ với nhau ?

17. Bảng sau đây cho ta tuổi (tính theo năm) và chiều cao thân răng của 15 con lừa. Tuổi được kí hiệu bởi X và chiều cao thân răng được ký hiệu Y :

X	0,06	0,08	0,08	0,18	0,28
Y	5,05	4,95	5,10	4,85	4,40

X	0,36	0,56	0,53	1,02	1,23
Y	5,25	4,4	5,15	4,3	4,15

X	1,77	1,94	2,14	2,44	2,44
Y	3,35	3,40	3,05	2,65	3,00

- a) Tìm đường hồi quy của Y đối với X.
- b) Tính sai số tiêu chuẩn của đường hồi quy.
- c) Kiểm định giả thiết rằng hệ số hồi quy lí thuyết khác không, với mức ý nghĩa $\alpha = 5\%$.

18. Đo đường kính X và chiều cao Y của 20 cây ta thu được số liệu sau :

X	2,3	2,5	2,6	3,1	3,4	3,7	7,3
Y	7	8	4	4	6	6	14

X	3,9	4	4,1	4,1	4,2	4,4
Y	12	8	5	7	8	7

X	4,7	5,1	5,5	5,8	6,2	6,9	6,9
Y	9	10	13	7	11	11	16

- Tìm đường hồi quy của Y đối với X .
- Tính sai số tiêu chuẩn của đường hồi quy.
- Tính tỉ số F để kiểm định giả thiết có hồi quy tuyến tính giữa Y với X .

19. Bảng sau đây cho ta một mẫu gồm 11 quan sát (x_i, y_i) từ tập hợp chính các giá trị của tập ĐLNN (X, Y) :

X	0,9	1,22	1,32	0,77	1,3	1,2
Y	-0,3	0,1	0,7	-0,28	-0,25	0,02

X	1,32	0,95	1,45	1,3	1,2
Y	0,37	-0,70	0,55	0,35	0,32

- Tìm đường hồi quy của Y đối với X .
- Tính sai số tiêu chuẩn của đường hồi quy.
- Tính tỉ số F để kiểm định sự đúng đắn của giả thiết : Có hồi quy tuyến tính của Y theo X .

20. Bảng sau đây ghi lại kỉ lục thế giới về chạy 1 dặm trong thế kỉ 20 :

Năm (X)	1911	1913	1915	1923
Thời gian (Y)	4 : 15.4	4 : 14.6	4 : 12.6	4 : 10.4

Năm (X)	1931	1933	1934	1937
Thời gian (Y)	4 : 09.2	4 : 07.6	4 : 06.8	4 : 06.4

X	1942	1942	1942	1943	1944
Y	4 : 06.2	4 : 06.2	4 : 04.6	4 : 02.6	4 : 01.6

X	1945	1954	1954	1957	1958
Y	4 : 01.4	3 : 59.4	3 : 58	3 : 57.2	3 : 54.5

X	1962	1964	1965	1966	1967
Y	3 : 54.4	3 : 54.1	3 : 53.6	3 : 51.3	3 : 51.1

X	1975	1975	1979	1980	1981
Y	3 : 51	3 : 49.4	3 : 49	3 : 48.9	3 : 48.8

X	1981	1981	1985		
Y	3 : 48.7	3 : 47.6	3 : 46.5		

- Tính hệ số tương quan giữa X và Y (đơn vị của Y là giây).
- Tìm đường thẳng hồi quy giữa Y và X.
- Tìm sai số tiêu chuẩn của đường thẳng hồi quy. Kiểm định giả thiết có hồi quy.
- Dựa trên phương trình này ước lượng kỷ lục thế giới năm 2050.
- Ước lượng năm mà kỷ lục thế giới là 3 phút 30 giây.
- Dựa trên phương trình hồi quy ước lượng kỷ lục thế giới năm 2500.

Kết quả này có hợp lí hay không ?

(Chú thích $3 : 48.7 = 3$ phút 48,7 giây).

21. Bảng sau đây cho số liệu về lượng cung tiền (Y) , tổng sản phẩm quốc gia (X) và lãi suất trái phiếu chính phủ (V) ở một quốc gia nhỏ ở Trung Mỹ trong 10 năm liên tiếp (1971–1980)

Năm	Y(triệu)	X(triệu)	V(%)
1971	57,2	485,5	4,01
1972	55,6	506,9	4,93
1973	54,4	540,2	4,18
1974	62,6	589,0	3,27
1975	67,2	637,4	4,35
1976	66,4	692,7	4,44
1977	70,4	745,3	4,40
1978	82,8	821,6	4,71
1979	102,4	906,3	3,33
1980	115,0	981,3	3,90

- a) Tìm phương trình mặt phẳng hồi quy của Y theo X và V.
b) Tính sai số tiêu chuẩn của mặt phẳng hồi quy.

22. Bảng sau đây cho số liệu về thu nhập trung bình một gia đình (Y) , số năm học trung bình (X) của người trên 25 tuổi và tỷ lệ dân số sống ở nông thôn (V) trong 9 vùng của nước Mỹ

STT	Y(đôla)	X	V(%)
1	6128	11,2	1,7
2	6211	10,5	2,1
3	6215	10,7	7,1
4	5154	10,7	18,4
5	4713	9,8	8,9
6	3793	8,8	17,3
7	4548	9,9	9,0
8	5660	12,0	8,3
9	6572	12,0	3,1

- a) Tìm phương trình mặt phẳng hồi quy của Y theo X và V.
b) Tính sai số tiêu chuẩn của mặt phẳng hồi quy.

23. Bảng sau đây cho số liệu về số tiền chi tiêu cho thực phẩm (Y), thu nhập sau thuế (X) (trong một tuần) và số người trong gia đình (V) của 10 gia đình ở nước Mỹ năm 1970.

STT	Y (đô la/tuần)	X (đô la/tuần)	V
1	25	80	2
2	36	70	4
3	40	100	4
4	47	150	3
5	49	180	3
6	50	200	2
7	53	190	3
8	55	130	5
9	60	225	4
10	65	175	5

- Tìm phương trình mặt phẳng hồi quy của Y theo X và V.
- Tính sai số tiêu chuẩn của mặt phẳng hồi quy.
- Giả sử một gia đình có thêm một người và thu nhập không đổi. Hãy dự đoán xem chi tiêu cho thực phẩm của gia đình sẽ tăng lên bao nhiêu ?

Đáp số và chỉ dẫn

- $r = 0,982$.
 - $r = -0,951$.
- $r = 0,985$.
- $r = 0,102$.
- $r = 0,951$.
- $y = 0,38x + 104,29$
 $s_{Y.X} = 23,3$.

6. $T = 1,85$; $c = 2,875$.

Chấp nhận giả thiết : Không có tương quan tuyến tính.

7. $T = 0,91$; $c = 2,101$.

Chấp nhận giả thiết không có tương quan tuyến tính.

8. $r = 0,222$;

$T = 0,755$;

$c = 2,201$

Chấp nhận giả thiết không có tương quan tuyến tính.

9. $\widehat{\eta^2} = 0,88$;

$r^2 = 0,1$.

X và Y có mối tương quan phi tuyến chặt, nhưng tương quan tuyến tính thì yếu.

10. $\widehat{\eta^2} = 0,947$;

$r^2 = 0,94$.

X và Y có tương quan tuyến tính chặt.

11. $\widehat{\eta^2} = 0,947$;

$r^2 = 0,08$.

Giữa X và Y có tương quan phi tuyến.

12. $T = 4,03$;

$c = 5,99$.

Chấp nhận H_0 .

13. $T = 5,308$;

$c = 5,991$.

Chấp nhận H_0 . Không có sự liên quan giữa hai dấu hiệu.

14. $T = 55,5$;

$$c = 10,645.$$

Bác bỏ H_0 . Có mối liên hệ.

15. $T = 10,17$;

$$c = 11,343.$$

Bác bỏ H_0 . Có tồn tại mối quan hệ.

16. $T = 6,65$;

$$c = 13,277.$$

Chấp nhận H_0 . Không có mối liên hệ.

17. $y = -0,934x + 5,162$;

$$s_{Y.X} = 0,307$$
 ;

$$|T| = 10,26$$
 ; $c = 2,16.$

Có tương quan tuyến tính.

18. $y = 1,677x + 1,045$;

$$s_{Y.X} = 2,22$$
 ;

$$F = 23,26.$$

Có mối tương quan tuyến tính.

19. $y = 1,548x - 1,739$;

$$s_{Y.X} = 0,29$$
 ;

$$F = 12,55.$$

Có mối tương quan tuyến tính.

20. a) $r = -0,989.$

b) $y = -0,4(x - 1900) + 260,2.$

c) $s_{Y.X} = 1,31.$

$$|T| = 40$$
 ; $c = 2,045.$ Bác bỏ $H_0.$

Hệ số hồi quy lí thuyết khác 0.

d) 3 phút 0,22 giây.

e) 2025.

f) 20,22 giây. Kết quả không hợp lí. Sử dụng phương trình hồi quy để dự báo trong một lần cận của năm 2000 thì mới chính xác.

21. a) $y = 17,0569 + 0,1114x - 4,9607v$

b) $\sum e_i^2 = 263,4297$; $s_{Y.X,V} = 6,124$

22. a) $y = 479,775 + 523,506x - 70,766v$

b) $\sum e_i^2 = 1270848$, $s_{Y.X,V} = 460,22$

23. a) $y = -1,2072 + 0,1824x + 6,2437v$

b) $\sum e_i^2 = 51,2539$, $s_{Y.X,V} = 2,7$

c) 6,2437